



جمهورية العراق
وزارة التعليم العالي والبحث العلمي
الجامعة المستنصرية
كلية التربية البدنية وعلوم الرياضة
الدراسات العليا _ الماجستير

البحث العلمي (البيانات الاصطناعية واستخدامها في البحث العلمي)

اعداد:

درر مؤيد إسماعيل
ايمان علاء الدين حسين
عباس سعد محمود
علي عبد الكريم حسين

واجب مقدم الى:

ا.د. غادة محمود
ا.م.د. حيدر مجيد

كجزء من متطلبات مادة البحث العلمي للدراسات العليا – الماجستير

المقدمة¹

في ظل التطورات المتسارعة التي يشهدها العالم الرقمي بات الذكاء الاصطناعي يلعب دور محوري في مختلف مجالات البحث العلمي فقد أتاح الذكاء الاصطناعي للباحثين الوصول الى أدوات وتقنيات متقدمة تساهم في تحليل البيانات وإدارة المراجع وتصميم التجارب واستخراج النتائج بدقة وسرعه، وبفضل هذه الأدوات لم يعد الباحث بحاجة الى قضاء ساعات طويلة في جمع وتحليل المعلومات والبيانات

وتعرف "البيانات الاصطناعية" وهي بيانات لا تجمع مباشرة من الواقع، وإنما تولد حاسوبياً بخوارزميات متقدمة بحيث تحاكي الخصائص الإحصائية للبيانات الحقيقية دون أن ترتبط بأفراد أو أحداث فعلية بعينها وتبرز أهميته هذا الموضوع في المجال الرياضي على وجه الخصوص نظراً لما تواجهه البحوث الرياضية من صعوبات مزمنة في الحصول على عينات كافية من الرياضيين النخبويين.

أو في توثيق الحالات النادرة كالإصابات الحرجة أو في إجراء دراسات طولية مكلفه زمنياً ومادياً، إذ توفر البيانات الاصطناعية بديل علمي واعداداً يتيح للباحث توليد كم كبير من البيانات الواقعية إحصائياً دون التقيد الكامل بمحدوديه العينات الحقيقيه أو مخاطر انتهاك خصوصية المشاركين

مدخل مفاهيمي إلى البيانات الاصطناعية

تعريف البيانات الاصطناعية

تُعرّف البيانات الاصطناعية بأنها مجموعة من البيانات يتم توليدها اصطناعياً باستخدام خوارزميات حاسوبية أو نماذج إحصائية أو نماذج تعلم آلي ولكنها تحتفظ بكافة الخصائص الإحصائية والعلاقات الرياضية للبيانات الواقعية"، بحيث تحاكي الأنماط والتوزيعات الإحصائية الموجودة في بيانات حقيقية أصلية، دون أن تحتوي على أي معلومة فعلية تخص أفراداً أو أحداثاً واقعية بعينها¹.

وتقوم الفكرة الجوهرية لهذا المفهوم على تدريب نموذج حاسوبي على مجموعة بيانات حقيقية تُعرف بمجموعة التدريب، بحيث يتعلم النموذج البنية الإحصائية الكامنة في تلك البيانات كالعلاقات بين المتغيرات وتوزيعاتها، ثم يُستخدم هذا النموذج بعد ذلك لتوليد عينات جديدة تحاكي هذه البنية دون أن تكون نسخة طبق الأصل من أي سجل حقيقي².

والمشكلة البحثية التقليدية : يواجه الباحثون ثلاث عقبات

1. شح البيانات

2. قيود الخصوصية وقوانين حماية البيانات

3. التكلفة العالية لعينات ميدانية

الحل الذكي لهذه المشاكل هي ان يأتي من خلال البيانات الاصطناعية كبديل رقمي ثوري يصنع خصيصاً لخدمة البحث العلمي

¹ معهد محترف في الإدارة؛ كيان متخصص في تصميم وتنفيذ برامج تحليل البيانات المندمجة بالذكاء التنافسي ، مقاله نشرت يوليو 2026

² فايد عبد الرزاق ؛ مجله روافد لدراسات والأبحاث في علوم الرياضة ، 2020.ص60

نشأة المفهوم وتطوره التاريخي

تعود الجذور الأولى لفكرة البيانات الاصطناعية إلى أساليب المحاكاة الإحصائية التقليدية كطريقة مونت كارلو (المستخدمة منذ منتصف القرن العشرين لتوليد بيانات عشوائية تحاكي توزيعات احتمالية معينة، غير أن المفهوم لم يتبلور بصورته الحديثة إلا مع ظهور نماذج التعلم العميق التوليدية، وعلى رأسها الشبكات التوليدية الخصومية (GANs) التي قدّمها Goodfellow وزملاؤه عام 2014 ، والتي مثّلت نقلة نوعية في القدرة على توليد بيانات اصطناعية عالية الواقعية¹.

الفرق بين البيانات الحقيقية والبيانات الاصطناعية والبيانات المعززة

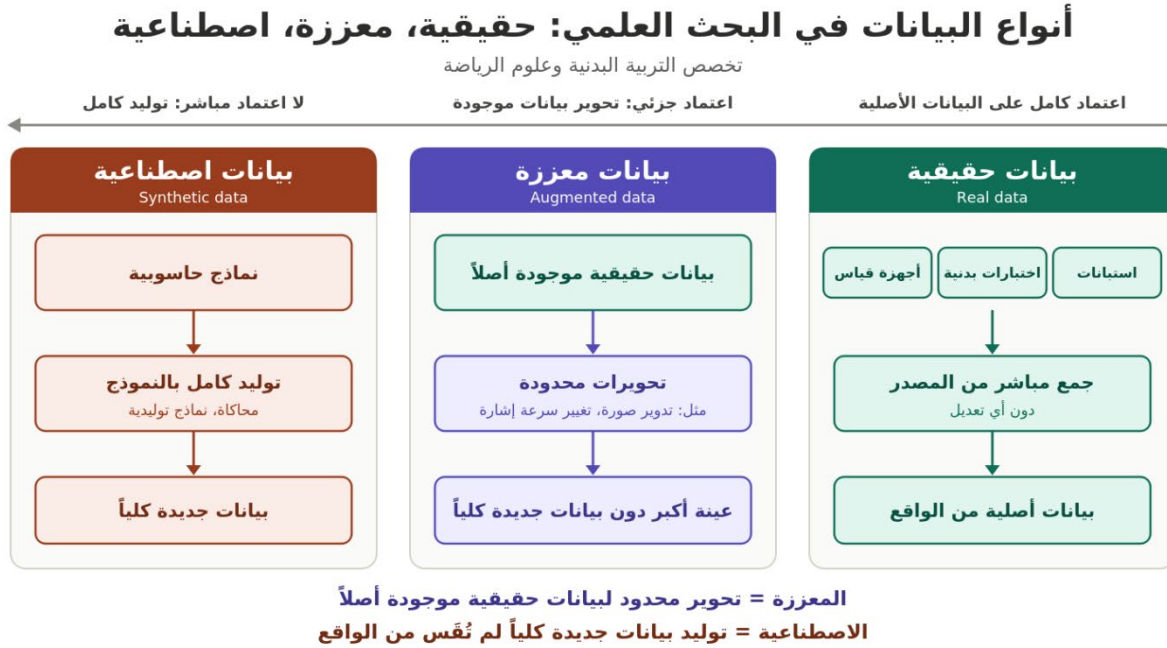
تُجمع البيانات الحقيقية مباشرة من مصادرها الأصلية كالاستبانات والاختبارات البدنية والأجهزة القياسية، بينما تُنتج البيانات الاصطناعية بالكامل من نماذج حاسوبية. أما البيانات المعززة (Augmented Data) فهي حالة وسيطة تقوم على إجراء تحويلات محدودة على بيانات

حقيقية موجودة أصلاً كتدوير صورة أو تغيير سرعة إشارة حركية (لزيادة حجم العينة دون توليد بيانات جديدة كلياً كما هو الحال في البيانات الاصطناعية²).

مثال رياضي: باحث قاس زمن ركض 30 متراً لأربعين لاعباً بساعة التوقيت، فهذه بيانات حقيقية. وإذا صوّر أحد اللاعبين بالفيديو ثم دوّر الصورة أو سرّعها قليلاً ليحصل على عدة نسخ من الفيديو نفسه، فهذه بيانات معززة، لأنها ما زالت مبنية على لاعب حقيقي. أما إذا درّب الحاسوب على أزمنة اللاعبين الأربعين ثم وُلدَ أزمنة 400 لاعب جديد لا وجود لهم، فهذه بيانات اصطناعية.

¹Goodfellow, I., Pouget-Abadie, J., et al.: Generative Adversarial Networks, Advances in Neural Information Processing Systems (NeurIPS), Vol. 27, 2014, p.2672

² Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(60)



الشكل (1) يوضح الفرق بين البيانات الاصطناعية والحقيقية والمعززة

أنواع البيانات الاصطناعية¹

تصنف البيانات الاصطناعية من حيث درجة اعتمادها على البيانات الحقيقية الى:

1. البيانات الاصطناعية بالكامل :

في هذا النوع يتم توليد البيانات بالكامل بواسطة خوارزميات دون الاعتماد المباشر على سجلات حقيقية يمكن تتبعها إلى أفراد أو كيانات فعلية. فنقوم النماذج أولاً بتعلم الخصائص الإحصائية العامة للبيانات الأصلية مثل التوزيعات، والعلاقات بين المتغيرات، والأنماط السلوكية ثم تُنشئ بيانات جديدة تحاكي هذه الخصائص دون أن تحتوي على أي سجل مطابق للواقع. ويُستخدم هذا النوع عندما تكون الخصوصية أولوية قصوى، مثل البيانات الطبية أو المصرفية، أو عند الرغبة في مشاركة البيانات لأغراض البحث أو التطوير

¹ خالد محمد الفرّج: معهد محترفي الإدارة IMB؛ الإمارات، الشارقة، محاضرة نشرت في المجلة، 2062، ص32

(Re-identification) دون مخاطر قانونية. ميزته الكبرى تكمن في تقليل احتمالية إعادة تحديد الهوية ، لكنه يتطلب نمذجة دقيقة لضمان الحفاظ على العلاقات الإحصائية المهمة (Risk)

2. البيانات الاصطناعية جزئياً:

في هذا المستوى، يتم الاحتفاظ بجزء من البيانات الحقيقية مثل البنية العامة أو بعض المتغيرات غير الحساسة، بينما يتم استبدال المتغيرات الحساسة أو القيم الحرجة ببيانات مولدة اصطناعياً ويُستخدم هذا النهج عندما تحتاج المؤسسة إلى الحفاظ على واقعية عالية للنموذج مع تقليل مخاطر الخصوصية، فمثلاً، يمكن الاحتفاظ بمتغيرات التوقيت أو التصنيفات العامة، مع توليد القيم المالية أو المعلومات الشخصية بصورة اصطناعية. ويوازن هذا النوع بين الدقة التحليلية والحماية، لكنه يتطلب حوكمة دقيقة لتحديد ما يُستبدل وما يُحتفظ به

3. البيانات الهجينة:

تمثل البيانات الهجينة مزيجاً مدروساً بين بيانات حقيقية وبيانات اصطناعية ضمن نفس مجموعة البيانات، بحيث يتم تعزيز البيانات الأصلية بسجلات مولدة لتحسين التوازن أو تمثيل الحالات النادرة. ويُستخدم هذا النوع كثيراً في حالات مثل كشف الاحتيال، حيث تكون الحالات الاحتمالية الحقيقية قليلة، فيتم توليد حالات ميزته أنه يحافظ على الأساس إضافية لتعزيز تدريب النموذج. كما يُستخدم في معالجة اختلال التوزيع الواقعي للبيانات، مع تحسين جودة التدريب، لكنه يتطلب مراقبة دقيقة لتجنب إدخال تحيزات غير مقصودة

تنوع البيانات الاصطناعية حسب الاستخدام¹

1. بيانات الجدولية الاصطناعية:

وتُعد الأكثر شيوعاً في بيئات الأعمال، حيث يتم توليد جداول تحاكي بيانات العملاء، والمبيعات، أو نماذج الاحتمالات GANS والمعاملات المالية، وسجلات العمليات، وتستخدم خوارزميات متقدمة مثل الإحصائية لإعادة إنتاج التوزيعات والعلاقات بين المتغيرات. ويفيد هذا النوع في تدريب نماذج التنبؤ، واختبار قواعد الأعمال، وبناء لوحات معلومات دون استخدام بيانات حساسة فعلية

2. بيانات الصور الاصطناعية:

وفيه يتم توليد صور رقمية تحاكي مشاهد حقيقية باستخدام تقنيات الرؤية الحاسوبية والمحاكاة ثلاثية الأبعاد، ويُستخدم هذا النوع بكثرة في تدريب نماذج التعرف على الصور، مثل اكتشاف العيوب الصناعية أو التعرف

¹ مصدر سبق ذكره

على الوجه. وفي سياق تحليل البيانات، تساعد صور المحاكاة على تحسين دقة النماذج عندما تكون الصور الواقعية محدودة أو مكلفة في الجمع

3. بيانات السلاسل الزمنية الاصطناعية:

يتم فيها توليد بيانات تحاكي تغيرات زمنية مثل المبيعات اليومية، أو استهلاك الطاقة، أو أسعار الأسهم، وتُستخدم لاختبار النماذج التنبؤية وتحليل الاتجاهات الموسمية، كما تساعد في محاكاة سيناريوهات مستقبلية لاختبار مرونة النماذج أمام تقلبات السوق

4. بيانات التفاعل السلوكي الاصطناعية

وهي تحاكي أنماط سلوك المستخدمين، مثل تصفح المواقع أو عمليات الشراء أو التفاعل مع التطبيقات، وتُستخدم في اختبار استراتيجيات التسويق، وتحليل تجربة المستخدم، وبناء نماذج توصية، خاصة عندما تكون بيانات السلوك الحقيقية محدودة أو محكومة بقيود خصوصية

الأسس والتقنيات العلمية لتوليد البيانات الاصطناعية¹

*المحاكاة الإحصائية والنمذجة الاحتمالية

وتقوم على افتراض توزيع احتمالي معين كالتوزيع الطبيعي يصف بيانات المتغير المراد محاكاته، ثم توليد قيم عشوائية من هذا التوزيع باستخدام أساليب كطريقة مونت كارلو، وتُعد من أبسط تقنيات توليد البيانات الاصطناعية وأقدمها، وتُستخدم غالباً في المتغيرات ذات التوزيع المعروف مسبقاً كقياسات اللياقة البدنية

*الشبكات التوليدية الخصومية (Generative Adversarial Networks – GANs)

¹ محمود عصمت العطيفي ؛ كتاب الأسس والتقنيات العلمية ، جامعة أسيوط ، مصر ، رقم الإيداع 7187، 2008

وتتكون من شبكتين عصبيتين تتنافسان في عملية تدريب واحدة؛ شبكة مولدة (Generator) تحاول إنتاج بيانات مزيفة تشبه البيانات الحقيقية، وشبكة مميّزة (Discriminator) تحاول التمييز بين البيانات الحقيقية والمولدة، وتستمر هذه المنافسة حتى تصل الشبكة المولدة إلى مستوى يصعب معه على الشبكة المميّزة التفريق بين النوعين، مما ينتج بيانات اصطناعية عالية الواقعية¹.

* نماذج الانتشار (Diffusion Models)

وتعمل عبر مرحلتين متعاكستين؛ مرحلة أولى يُضاف فيها تشويش عشوائي تدريجي إلى البيانات الحقيقية حتى تفقد بنيتها الأصلية بالكامل، ومرحلة ثانية يتعلم فيها النموذج كيفية عكس هذا التشويش تدريجياً لإعادة بناء بيانات جديدة، وقد أثبتت هذه النماذج تفوقاً ملحوظاً في توليد بيانات عالية الجودة مقارنة بالشبكات التوليدية الخصومية في بعض التطبيقات الحديثة².

.

* الشبكات العصبية المتغايرة الترميز (Variational Autoencoders – VAE)

وتقوم على ضغط البيانات الحقيقية إلى تمثيل مكثف في فراغ كامن (Latent Space)، ثم إعادة بنائها منه، مما يتيح توليد عينات جديدة من خلال أخذ نقاط عشوائية من هذا الفراغ الكامن وتحويلها إلى بيانات جديدة تحاكي خصائص البيانات الأصلية.

¹ Goodfellow, I., Pouget-Abadie, J., et al.: Generative Adversarial Networks, Advances

in Neural Information Processing Systems (NeurIPS), Vol. 27, 2014, p.2672

²مصدر سبق ذكره

علاقة البيانات الاصطناعية بالبحث الرياضي (Sports Analytics)¹

يعتمد البحث الرياضي الحديث بشكل كبير على نماذج التعلم الآلي والذكاء الاصطناعي لتوقع الإصابات، تحسين الأداء البدني، وتحليل التكتيكات. ومع ذلك، تواجه الأبحاث الرياضية عقبات رئيسية مثل: شح البيانات الخاصة بالرياضيين النخبة (بسبب الخصوصية والسرية التجارية للأندية)، وندرة وقوع إصابات خطيرة مما يجعل نماذج التنبؤ غير دقيقة.

مثال رياضي: لنفترض أن نادياً يملك بيانات 20 لاعباً فقط، أصيب منهم لاعبان بتمزق عضلي خلال الموسم. هذا العدد لا يكفي لتدريب نموذج يتنبأ بالإصابات، كما أن النادي لا يستطيع نشر بيانات لاعبيه لأنها سرية.

تأتي البيانات الاصطناعية لتقدم حلاً جذرياً لهذه المشكلات عبر:

- توليد بيانات سيناريوهات نادرة: مثل المواقف التكتيكية المعقدة في كرة القدم أو حركات البيوميكانيك الخطرة التي تؤدي للإصابات دون تعريض الرياضيين الحقيقيين للخطر.

مثال رياضي: في كرة السلة قد تحدث إصابة الركبة بعد هبوط اللاعب من قفزة مع انحناء الركبة إلى الداخل، وهي حالة نادرة التسجيل. يولد الباحث 1000 حالة هبوط اصطناعية بزوايا ركبة مختلفة، بعضها خطر، لتدريب النموذج دون أن يهبط أي لاعب حقيقي بطريقة تعرضه للإصابة.

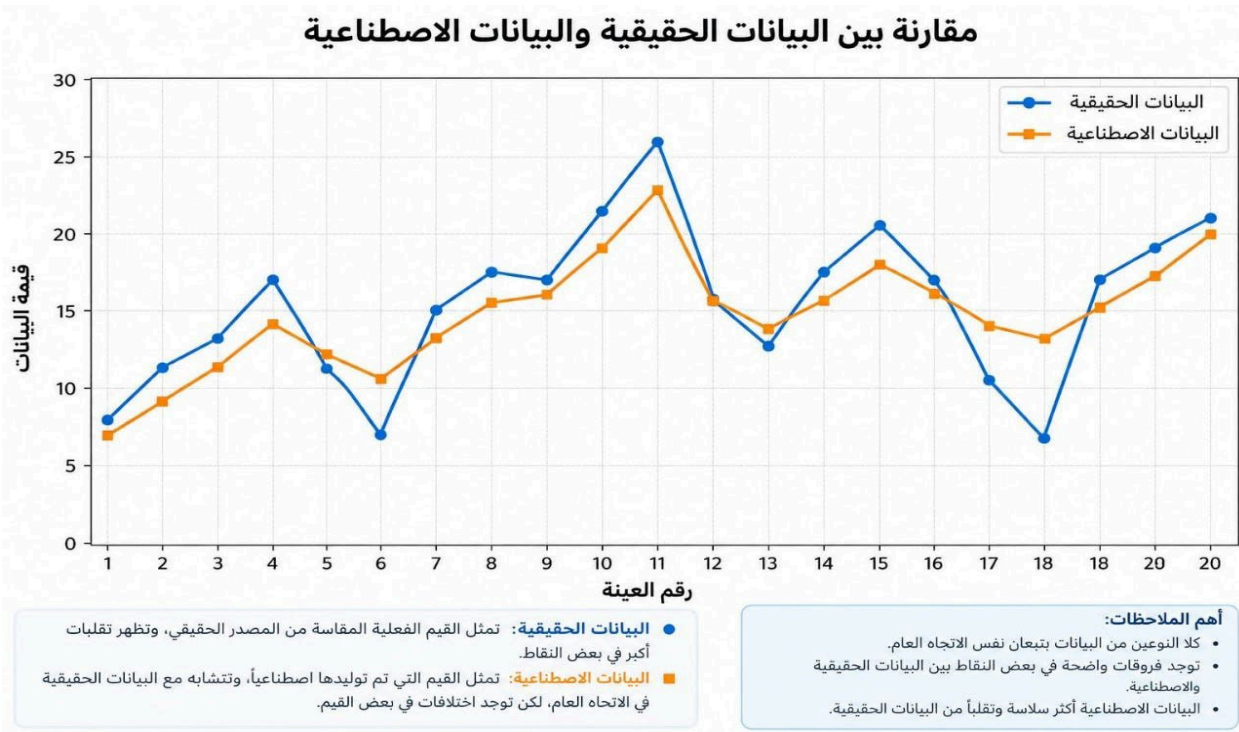
- زيادة حجم العينات البحثية (Data Augmentation): لتدريب نماذج الذكاء الاصطناعي الرياضية بدقة أكبر حتى مع توفر قواعد بيانات أصلية محدودة

مثال رياضي: باحث لديه اختبارات لياقة لـ 30 لاعباً ويريد تدريب نموذج يصنف اللاعب إلى ضعيف أو متوسط أو جيد. بما أن 30 حالة قليلة، يضيف 270 حالة اصطناعية لتصبح العينة 300، فيتدرب النموذج على أمثلة أكثر وتزداد دقته.

¹ Cordeiro, M. C., O Cathain, C., Daly, L., Kelly, D. T., & Rodrigues, T. B. (2025). A synthetic data-driven machine learning approach for athlete

- **مشاركة البيانات بأمان:** تمكين الباحثين والأكاديميين من مشاركة مجموعات البيانات الرياضية عالمياً دون خرق قوانين الخصوصية أو عقود الأندية

مثال رياضي: يرغب باحثان من جامعتين في تبادل بيانات لياقة لاعبي أحد الأندية، لكن النادي يمنع إعطاء أسماء لاعبيه وقياساتهم الأصلية. فيتبادلان ملفاً من بيانات اصطناعية تشبه الأصلية إحصائياً، ولا يتضمن أي لاعب حقيقي.



الشكل (2) يوضح علاقة البيانات الحقيقية والاصطناعية بالبحث الرياضي

مجالات استخدام البيانات الاصطناعية في البحث الرياضي والعلوم الرياضية

* **تحليل الأداء الحركي والبايوميكانيكي**

تُستخدم نماذج المحاكاة الحركية لتوليد بيانات اصطناعية تصف مسارات المفاصل وزوايا الحركة وقوى رد الفعل الأرضية في أداءات رياضية متعددة، مما يتيح للباحثين دراسة أثر تغيير متغير حركي معين كزاوية الانطلاق في الوثب الطويل على النتيجة النهائية للأداء دون الحاجة لتكرار التجربة ميدانياً مئات المرات¹.

مثال رياضي: لتحليل حركة رمي الرمح، بدلاً من تصوير 500 رمية حقيقية تُرهق اللاعب، نبني نموذجاً حاسوبياً يُولد زوايا مفصلي الكتف والمرفق وقوة الرمي في 500 رمية اصطناعية مختلفة، ثم يدرس الباحث أي زاوية تعطي أبعد مسافة.

***التنبؤ بالإصابات الرياضية وإدارة المخاطر** نظراً لصعوبة الحصول على بيانات كافية عن حالات الإصابة الحرجة أو النادرة لدى الرياضيين النخبويين، تُستخدم البيانات الاصطناعية لتوليد سيناريوهات إصابة متنوعة تُدرَّب عليها نماذج التعلم الآلي، بما يُحسِّن دقة أنظمة التنبؤ المبكر بمخاطر الإصابة قبل حدوثها فعلياً².

مثال رياضي: إصابة العضلة الخلفية للفخذ شائعة عند لاعبي كرة القدم، لكن حالاتها المسجلة قليلة في الفريق الواحد. يُولد الباحث 500 حالة اصطناعية تحاكي الحالات الحقيقية القليلة (حمل التدريب الأسبوعي وعدد السرعات العالية وعدد المباريات)، لتدريب نموذج يتنبأ المدرب إلى اللاعب الأكثر عرضة للإصابة.

***تحسين أنظمة الرؤية الحاسوبية لتتبع اللاعبين**

تحتاج نماذج الرؤية الحاسوبية المستخدمة في تتبع حركة اللاعبين وتحليل المباريات إلى كميات ضخمة من البيانات المصورة المُؤسَّمة، ولذلك تُستخدم بيانات المحاكاة الافتراضية لتوليد مقاطع فيديو اصطناعية لملاعب

¹أحمد فوزي السيد؛ توظيف تقنيات الذكاء الاصطناعي في تحليل الأداء الرياضي: رؤية تحليلية، مجلة علوم الرياضة، كلية

التربية الرياضية، جامعة حلوان، المجلد 15، العدد 2، 2023، ص.30

²زينب علي حسين؛ بناء نموذج إحصائي تنبؤي للإصابات الرياضية باستخدام تقنيات التعلم الآلي، رسالة ماجستير غير

منشورة، كلية التربية الرياضية للبنات، جامعة الزقازيق، 2021، ص.55

ولاعبين افتراضيين في أوضاع وإضاءة وزوايا تصوير متنوعة، مما يُسرّع من عملية تدريب هذه النماذج ويقلل كلفتها¹.

مثال رياضي: تحتاج الكاميرا الذكية إلى آلاف الصور لتتعلم تمييز لاعبي الفريقين في مباراة كرة قدم. بدلاً من تأشير الصور يدوياً، يولّد الحاسوب صوراً اصطناعية لملاعب فيه لاعبون بقمصان مختلفة وإضاءة مختلفة (نهار وليل) وزوايا تصوير متعددة، فتتعلم الكاميرا التمييز بدقة أعلى.

*محاكاة المباريات والسيناريوهات التكتيكية

يمكن توليد بيانات مباريات افتراضية تحاكي أنماط لعب فرق مختلفة، بما يتيح للمدربين والباحثين اختبار فاعلية خطط تكتيكية معينة إحصائياً قبل تطبيقها الفعلي في المنافسات الرسمية.

مثال رياضي: مدرب يريد معرفة ما يحدث لو لعب فريقه بطريقة 3-3-4 ضد فريق يضغط عالياً. يولّد الحاسوب 1000 مباراة افتراضية بين الأسلوبين، ويعرض متوسط عدد الأهداف وعدد التمريرات الناجحة لكل خطة، قبل تجربتها في مباراة حقيقية.

مزايا استخدام البيانات الاصطناعية في البحث الرياضي

- توسيع حجم العينة البحثية في الرياضات والفئات التي يصعب الوصول إليها ميدانياً بأعداد كبيرة².
- تمكين الباحثين من إجراء تجارب افتراضية متعددة السيناريوهات دون التعرض الفعلي لمخاطر الإصابة أو الإرهاق البدني للمشاركين.

¹ اسارة محمد الأنصاري؛ البيانات الضخمة وتطبيقاتها في الكشف عن المواهب الرياضية، مجلة التربية الرياضية للنبات،

جامعة الإسكندرية، المجلد 10 ، العدد 1، 2022، ص.40

² أحمد فوزي السيد؛ توظيف تقنيات الذكاء الاصطناعي في تحليل الأداء الرياضي: رؤية تحليلية، مجلة علوم الرياضة، كلية

التربية الرياضية، جامعة حلوان، المجلد 15 ، العدد 2، 2023، ص.136

- خفض الكلفة الزمنية والمالية لعمليات جمع البيانات الميدانية والمخبرية.
- تسهيل مشاركة البيانات بين الباحثين والمؤسسات دون انتهاك الاتفاقيات الخاصة بخصوصية الرياضيين¹.
- تحسين كفاءة نماذج الذكاء الاصطناعي المستخدمة في التحليل الرياضي من خلال تزويدها بكميات تدريب أكبر وأكثر تنوعاً

الاعتبارات الأخلاقية لاستخدام البيانات الاصطناعية في البحث الرياضي²

- الشفافية العلمية:** الإفصاح الصريح في منهجية البحث عن كون البيانات كلياً أو جزئياً اصطناعية، ونسبتها إلى إجمالي حجم العينة .
- "صون الخصوصية":** التأكد من عدم إمكانية استنتاج هوية أي رياضي حقيقي من البيانات الاصطناعية المنشورة، ولو بصورة غير مباشرة.
- "الحد من التحيز":** مراجعة النموذج التوليدي دورياً للتأكد من عدم تضخيمه لتحيزات موجودة أصلاً في البيانات الحقيقية كتحييز النوع أو العمر أو مستوى الأداء.
- "المسؤولية العلمية":** عدم استخدام البيانات الاصطناعية كبديل كامل عن التحقق الميداني في القرارات التي تمس صحة الرياضيين وسلامتهم مباشرة، كتشخيص إصابة فردية بعينها

خطوات التحقق من صحة البيانات الاصطناعية قبل اعتمادها بحثياً³

¹ محمد لطفي السيد؛ علم البيانات الرياضي: الأسس والتطبيقات، دار الفكر العربي، القاهرة، ط1، 2023، ص.90

² عبدالله ناصر المعولي؛ دراسة تحليلية عن الاعتبارات الأخلاقية وضوابط استخدام الذكاء في البحث العلمي، جامعة السلطان قابوس، سلطنة عمان، مجله الاكاديمية للأبحاث والنشر العلمي، 2026، رقم الإيداع 6495_2706

³ محمد نصر الدين؛ الإحصاء الاستدلالي التطبيقي في بحوث التربية الرياضية، دار الفكر العربي، القاهرة، ط2، 2020،

لا يجوز اعتماد أي مجموعة بيانات اصطناعية في بحث رياضي دون إخضاعها لإجراءات تحقق منهجي دقيقة، من أهمها:

المقارنة الإحصائية الوصفية: بين البيانات الاصطناعية والبيانات الحقيقية الأصلية المتوسطات، الانحرافات المعيارية، التوزيعات للتأكد من تطابقها بدرجة مقبولة

اختبار الأداء التنبؤي: بتدريب نموذج إحصائي أو نموذج تعلم آلي على البيانات الاصطناعية واختباره على بيانات حقيقية مستقلة، ومقارنة دقته بنموذج مماثل مدرب على بيانات حقيقية بالكامل.

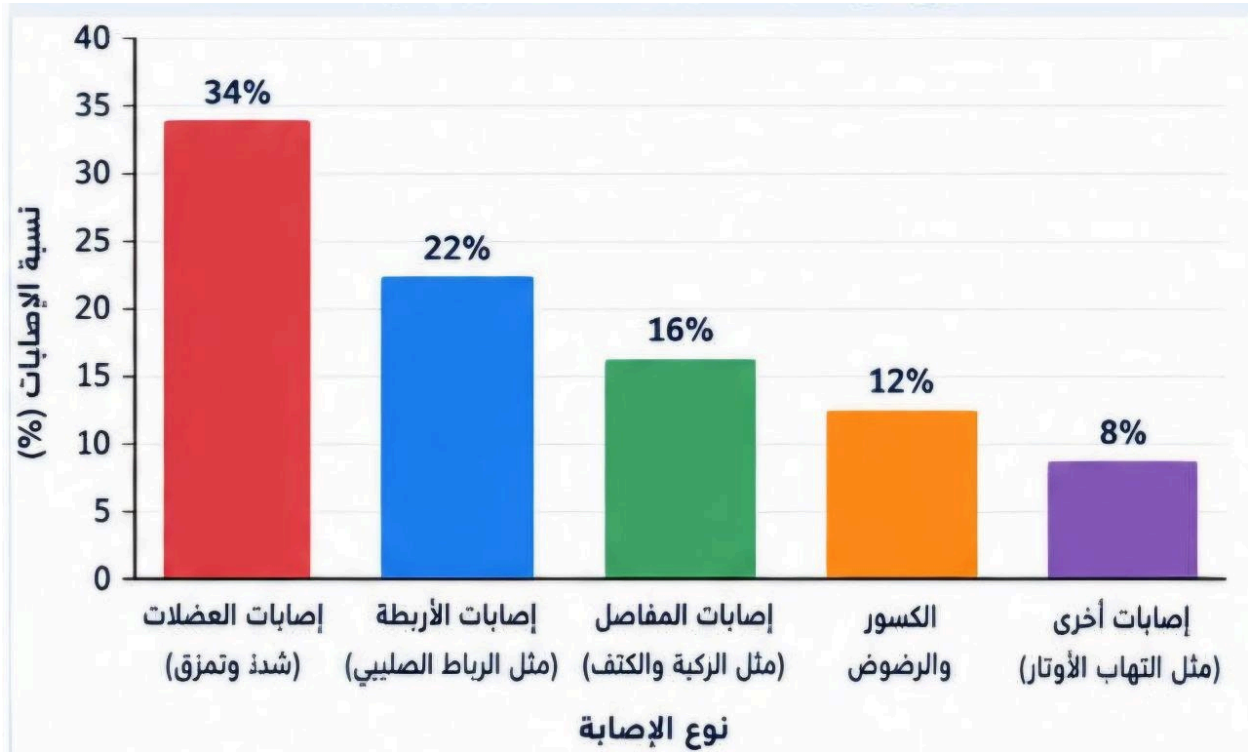
التحقق البصري للبيانات الحركية والمرئية تحديداً: من خلال عرض عينات من البيانات المؤدة على خبراء ميدانيين مدربين أو محكمين لتقييم مدى واقعيته الحركية أو الفنية.

مراجعة الخصوصية التفاضلية: بإجراء اختبارات هجومية محاكية كمحاولة إعادة تمييز الهوية للتأكد من عدم إمكانية تتبع أي سجل اصطناعي إلى فرد حقيقي بعينه

أبرز التطبيقات العملية للبيانات الاصطناعية في الرياضة¹

- **التنبؤ بالإصابات والوقاية منها:** استخدام بيانات إصابات اصطناعية لتدريب نماذج تتعرف على الأنماط المبكرة للإجهاد العضلي أو الإصابات الخطرة في الملاعب (مثل قطع الرباط الصليبي).
- **تحليل الأداء التكتيكي والفردية:** توليد بيانات تتبع (Tracking Data) وهمية لحركات اللاعبين والكرة (سواء عبر نظام GPS أو الكاميرات الذكية) لاختبار خطط لعب مختلفة في بيئات افتراضية قبل تطبيقها على أرض الواقع.
- **تحسين دقة أبحاث البايوميكانيك:** سد النقص في قياسات الحركة المعقدة للاعبين أثناء الأداء المهارى العالي.

¹ Warmenhoven, J., Impellizzeri, F. M., Shrier, I., Vigotsky, A. D., Lolli, L., Menaspà, P., Coutts, A. J., Fanchini, M., & Hooker, G. (2025). Synthetic data for sharing and exploration in high-performance sport.



الشكل (3) يوضح أنواع الإصابات الرياضية الأكثر شيوعاً

الاستنتاجات

- البيانات الاصطناعية هي بيانات يولدها الحاسوب بحيث تشبه البيانات الحقيقية في المتوسط والتوزيع والعلاقات بين المتغيرات، لكنها لا تعود لأي لاعب أو حدث حقيقي بعينه.
- أهم سبب لاستخدامها في التربية الرياضية هو صعوبة الحصول على عينات كافية من الرياضيين النخبة، وندرة الإصابات الخطيرة، وارتفاع كلفة الدراسات الطويلة.
- البيانات الاصطناعية تختلف عن البيانات الحقيقية التي تُجمع مباشرة من الميدان، وتختلف أيضاً عن البيانات المعززة التي تُعدّل فيها بيانات حقيقية موجودة أصلاً تعديلاً بسيطاً.

- للبيانات الاصطناعية ثلاثة أنواع: الكاملة وهي الأكثر أماناً للخصوصية، والجزئية التي تُستبدل فيها البيانات الحساسة فقط، والهجينة التي تجمع بين سجلات حقيقية وأخرى اصطناعية وتتاسب العينات الصغيرة.
- توجد عدة تقنيات لتوليدها، منها المحاكاة الإحصائية والشبكات التوليدية الخصومية ونماذج الانتشار والمحاكاة الفيزيائية، ولا توجد تقنية واحدة تصلح لكل الحالات. يعتمد الاختيار على نوع البيانات وحجمها وإمكانات الباحث.
- لا تصلح البيانات الاصطناعية للبحث العلمي إلا إذا حققت أربعة معايير: أن تشبه الحقيقية إحصائياً، وأن تكون متنوعة، وأن تحمي خصوصية الرياضيين، وأن تعطي النتائج التحليلية نفسها.
- تقيد الباحث الرياضي في توسيع حجم العينة، وإجراء تجارب افتراضية دون خطر على اللاعبين، وتقليل الوقت والكلفة، ومشاركة البيانات بأمان، وتحسين دقة نماذج الذكاء الاصطناعي.
- تُستخدم عملياً في تحليل الحركة (البايوميكانيك)، والتنبؤ بالإصابات والوقاية منها، وتتبع اللاعبين بالكاميرات، ومحاكاة المباريات والخطط التكتيكية.
- يجب على الباحث الالتزام بالأخلاقيات عند استخدامها: أن يفصح عن نسبة البيانات الاصطناعية في العينة، وأن يحمي هوية اللاعبين، وأن يتجنب تضخيم التحيز، وألا يعتمد عليها وحدها في القرارات التي تمس صحة اللاعب وسلامته.
- لا يجب التحقق من صحة البيانات الاصطناعية قبل اعتمادها، بمقارنتها إحصائياً مع الحقيقية، واختبار دقتها التنبؤية، وعرضها على خبراء ميدانيين، وفحص خصوصيتها.

التوصيات

وفي ضوء ما تقدم، يمكن استخلاص التوصيات الآتية:

تشجيع طلبة الدراسات العليا في التربية الرياضية على التعرف على أساسيات البيانات الاصطناعية وتقنيات توليدها ضمن مناهج البحث العلمي.

إلزام الباحثين بالإفصاح الصريح عن نسبة البيانات الاصطناعية المستخدمة في أي دراسة منشورة.

تشجيع التعاون البيئي بين كليات التربية الرياضية وأقسام علوم الحاسوب لتطوير نماذج توليد بيانات متخصصة بالمجال الرياضي.

إجراء مزيد من الدراسات العربية التطبيقية التي تختبر مدى واقعية البيانات الاصطناعية المولدة مقارنة بالبيانات الحقيقية في السياق الرياضي المحلي.

المراجع

- 1-فارس ألبياتي؛ الذكاء الاصطناعي في البحث العلمي : ط1، بغداد، 15، 2024
- 2-معهد محترف في الإدارة؛ كيان متخصص في تصميم وتنفيذ برامج تحليل البيانات المندمجة بالذكاء التنافسي ، مقاله نشرت يوليو 2026
- 3-فايد عبد الرزاق ؛ مجله روافد لدراسات والأبحاث في علوم الرياضة ، 2020.ص60
- 4-خالد محمد الفرج: معهد محترف في الإدارةIMB؛ الامارات ، الشارقة ، محاضرة نشرت في المجلة ، 2062 ص32،
- 5-محمود عصمت ألعطيفي ؛ كتاب الأسس والتقنيات العلمية ، جامعة أسيوط ، مصر ، رقم الإيداع 7187، 2008
- 6-أحمد فوزي السيد؛ توظيف تقنيات الذكاء الاصطناعي في تحليل الأداء الرياضي: رؤية تحليلية، مجلة علوم الرياضة، كلية التربية الرياضية، جامعة حلوان، المجلد15 ، العدد2، 2023، ص.30
- 7-زينب علي حسين؛ بناء نموذج إحصائي تنبؤي للإصابات الرياضية باستخدام تقنيات التعلم الآلي، رسالة ماجستير غير منشورة، كلية التربية الرياضية للبنات، جامعة الزقازيق، 2021، ص.55
- 8-عبدالله ناصر ألمعولي ؛ دراسة تحليلية عن الاعتبارات الأخلاقية وضوابط استخدام الذكاء في البحث العلمي، جامعة السلطان قابوس ، سلطنة عمان ، مجله الاكاديمية للأبحاث والنشر العلمي، 2026، رقم الإيداع 6495_2706
- 9-محمد نصر الدين؛ الإحصاء الاستدلالي التطبيقي في بحوث التربية الرياضية، دار الفكر العربي، القاهرة، ط2، 2020، ص.142

المصادر الأجنبية

- 1-¹Goodfellow, I., Pouget-Abadie, J., et al.: Generative Adversarial Networks, Advances in Neural Information Processing Systems (NeurIPS), Vol. 27, 2014, p.2672
- 2-¹ Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(60)
- 3-¹Goodfellow, I., Pouget-Abadie, J., et al.: Generative Adversarial Networks, Advances in Neural Information Processing Systems (NeurIPS), Vol. 27, 2014, p.2672
- 4-¹ Cordeiro, M. C., O Cathain, C., Daly, L., Kelly, D. T., & Rodrigues, T. B. (2025). A synthetic data-driven machine learning approach for athlete
- 5-¹ Warmenhoven, J., Impellizzeri, F. M., Shrier, I., Vigotsky, A. D., Lolli, L., Menaspà, P., Coutts, A. J., Fanchini, M., & Hooker, G. (2025). Synthetic data for sharing and exploration in high-performance sport.