

تطبيق التعلم الجماعي في صنع القرار الطبي

طالب الدكتوراه م. نهله سعيد سعد الدين

كلية الهندسة المعلوماتية - جامعة البعث

إشراف الدكتورة يسر السيد سليمان الأتاسي

الملخص:

لوحظ أنه مع توفر المزيد من البيانات، فإن الزيادة في استخدام خوارزميات التعلم الآلي ستسمح للمؤسسات بالنظر في نطاق أوسع بكثير من مجموعات البيانات أو المدخلات مما كان ممكناً في السابق. وهذا ما يوفر فرصة لاتخاذ قرارات أفضل - من خلال الجمع بين ذكاء الإنسان والآلة بطريقة ذكية.

تجدر الإشارة إلى أن مدى الاستخدام المستقبلي للخوارزميات في صنع القرار سيختلف حسب مجال تطبيقها، إلا أن المشكلة الأساسية هي الافتراض أن الخوارزميات قريبة من الكمال، أو ستحل محل البشر تماماً في جميع عمليات صنع القرار.

تزداد أهمية الخوارزميات في اتخاذ القرار، وقد تعزز أيضاً القرار البشري وتدعمه.

نعرض في هذه المقالة دراسة حول تطبيق صنع القرار باستخدام التعلم الجماعي (Ensemble Learning)، كما نحاول تسليط الضوء على صنع القرار في المجال الطبي بهدف مساعدة الأطباء عموماً في صنع القرار المناسب للحالة المدروسة.

الكلمات المفتاحية: التعلم الجماعي - خوارزميات - صنع قرار - تعلم الآلة - شبكات

InceptionV3- ResNet - تصنيف الصور.

Apply Ensemble Learning in Medical Decision-Making

Abstract:

It has been observed that as more data becomes available, the increase in the use of machine learning algorithms will allow organizations to consider a much wider range of data sets or inputs than was previously possible. This provides an opportunity to make better decisions – by combining human and machine intelligence in an intelligent way.

It should be noted that the extent of future use of algorithms in decision-making will vary according to their field of application, but the main problem is the assumption that algorithms are close to perfect, or will completely replace humans in all decision-making processes.

Algorithms are becoming increasingly important in decision-making, and they may also enhance and support human decision-making.

In this article, we present a study on the application of decision-making using group learning, and we try to shed light on decision-making in the medical field with the aim of assisting physicians in general in making the appropriate decision for the case studied.

Keywords: Ensemble Learning – Algorithms – Decision Making – Machine Learning – ResNet Networks – InceptionV3 – Image Classification.

1 - مقدمة:

صنع القرارات وظيفية ضرورية للغاية لأي شخص ولأي مجموعة ولأي مجتمع بشكل عام. فبدون القرارات، لا يحدث شيء فعلياً، كما تتحكم القرارات في مصائرنا كأفراد مستقلين أو كأفراد داخل مجموعات (مؤسسة أو شركة أو المجتمع ككل).

نحن نتخذ قرارات كل يوم. هذا جزء من عملنا وحياتنا، وفي معظم الأوقات تسيير الأمور في نصابها الصحيح مع تلك القرارات. فإذا كنا بحاجة إلى اتخاذ قرارات كبيرة وطويلة الأمد سيكون لها تأثير على ما نفعله في المستقبل، فسيكون من الجيد أن يكون لدينا نوع من نموذج صنع القرار الذي يساعدنا ويوفر أساساً ثابتاً للقرار الذي يجب أن نتخذه.

القرارات مهمة، لكن نلاحظ أنه نادراً ما يتم الاهتمام بكيفية صنع القرار. فبدلاً من اختيار نموذج اتخاذ القرار الصحيح للموقف المطروح، إما أن نتراجع عن القواعد أو نقرر ببساطة بأنفسنا بمعزل عن الآخرين. إلا رد الفعل السريع يمكن أن يكون له عواقب سلبية اقتصادية وثقافية وشخصياً وصحياً داخل المجتمع أو المؤسسات.

قرارات قرارات، هذا هو واقع الحياة، هل أنا أتخذ القرار الصحيح؟، هل هذا هو الخيار الأفضل؟

توفر خوارزميات لكل أنواع المشكلات التي من الممكن أن تحدث في العالم يتيح لنا القول إن هناك خوارزمية لذلك (لصنع القرار) أيضاً. وفعلياً عندما نتخذ أي قرار في حياتنا، فإننا نريد أفضل نتيجة بأقل قدر من الأسف.

كما نعلم فإن الخوارزميات تستخدم منذ سنوات كثيرة. إلا أن التوافر المتزايد للبيانات، وإمكانية الوصول إلى أجهزة الكمبيوتر ذات قوة المعالجة الكبيرة، أثر في تطوير خوارزميات جديدة للمساعدة في التنبؤ بالسلوك وأتمتة القرارات. وأدى ذلك إلى تمكين اتخاذ قرارات أكثر كفاءة ومستنيرة في الحكومات والشركات والهيئات والمؤسسات عموماً.

عند اتخاذ قرار ما من الممكن أن يكون لدينا خيارات متعددة للاختيار بينها، فأياً منها يجب أن نختاره حتى يكون لدينا أكبر فرصة لاختيار الخيار الأفضل؟

نواجه هذه الحالة في عدة حالات، مثلاً في عملية التوظيف لاختيار المرشح المناسب والأفضل بين المرشحين المتقدمين، أو اختيار منزل بين عدة خيارات متاحة أو.... الخ،

هذه العملية مقسمة إلى مرحلتين، مرحلة الاستكشاف ومرحلة الاختيار. أولاً يجب استكشاف عدد محدد من الخيارات لإنشاء خط الأساس الخاص بنا الذي سنعتمده كمعيار تقاضل بين الخيارات، ثم اختيار الخيار الأول الذي يتجاوز خط الأساس الذي تم تحديده. كما أنه في السنوات الأخيرة، تركزت الدراسات حول فرص استخدام خوارزميات التعلم الآلي في صنع القرار. وكما نعلم التعلم الآلي هو نوع من الذكاء الصناعي الذي يزود أجهزة الحاسوب بالقدرة على الوصول للنتائج تلقائياً عند تلقيها بيانات جديدة، دون أن تكون مبرمجة بشكل صريح على تلك البيانات. وهناك العديد من الأمثلة على الاستخدام الحالي والمستقبلي لخوارزميات التعلم الآلي في صنع القرار، نذكر في هذا السياق:

- في المجال المالي: يتم استخدام التعلم الآلي لأتمتة قرارات التداول واكتشاف فرص الاستثمار للعملاء بشكل سريع.
- في المجال التعليمي: يتم استخدام التعلم الآلي لتقييم أداء التدريس في المدارس. كما يمكن استخدام التعلم الآلي لدعم كفاءة التعلم عن طريق اختيار تقييمات وموارد تعلم أخرى للطلاب.
- في المجال الطبي: اكتسب التشخيص المعتمد على التعلم الآلي اهتماماً في السنوات الأخيرة، ويرجع ذلك إلى نتائج استدلالها الأسرع وقدرتها على أداء المهام المعقدة، والتي تتطلب خبرة وتجربة متخصصة. يتم استخدام التعلم الآلي لتحسين دقة تشخيص الأمراض المختلفة من خلال اكتشاف الأنماط، كما يمكن استخدام التعلم الآلي للتنبؤ بالاستجابة لمسارات علاج معينة، ما يتيح اتخاذ قرارات أكثر فائدة حول خيارات العلاج المخصصة (وهو المجال الذي تم العمل عليه في هذه الدراسة).

2- هدف البحث:

نحاول من خلال الدراسة دعم المجموعات في صنع القرار من خلال تطبيق التعلم الجماعي في عملية صنع القرار الطبي (التشخيص كما تم ذكره في القسم السابق) أصبح الذكاء الصناعي منتشراً في معظم جوانب صنع القرار في العالم. كما يتزايد استخدام الخوارزميات بسرعة بسبب توفر كميات كبيرة من البيانات وتحليلها من قبل

الحكومات والشركات والمؤسسات عموماً للاستفادة منها في عمليات صنع القرار العام. نقدم في هذه الدراسة تطبيقاً لتجميع التعلم لعدد من خوارزميات تعلم الآلة المستخدمة حديثاً في جانب صنع القرار.

يمكن لعملية صنع القرار باستخدام الخوارزميات القضاء على التحيز أو التمييز في عملية صنع القرار، وذلك يعتمد على جودة وتمثيل البيانات الأساسية التي تستخدمها الخوارزمية، بحيث تتأثر جودة البيانات بعدد من العوامل التي تشمل دقة التوقيت والتفصيل وجودة البيانات الوصفية ... الخ.

نحاول في هذه الدراسة أيضاً توضيح ما إذا كان نظام الذكاء الصناعي يمكن أن يعمل كصانع قرار أساسي، أو كمساعد وداعم مهم لصانع القرار البشري، بحيث إن الجمع بين أساليب الذكاء الصناعي في دعم القرار يمكن من استخدام كميات كبيرة من البيانات لإنتاج معلومات صنع القرار ذات الصلة بالمجال المدروس. كما نعمل في هذا البحث على توضيح كيف يمكن دمج منهجيات الذكاء الصناعي المختلفة في أنظمة صنع القرار، بحيث يتم تقديم إطار عمل لوصف تطبيقات الذكاء الصناعي في أنظمة صنع ودعم القرار لاحقاً.

أحد تطبيقات التعلم الآلي الأكثر شهرة هو في مجال صنع القرار، بحيث يتم تنفيذ أنظمة صنع القرار الحديثة على مجموعات بيانات ضخمة، وتعمل بشكل خاص على التعلم العميق وهو شكل من أشكال التعلم الآلي، ولكن ليس في قالب الأنظمة الخبيرة التقليدية التي تتم برمجتها يدوياً لإيجاد حلول لمشاكل محددة.

في نظم دعم القرار الأكثر نموذجية، يمكن القول إن برامج التشخيص والتنبؤ بالحالة تقترب من المعيار الأفضل من المعيار البشري، بحيث توجد في الوقت الحالي أنظمة ذكاء صناعي يمكنها اكتشاف مرض الزهايمر المبكر بدقة تزيد عن 80% حتى عقد من الزمن قبل ظهور الأعراض لأول مرة، وهو إنجاز يتفوق بشكل كبير على أفضل أخصائي علم الأمراض البشري الذي يحاول أي شيء مشابه Amoroso et al. (2017).

تقول الدراسة [1] أنه لا ينبغي استخدام أدوات القرار الخوارزمية في المسائل الكبيرة أو القرارات الحرجة فيما يتعلق بالسلامة ما لم تكن الأنظمة المعنية أفضل من الإنسان في المجال ذي الصلة أو المجال الفرعي لصنع القرار. أوصت الدراسة بثلاث استراتيجيات لمعالجة مشكلة التحكم، وأكثرها واعدة تتضمن تعاوناً بين أدوات خوارزمية عالية الكفاءة وأشخاص بشريين. كما حددت الدراسة ستة مبادئ رئيسية يجب أن تعكسها جميع أنظمة الإنسان والآلة في تصميمها.

3- عملية صنع القرار:

يشير مصطلح صنع القرار إلى اتخاذ الخيارات بين مسارات العمل البديلة - والتي قد تشمل أيضاً عدم اتخاذ أي إجراء. يستخدم الأفراد في جميع المؤسسات المعلومات التي يجمعونها لاتخاذ مجموعة واسعة من القرارات، وقد تؤثر هذه القرارات على حياة الآخرين وتغير مسار العمل. مثلاً، أدت القرارات التي اتخذها التنفيذيون والشركات الاستشارية لشركة Enron في النهاية إلى خسارة 60 مليار دولار للمستثمرين، وآلاف الموظفين بدون وظائف، وخسارة جميع صناديق التقاعد للموظفين.

أنواع القرارات:

قرارات مبرمجة: هناك قرارات بسيطة واعتيادية مثل ماذا ترتدي، وماذا تأكل، والطريق الذي تسلكه وأنت تذهب من وإلى المنزل. ربما لا تقضي الكثير من الوقت في هذه القرارات ويسمى هذا النوع من القرارات المباشرة بالقرارات المبرمجة، وهي قرارات تحدث بشكل متكرر بما يكفي لتطوير استجابة آلية لها. تسمى الاستجابة الآلية التي نستخدمها لاتخاذ هذه القرارات بقاعدة القرار.

قرارات غير مبرمجة: تتطلب القرارات الفريدة والمهمة التفكير الواعي وجمع المعلومات والنظر بعناية في البدائل، وتسمى هذه القرارات قرارات غير مبرمجة.

يعد اتخاذ القرارات الاستراتيجية والتكتيكية والتشغيلية جزءاً من وظيفة التخطيط في نموذج P-O-L-C (التخطيط / Planning - التنظيم / Organizing - القيادة / Leading - التحكم / Controlling).

نتخذ مئات القرارات كل يوم. قد يبدو اتخاذ القرار وكأنه مهمة بسيطة، إلا أن هناك عوامل رئيسية نحتاج إلى أخذها في عين الاعتبار للتأكد من أن القرار النهائي يؤدي إلى النتيجة التي نريدها.

عملية صنع القرار بالشكل الأبسط هي:

- الاعتراف بضرورة اتخاذ القرار: فهم لماذا نتخذ القرار، هل نفهم فعلياً المشكلة المطروحة أم هناك حاجة للحصول على مزيد من المعلومات؟

- تحديد جميع البدائل المختلفة المتاحة: محاولة العثور على مجموعة من الحلول للمشكلة حتى يتم مراجعتها واختيار الأفضل منها. وفقاً للطبيعة البشرية، فإنه يتم القفز إلى الحل الأول الذي نفكر فيه فوراً، ولكن هذا ليس دائماً المسار الأكثر إنتاجية للعمل.

- التعرف على البدائل وتقييم النتائج واختيار أفضل خيار: مراجعة كل بديل تم اكتشافه وتقييم ما إذا كان تنفيذ هذه الأفكار سيكون له تأثير إيجابي أو سلبي على الأهداف النهائية.

- تنفيذ القرار وتقييم النتيجة: نتأكد من تقييم النتيجة للتأكد من أن القرار الذي تم اتخاذه كان الاختيار الصحيح. لا تتردد في إجراء تغييرات إذا رأيت مجالاً للتحسين.

القرارات التي نتخذها لا تتطوي دائماً على مشاكل. فقد تأتي بعض مواقف اتخاذ القرار كنتيجة للفرصة، حيث تدرك أن لديك فرصة لإجراء تغيير ويجب عليك تحديد أفضل طريقة للاستفادة منها.

4- التعلم الجماعي:

ظهر مؤخراً التعلم الجماعي أو التعلم القائم على المجموعات أو تعلم أنظمة التصنيف المتعددة وبتعريف هذا التعلم فهو مفهوم استخدام نماذج عصبية أو نماذج تدريب متعددة معاً لإنشاء خوارزمية أكثر دقة من الخوارزمية الأصلية التي تعمل بمفردها. ويعتبر أحد الحلول الملائمة لبناء نماذج تصنيف قوية ودقيقة [2]. فبدلاً من استخدام خوارزمية واحدة لبناء نموذج، يتم استخدام طرق المجموعات بشكل واسع لبناء مجموعة من المصنفات والجمع بينها، سواء كانت هذه المصنفات ضعيفة أو قوية، وتسمى هذه المجموعة

مجموعة المتعلمين الأساسيين. تتعلق النقاط الأساسية لفعالية طريقة التجميع بالاختيار الدقيق لكل من المتعلمين الأساسيين [3] وطريقة الدمج لإنتاج النتيجة النهائية. من الطرق الجماعية الشهيرة وشائعة الاستخدام، نذكر:

- حساب المتوسط (البسيط أو الموزون)
- التصويت (الأغلبية، الإجماع، التعددية، أو حتى الأصوات المرجحة)

في [4]، تم اقتراح مخطط تدريب مشترك قائم على مجموعة التصويت فيما يتعلق بمشاكل التصنيف الثنائي. الأساس في الخوارزمية المقترحة هو بناء متعلم مجموعة يختار من بين خمس مصنفات مختلفة، بناءً على وظيفة موضوعية جديدة تقيس فعالية أي زوج من المصنفات لكل مجموعة بيانات، تعمل بموجب نظام التصويت. يتم التحقق من فعالية الآلية المقترحة لاختيار المشاركين في المجموعة من خلال العديد من التجارب - من خلال مخطط التدريب الذاتي المعروف - ومخطط التدريب المشترك، حيث تم تطبيق الأزواج العشرة المختلفة من الخوارزميات في نفس متعلم التصويت والمصنفات الفردية الخمسة، على عدد كبير من مجموعات البيانات المعيارية عبر خمس قيم نسبية معنونة منفصلة. تم تأكيد النتائج التي تم الحصول عليها فيما يتعلق بمقياسين تصنيفين معروفين إحصائياً.

تم ملاحظة أن عدداً كبيراً من القرارات التي تم تصنيفها بشكل خاطئ في البداية بناءً على التنبؤات الفردية تم إرجاعها نحو التسمية الصحيحة. بينما في الوقت نفسه، لم تتأثر العديد من الحالات التي يختلف فيها المشتركان حول التسمية الثنائية. يحدث هذا لأن قيمة ثقة كبيرة وصحيحة مقترنة بقيمة أصغر غير صحيحة تظل كما هي في ظل نظام التصويت المقترح.

تم في النهاية اقتراح تطبيق اعتماد مراحل معالجة مسبقة أكثر تخصصاً موجهة نحو مشاكل أكثر تحديداً لتعزيز أداء الاستراتيجية المقترحة في هذه الورقة، ومع ذلك فقد سجلت الخوارزمية المقترحة أداءً قوياً ودقيقاً. كما أوصت الدراسة بإمكانية استخدام الشبكات العصبية العميقة (DNNs) [5] لتعزيز أداء التنبؤ، مع تغذية إما بالبيانات

الأولية أو أنواع عامة أخرى من مجموعات البيانات. فقد أثبتت شبكات الذاكرة طويلة المدى (LSTM) فعاليتها بدرجة كافية عند دمجها مع طرق SSL لبناء أنظمة دعم القرار الطبي [3].

في البحث [6] [7] ، تم اقتراح خوارزمية هجينة للتعلم الآلي تتكون من أشجار القرار والشبكات العصبية التي يمكنها تصنيف البيانات بكفاءة وفعالية كبيرة. كان هيكل الخوارزمية الهجينة معتمداً على شجرة قرار ثنائية بحيث كل عقدة في الشجرة عبارة عن شبكة عصبية مدربة على تصنيف فئة معينة من المخرجات باستخدام التصنيف الثنائي. البيانات التي تم استخدامها في هذه الدراسة لتدريب واختبار قدرة الخوارزمية كانت مجموعة بيانات لصيانة طائرة بوينج 737 التابعة لإدارة الطيران الفيدرالية (FAA) والتي تكونت من 137236 سجلاً مميزاً يتكون كل منها من 72 بارامتر. تم العمل بذلك عن طريق تصنيف التناقض أو سبب الحادث فيما إذا كان الحادث قد وقع أثناء عمليات الصيانة المجدولة أم لا ثم تصنيف التفاصيل المحددة المتعلقة بالحادث. أشارت النتائج إلى أن الخوارزمية الهجينة قادرة على تصنيف الحوادث بشكل فعال بدقة عالية. أشارت الدراسة أيضاً إلى قابلية تطبيق الخوارزمية في سيناريوهات العالم الحقيقي بحيث عرضت فوائد الجمع بين أشجار القرار والشبكات العصبية بدلاً من استخدامها بشكل فردي كونها أعطت تنبؤاً أكثر دقة عند مقارنتها باستخدام نموذج عصبي واحد فقط للتنبؤات.

كما تم اقتراح نموذج هجين أيضاً في الدراسة [8] بحيث دمج نموذج الانتروبي الأقصى، وآلة نقل الدعم (Support Vector Machine)، وناظرة بايز لتلخيص نص متعدد المستندات لتحسين دقة التصنيف.

اقترح الباحثون في الدراسة [9] أيضاً نموذجاً هجيناً للتعلم الآلي للمشكلات متعددة التصنيف. تتكون الطريقة من مصنف شجرة القرار C4.5 ونهج واحد مقابل الكل. تم عرض فعالية الطريقة الهجينة على مجموعات البيانات مفتوحة المصدر لتجزئة الصور والأمراض الجلدية والتصوير للمفاوي.

في الدراسة [10] جمعت شبكة عصبية مع خوارزمية جينية لزيادة أداء الشبكات العصبية وتم تطبيق الطريقة الهجينة على مجموعة بيانات أمراض القلب.

5- دراسة حول الاستخدام الحالي للخوارزميات والتعلم الآلي في صنع القرار:

في طريقة المجموعة، تم دمج الشبكة العصبية (MLP) مع AdaBoost و Bagging و Dagging و MultiBoost و Rotation Forest و Random SubSpace. كما تم تطوير نموذج تنبؤي هجين آخر بواسطة Chen et al [11]. وفي هذه الطريقة، تم دمج مجموعات K-mean مع شجرة قرار J48 لتشخيص مرض السكري من النوع 2.

كما استخدمت دراسة أخرى مبنية على التنبؤ أسلوب التعلم الجماعي. اقترح المؤلفون نهج التعلم الجماعي المكون من طبقتين (TLEL) بناءً على أشجار القرار [12]. تستخدم الطبقة الخارجية نماذج Random Forest مختلفة للتدريب بينما الطبقة الداخلية عبارة عن تكامل لشجرة القرار والتعبئة لبناء نموذج Random Forest. استخدمت دراسة أخرى [13] التعلم الجماعي من أجل أداء تنبؤي أفضل للحياة المتبقية (RUL) لمحركات الطائرات واستخدمت عدة طرق مثل المتعلمين الأساسيين المتعددين، بما في ذلك الغابات العشوائية (RFs) والشبكات العصبية التكرارية (RNN)، نظام الاستدلال الضبابي القائم على الشبكة التكيفية (ANFIS) كما استخدمت الطريقة أيضاً تحسين سرب الجسيمات (PSO) والتحسين التريبيعي المتسلسل (SQP) لتحقيق أفضل مزيج من الأوزان لاستخدامها في المتعلمين الأساسيين.

كما كان هناك امتداد للدراسة السابقة بواسطة Li et al [14]. واستخدم في الدراسة البيان غير الدوري الموجه (DAG) المهجن بذاكرة طويلة المدى (LSTM) وشبكة عصبية تلافيفية (CNN) للتنبؤ بـ RUL. قام باختبار الطريقة باستخدام مجموعة بيانات توفرها وكالة ناسا.

6- خوارزميات تعلم المجموعات شائعة الاستخدام

Bagging /التعبئة:

هي واحدة من أقدم الخوارزميات المبنية على المجموعات وأكثرها سهولة وربما أبسطها، مع أداء جيد بشكل مدهش (Breiman 1996) [15]. يتم الحصول على تنوع المصنفات في التعبئة باستخدام نسخ متماثلة لبيانات التدريب. بحيث يتم الحصول على مجموعات فرعية مختلفة من بيانات التدريب بشكل عشوائي (مع الاستبدال) من مجموعة بيانات التدريب بأكملها. يتم استخدام كل مجموعة فرعية من بيانات التدريب لتدريب مصنف مختلف من نفس النوع. يتم بعد ذلك دمج المصنفات الفردية عن طريق التصويت بالأغلبية البسيطة لقراراتهم. بالنسبة لأي حالة، فإن التصنيف الذي يختاره أكبر عدد من المصنفات هو قرار المجموعة.

:Boosting

يتم هنا إنشاء مجموعة من المصنفات من خلال إعادة تجميع البيانات، والتي يتم دمجها بعد ذلك عن طريق التصويت بالأغلبية. بحيث يتم توجيه عملية إعادة التشكيل بشكل استراتيجي لتوفير بيانات التدريب الأكثر إفادة لكل مصنف متتالي. كل تكرار للتعزيز يقوم بإنشاء ثلاث مصنفات ضعيفة: المصنف الأول C1 يتم تدريبه بمجموعة فرعية عشوائية من بيانات التدريب المتاحة. يتم اختيار المجموعة الفرعية لبيانات التدريب للمصنف الثاني C2 على أنها المجموعة الفرعية الأكثر فائدة، بالاعتماد على C1. بحيث يتم تدريب C2 على بيانات التدريب التي تم تصنيف نصفها فقط بشكل صحيح بواسطة C1، والنصف الآخر مصنف بشكل خاطئ. يتم تدريب المصنف الثالث C3 مع الحالات التي يختلف عليها C1 وC2. وفي النهاية يتم الجمع بين المصنفات الثلاثة من خلال تصويت الأغلبية الثلاثية.

:AdaBoost

يمكن القول إن AdaBoost (Adaptive Boosting) هي من أكثر الخوارزميات القائمة على المجموعات شهرةً ولها العديد من الأنواع، مثل AdaBoost.M1 لمشاكل التصنيف، AdaBoost.M2 للمصنفات الضعيفة، AdaBoost.R لمشاكل regression.

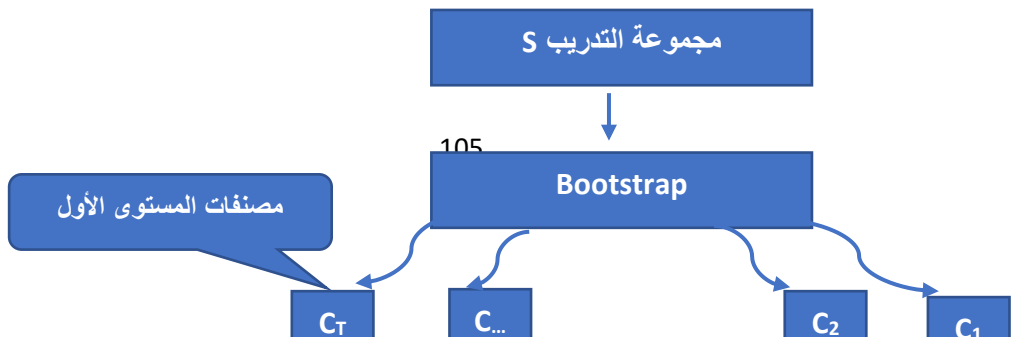
التعميم المكس / Stacked Generalization

في التعميم المكس λ Wolpert (أو التكديس)، يتم تدريب مجموعة من المصنفات أولاً باستخدام عينات من بيانات التدريب، وإنشاء مصنفات المستوى الأول، والتي تُستخدم مخرجاتها بعد ذلك لتدريب مصنفات الطبقة الثانية (Wolpert 1992) [16]. الفكرة الأساسية هي معرفة ما إذا تم تعلم بيانات التدريب بشكل صحيح. على سبيل المثال، إذا تعلم مصنف معين بشكل غير صحيح منطقة معينة من مساحة الميزة، وبالتالي أخطأ باستمرار في تصنيف الحالات القادمة من تلك المنطقة، فقد يكون مصنف المستوى الثاني قادراً على تعلم هذا السلوك بالتعاون مع السلوكيات المكتسبة للمصنفات الأخرى، وبالتالي يمكنه تصحيح مثل هذا التدريب غير المناسب. يتم تقسيم مجموعة بيانات التدريب بأكملها إلى T كتلة، ويتم تدريب كل مصنف من المستوى الأول بدايةً على مجموعة مختلفة من $T-1$ كتلة من بيانات التدريب. ثم يتم تقييم كل مصنف على الكتلة T (الاختبار الزائف)، ولا يتم رؤيته أثناء التدريب. تشكل مخرجات هذه المصنفات على كتل التدريب الزائفة مع الـ Labels الصحيحة الفعلية لتلك الكتل مجموعة بيانات التدريب لمصنف المستوى الثاني.

يمكننا تخيل إجراء التكديس على أنه يحتوي على مستويين: المستوى الأول والمستوى الثاني:

- المستوى الأول: بيانات هذا المستوى هي مدخلات مجموعة بيانات التدريب وتتعلم نماذج هذا المستوى القيام بتنبؤات من هذه البيانات.
- المستوى الثاني: بيانات المستوى الثاني تأخذ مخرجات نماذج المستوى الأول كمدخلات، ويقوم النموذج الوحيد في هذا المستوى (أو المتعلم الفوقي / meta-learner) بتعلم عمل تنبؤات من هذه البيانات.

نوضح في الشكل الآتي التعميم المكس:



الشكل (1) توضيح التعميم المكّس

مجموعة التعميم المكّس (Stacked Generalization Ensemble):

المجمّع هو نموذج يحسب المتوسط للتنبؤات من عدة نماذج مدربة، ويتمثل أحد قيود هذا النهج هو أن كل نموذج يساهم بنفس المقدار في تنبؤ المجموعة، بغض النظر عن مدى جودة أداء النموذج.

أحد أشكال هذا النهج يُطلق عليه اسم مجموعة المتوسط الموزون (Weighted Average Ensemble) يقوم بوزن مساهمة كل عضو في المجموعة من خلال الثقة أو الأداء المتوقع للنموذج على عينة بيانات محددة. وهذا يسمح للنماذج ذات الأداء الجيد بالمساهمة أكثر من النماذج ذات الأداء الأقل. يوفر نموذج مجموعة المتوسط الموزون تحسناً على نموذج مجموعة المتوسط العادي.

التحسين الآخر الذي يمكن تقديمه لهذا النهج هو استبدال نموذج المجموع الموزون الخطي (مثل الانحدار الخطي) المستخدم لدمج تنبؤات النماذج الفرعية بأي خوارزمية تعلم. وهذا النهج يسمى التعميم المكّس (التكديس).

في التكديس، تأخذ خوارزمية التعلم مخرجات النماذج الفرعية كمدخلات وتحاول تعلم أفضل طريقة تجمع بين تنبؤات الإدخال للحصول على تنبؤ خرج أفضل ما يمكن.

7- أهمية نموذج التعلم الجماعي:

يكون عادةً نموذج التعلم الجماعي أكثر تعقيداً من استخدام نموذج واحد للتعلم الآلي، وقد يتطلب هذا أسلوباً أكثر تعقيداً لإعداد النموذج، وموارد حسابية أكثر لتدريب النموذج، ومع ذلك، يمكن القول إنه لدينا سببان لاستخدام خوارزميات تعلم المجموعات، وهما:

متانة أفضل: تقدم العديد من خوارزميات التعلم الآلي تنبؤات مختلفة في كل مرة يتم فيها تدريب النموذج على نفس البيانات أو حتى على بيانات مختلفة قليلاً. يشار إلى هذا باسم التباين في التنبؤات أو استقرار النموذج. يمكن أن تعطي خوارزميات التعلم الجماعي تنبؤات أكثر استقراراً من نموذج واحد.

تنبؤات أفضل: هناك بعض المجالات ومهام التنبؤ يكون الأداء التنبؤي هو الأكثر أهمية من أي شيء آخر. وهذا هو الحال بشكل خاص في التعلم الآلي التنافسي حيث تكون المهارة التنبؤية هي النتيجة الوحيدة التي يتم أخذها في الاعتبار وهذا ما توفره خوارزميات التعلم الجماعي فهي توفر مهارة تنبؤية أفضل من استخدام نموذج واحد.

تستخدم المجموعات بشكل شائع لتثبيت التنبؤات التي قدمتها نماذج التعلم العميق بالنظر إلى كل من الطبيعة العشوائية لبنية النموذج التي تمت تهيئتها باستخدام أوزان عشوائية.

يشير تنوع المجموعة إلى الاختلافات في التنبؤات أو أخطاء التنبؤ التي تحدثها النماذج التي تساهم في المجموعة. ومن الناحية المثالية، سيكون كل عضو في المجموعة مستقلاً، ما يعني أنه سيقدم تنبؤات بطريقة مختلفة تماماً عن جميع أعضاء المجموعة الآخرين إلا أن هذا غير ممكن لأن لدينا مجموعة بيانات تدريب واحدة فقط.

تشير طرق التجميع إلى الطريقة المحددة التي يتم بها استخدام الإخراج من كل عضو في المجموعة لإجراء التنبؤ النهائي. يمكن أن يتضمن ذلك استخدام طريقة إحصائية بسيطة كوسيلة للتصويت، على الرغم من أنه قد يتضمن وزن التنبؤات بمهارة النموذج المساهم أو حتى استخدام نموذج آخر لمعرفة أفضل طريقة للجمع بين التنبؤات.

8- الدراسة العملية:

تتكون المهمة المختارة من تصنيف صور شعاعية طبية لسبعة أنواع مفاصل وعظام في الجسم إلى فئتين. يرجع اختيار هذه المهمة بشكل أساسي إلى توفر هذه المجموعة من البيانات، بالإضافة إلى مجال المشكلة، وهو عام بما يكفي ليكون مناسباً للتطبيق المطلوب.

حسب الاحصائيات الأخيرة المتوفرة في مجال الحالات المتعلقة بمجال بحثنا (صور شعاعية) فإن أمراض العضلات والعظام تؤثر على أكثر من 1.7 مليار شخص في جميع أنحاء العالم، وهي السبب الأكثر شيوعاً للألم والإعاقة الشديدة وطويلة الأجل، بحيث تزداد 30 مليون زيارة لقسم الطوارئ سنوياً وتتزايد بسبب هذا النوع من الحالات [17][18].

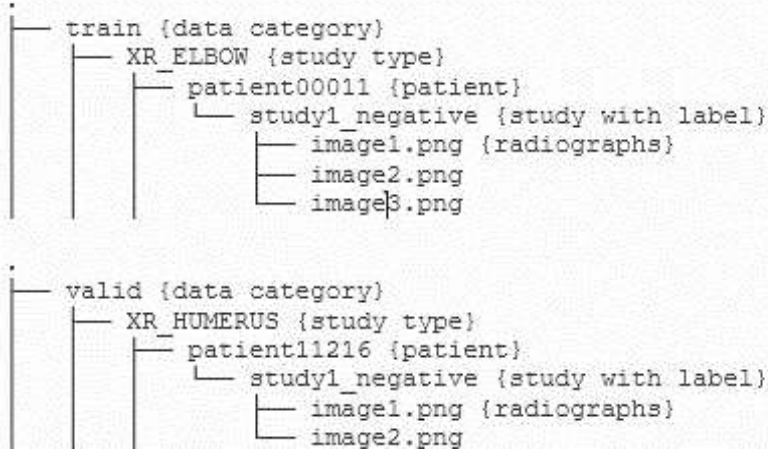
مهمة الكشف عن المرض، أو بمعنى آخر تحديد ما إذا كانت دراسة التصوير الشعاعي طبيعية أو غير طبيعية، هي عملية إشعاعية مهمة بحيث تعتبر دراسة تقسّر على أنها تستبعد المرض بشكل طبيعي ويمكن أن تلغي حاجة المرضى للخضوع لمزيد من الإجراءات التشخيصية أو التدخلات.

9- عينة البيانات المدروسة:

نعمل من خلال بحثنا على MURA Dataset، وهي مجموعة بيانات كبيرة للكشف عن التشوهات في التصوير الشعاعي العضلي الهيكلي تتألف من 14863 دراسة (وتحتوي كل دراسة على واحد أو أكثر من المشاهدات (الصور)) من 12173 مريضاً، بإجمالي 40,561 صورة إشعاعية متعددة الرؤية. تحتوي مجموعة البيانات على 9,045 دراسة طبيعية و 5818 دراسات إشعاعية عضلية هيكلية غير طبيعية [19].

ينتمي كل منها إلى واحد من سبعة أنواع من الدراسات التصويرية للأطراف العلوية: الكوع، الإصبع، الساعد، العضد، الكتف والمعصم واليد. تم تصنيف كل دراسة يدوياً على أنها طبيعية أو غير طبيعية من قبل أطباء الأشعة المعتمدين من مستشفى ستانفورد في وقت التفسير الشعاعي السريري في بيئة الأشعة التشخيصية بين 2001 و 2012.

أغلب الدراسات التي تمت على نفس عينة البيانات تقسم الصور إلى 36808 صورة تدريبية و3197 صورة للتحقق منها. بحيث تأتي عينة البيانات كمجموعة بيانات منقسمة إلى مجموعتين: مجموعة التدريب ومجموعة التحقق في مجلدات منفصلة. كل مجموعة من مجموعات البيانات (التدريب والتحقق) لها مجلدات فرعية.



الشكل (2): طريقة تقسيم عينة البيانات المدروسة

كما نرى من الشكل أنه يتم تقسيم البيانات ضمن المجموعتين المذكورتين سابقاً (في مجلد منفصل لكل مجموعة منها)، كما أنه يوجد مجلد خاص بكل فئة من الفئات السبعة والمذكورة أيضاً، ضمن كل فئة يوجد مجلد خاص لكل مريض (حالة دراسة مستقلة) ويتم تحديد حالة الدراسة مع وضع الـ label أو التصنيف لها، بحيث أن التصنيفين المحددين هما positive وnegative، وضمن المجلد الخاص بحالة الدراسة هناك مجموعة من المشاهدات (الصور).

تقسيم البيانات وزيادتها:

كما ذكرنا سابقاً عن البيانات المدروسة وتصنيفها إلى فئتين وفقاً للتسميات المقدمة من اختصاصي الأشعة. قمنا بتقسيم 80% بشكل عشوائي في مجموعة التدريب (Train Set) من مجموعة البيانات بالكامل، بينما بقي الجزء الباقي 20% في مجموعة الاختبار (Test Set).

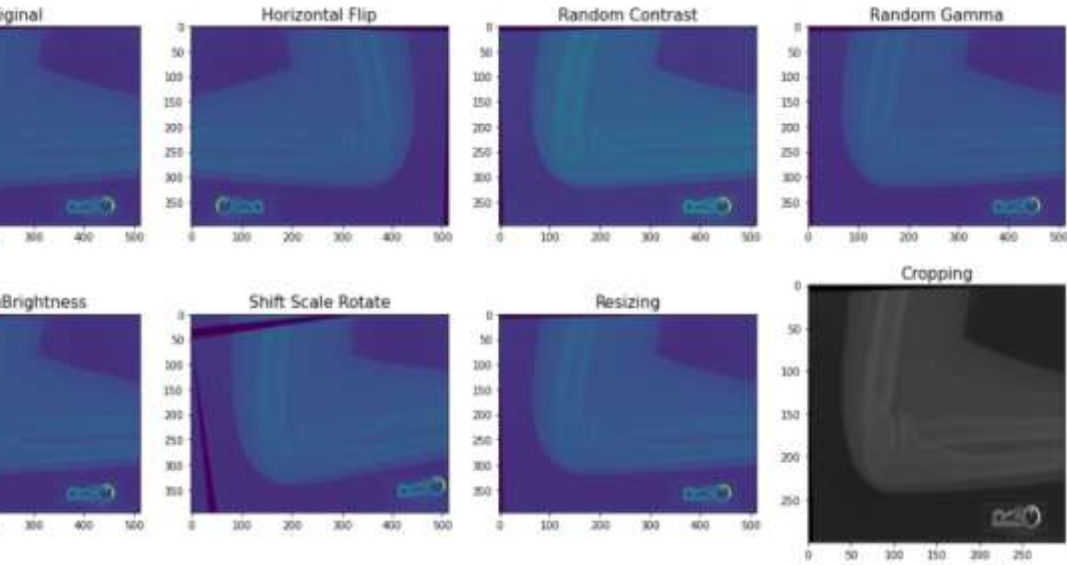
تم تقييم أداء المصنف (النموذج) وفقاً لمتوسط دقة التصنيف الناتج، كما تم الاعتماد على مقياس كوهين كبا كمقياس جديد في عملية التصنيف (هناك شرح له في قسم النتائج والتقييم) وذلك فترة تشغيل تبلغ 43 تكرار.

المعالجة المسبقة للبيانات:

قمنا بمعالجة كل صورة بحيث يكون لها نفس المتوسط الانحراف المعياري للصورة في مجموعة التدريب ImageNet ، تم استخدام صور متغيرة الحجم وتحويلها لحجم ثابت $224 \times 224 \times 3$. كما قمنا بزيادة البيانات أثناء التدريب عن طريق تطبيق بعض التحويلات باستخدام ImageDataGenerator, وما قمنا بتطبيقه فعلياً هو (Horizontal Flip, Random Contrast, Random Gamma, Random Brightness, Shift Scale Rotate, Resizing, Cropping)

كما اعتمدنا على Batch Size = 32

يوضح الشكل الآتي مثال على نتائج التحويلات المطبقة على إحدى الصور:



الشكل (3): مثال على نتيجة تطبيق تحويلات المعالجة المسبقة للبيانات على صورة من عينة البيانات (Data Augmentation)

مسابقة MURA

نشرت مجموعة التعلم الآلي في جامعة ستانفورد ورقة بحثية [19] تتعلق بمشكلة تصنيف الصور الشعاعية الطبية وقدمت واحدة من أكبر مجموعة بيانات صور شعاعية عامة في العالم تسمى MURA. MURA هي اختصار للتصوير الشعاعي العضلي الهيكلي (Musculoskeletal Radiographs). استخدمت مجموعة ML بجامعة ستانفورد خوارزمية DenseNet169 لتدريب شبكة عصبية عميقة يمكنها الكشف عن التشوهات في الصور الشعاعية بدقة أقرب إلى كبار أطباء الأشعة. وكان الهدف من مسابقة MURA هو تصنيف كل دراسة إلى طبيعية أو غير طبيعية (تنبؤات ثنائية)، وليس كل صورة. وكان أفضل أداء في جامعة ستانفورد. في مجموعة البيانات هذه، قام الفريق الذي نشرها بتدريب شبكة تلافيفية متصلة بكثافة DenseNet169 مكونة من 169 طبقة للكشف عن التشوهات وتحديد مكانها.

10- مواد وطرق البحث:

قمنا بتطبيق ثلاثة من نماذج شبكة Resnet المتوفرة حتى تاريخ إعداد هذا البحث من خلال تدريب نموذج لكل منها ثم قمنا بمقارنة أداء هذه النماذج. نماذج الشبكات التي قمنا بتجريبها هي ثلاثة نماذج، وهي ResNet152V2, InceptionV3, InceptionResNetV2. نود أن يحقق نموذجنا أداءً مشابهاً لأداء أطباء الأشعة. بالنسبة للدراسات التي أجريت على اليد والرسغ والمعصم، فإن نتائج النموذج الخاصة بنا أعلى قليلاً من ناحية الدقة من تلك النتائج الخاصة بأطباء الأشعة؛ وفي دراسات الكوع والساعد والعضد والكتف، كانت الدقة الخاصة بنموذجنا أقل قليلاً من درجات أخصائي الأشعة، ولكنها ليست ذات دلالة إحصائية، مما يشير ويؤكد إلى أن مجموعة البيانات تمثل مشكلة تحدي جيدة للبحث في المستقبل.

:Residual Networks (ResNet)

بعد أول بنية قائمة على CNN (AlexNet) والتي فازت بمسابقة ImageNet 2012، تستخدم كل بنية فائزة لاحقة المزيد من الطبقات في شبكة عصبية عميقة لتقليل معدل الخطأ، ويعمل هذا مع عدد أقل من الطبقات، ولكن عندما نزيد عدد الطبقات، توجد مشكلة شائعة في التعلم العميق مرتبطة بما يسمى Vanishing/Exploding gradient. يؤدي هذا إلى أن يصبح التدرج / 0 gradient أو كبيراً جداً. وبالتالي عندما نزيد عدد الطبقات، يزداد أيضاً معدل خطأ التدريب والاختبار.

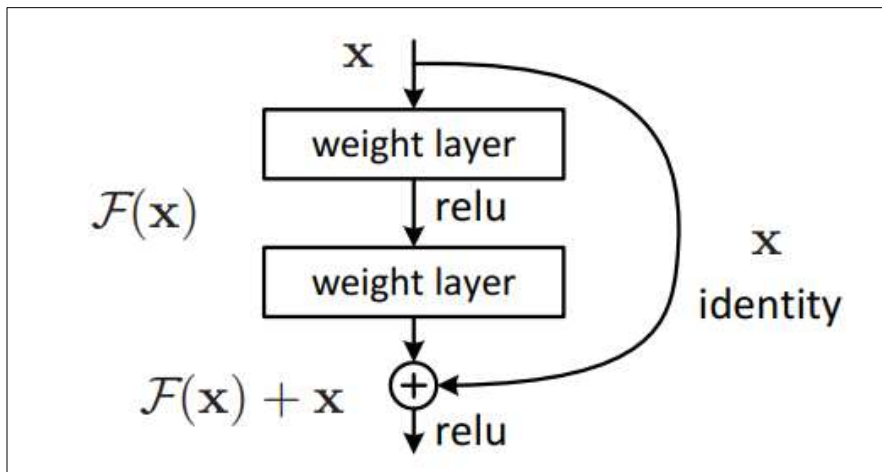
تم تقديم ResNet والتي اقترحها باحثون في Microsoft Research في عام 2015، كبنية جديدة تسمى الشبكة المتبقية (Residual Network).

الكتلة المتبقية (Residual Block):

من أجل حل مشكلة التدرج المتلاشي / المتفجر (vanishing/exploding gradient)، قدمت هذه البنية مفهوماً جديداً يسمى الشبكة المتبقية. في هذه الشبكة نستخدم تقنية تسمى تخطي الاتصالات. يتخطى الاتصال من خلال تخطي التدريب من عدة طبقات ويتصل مباشرة بطبقة الخرج.

النهج المستخدم في هذه الشبكة هو أنه بدلاً من أن تتعلم الطبقات التعيين الأساسي (mapping)، نسمح للشبكة بتناسب التعيين المتبقي.

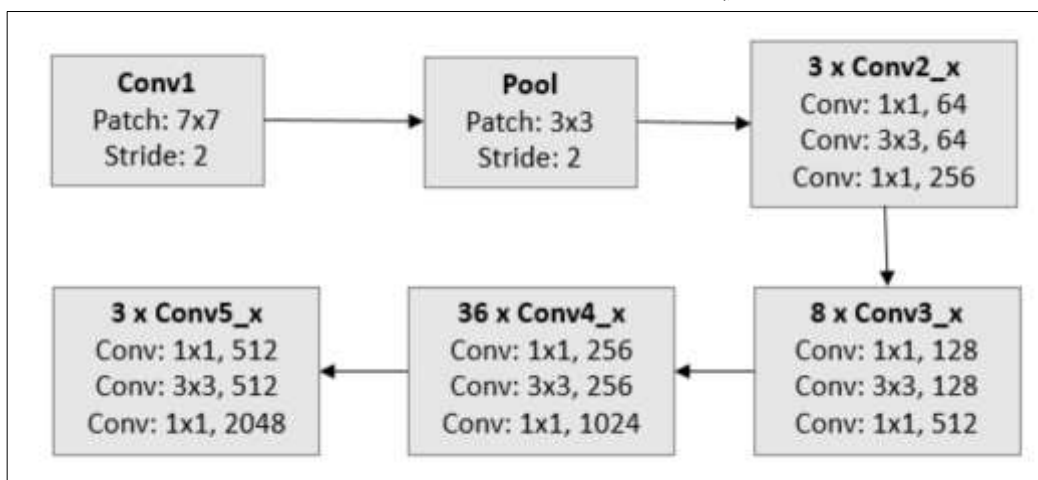
لذلك، بدلاً من قول $H(x)$ ، التعيين الأولي، تقوم الشبكة بما يلي $F(x) = H(x) - x$ والذي يعطي $H(x) = F(x) + x$. والشكل الآتي يوضح بنية الكتلة المتبقية (Residual Block)



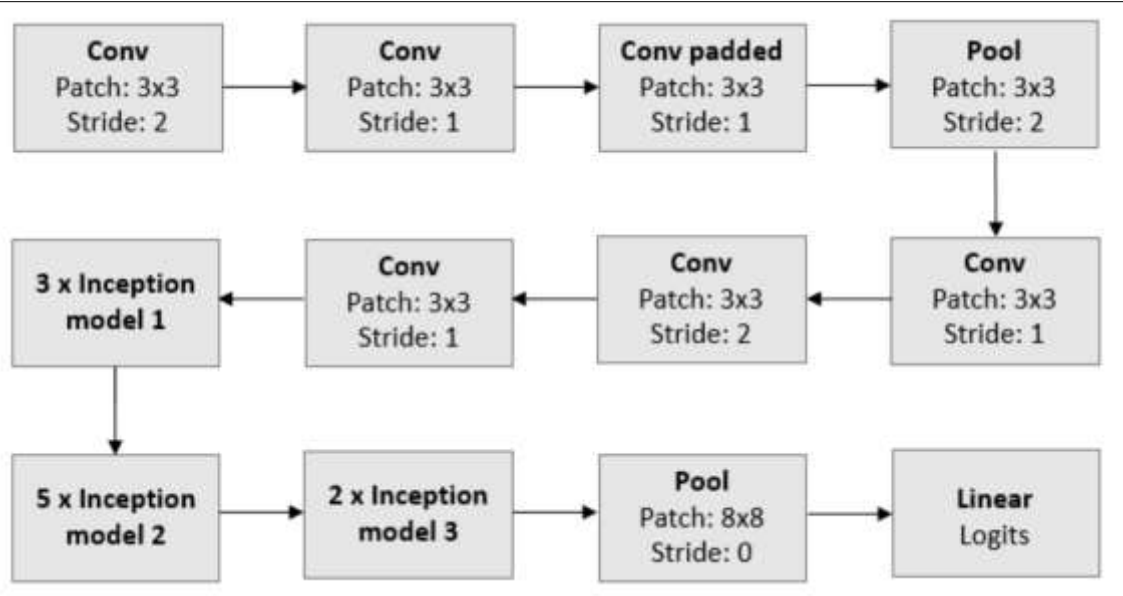
الشكل (4): بنية الكتلة المتبقية (Residual Block)

بنية الشبكة:

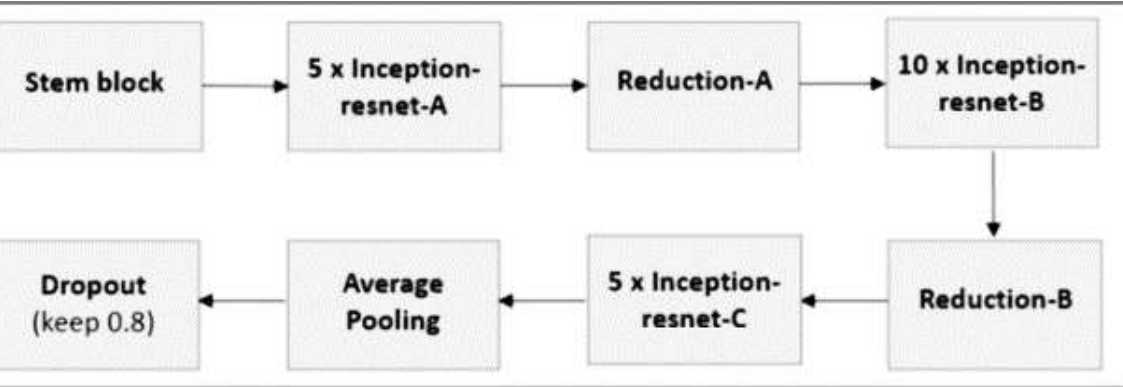
تستخدم هذه الشبكة بنية شبكة عادية مكونة من عدد من الطبقات مستوحاة من VGG-19 حيث يتم إضافة اتصال الاختصار (shortcut connection) بحيث تقوم اتصالات الاختصار هذه بتحويل البنية إلى شبكة متبقية.



الشكل (5): بنية شبكة Resnet152



الشكل (6): بنية شبكة InceptionV3



الشكل (6): بنية شبكة Inception-Resnet-v2

التدريب والاختبار العملي:

بيئة التكوين وإعداد البارامترات:

تم تطوير جميع التجارب وتنفيذها على منصة Google Colaboratory السحابية باستخدام Keras مع TensorFlow, كما تم التعامل مع منصة Kaggle لما تقدمه من ميزات للتعامل مع نماذج الشبكات العصبية أيضاً وبسبب القيود التي بدأت

تقرضها منصة Google Colab من حيث محدودية الموارد المتاحة وتحديد موارد الجلسة الواحدة وتقييد مدتها.

Colaboratory أو Colab ، هو أحد منتجات Google Research ، والذي يسمح للمطورين بكتابة وتنفيذ كود Python من خلال متصفحهم. يعد Google Colab أداة ممتازة لمهام التعلم العميق. كما يعتبر فعلياً كإصدار متخصص من Jupyter Notebook، والذي يعمل على السحابة ويوفر موارد حوسبة مجانية مثل GPUs وTPUs.

هناك عدة أسباب لاختيار استخدام Google Colab بدلاً من Jupyter Notebook العادي، وأهمها:

- وجود مكتبات مثبتة مسبقاً (Keras, TensorFlow, and PyTorch مثلاً)
 - يتم الحفظ على السحابة (يتم حفظ جميع دفاتر Google Colab ضمن حساب Google Drive، تماماً مثل ملفات Google Docs و Google Sheets)
 - التعاون (وهي ميزة أخرى يقدمها Colab فإذا كنت تعمل مع مطورين متعددين في مشروع ما، فمن الأفضل استخدام Colab. تماماً مثل التعاون في مستند محرر مستندات Google، كما يمكن البرمجة المشتركة مع مطورين متعددين باستخدام Colab notebooks)
 - استخدام GPU وTPU مجاناً (تتيح Google Research استخدام وحدات معالجة الرسومات (GPU) و (TPU) المخصصة لمشاريع التعلم الآلي كون تسريع GPU وTPU يحدثان فرقاً كبيراً حتى في المشاريع الصغيرة)
- أما بالنسبة لمنصة Kaggle فهي تابعة لشركة Google LLC، وهي مجتمع عبر الإنترنت لعلماء البيانات وممارسي التعلم الآلي. يسمح Kaggle للمستخدمين بالعثور على مجموعات البيانات ونشرها، واستكشاف النماذج وإنشائها في بيئة علم البيانات المستندة إلى الويب، والعمل مع علماء البيانات الآخرين ومهندسي التعلم الآلي، والدخول في مسابقات لحل تحديات علوم البيانات.

وقد كانت بدايتها في عام 2010 من خلال تقديم مسابقات التعلم الآلي وتقدم الآن أيضاً منصة بيانات عامة وبيئة عمل قائمة على السحابة (cloud-based workbench) لعلوم البيانات وتعليم الذكاء الاصطناعي.

خدمات Kaggle (إضافةً إلى الخدمات التي يقدمها Colab):

- مسابقات التعلم الآلي: كانت أول منتج لشركة Kaggle, بحيث تشر الشركات المشكلات ويتنافس المتسابقون لبناء أفضل خوارزمية، عادةً تكون بجوائز نقدية قيّمة.
- Kaggle Kernels: وهي بيئة عمل قائمة على السحابة لعلوم البيانات والتعلم الآلي, تسمح لعلماء البيانات بمشاركة الكود والتحليل في Python و R و R Markdown, وقد تمت مشاركة أكثر من 150 ألف نواة على Kaggle تغطي عدة مشاريع ومواضيع هامة.
- منصة لمجموعات البيانات العامة: يشارك أعضاء مجتمع Kaggle مجموعات البيانات مع بعضهم البعض. ونجد على هذه المنصة مجموعات بيانات عن كل شيء.
- Kaggle Learn: وهي منصة لتعليم الذكاء الاصطناعي.

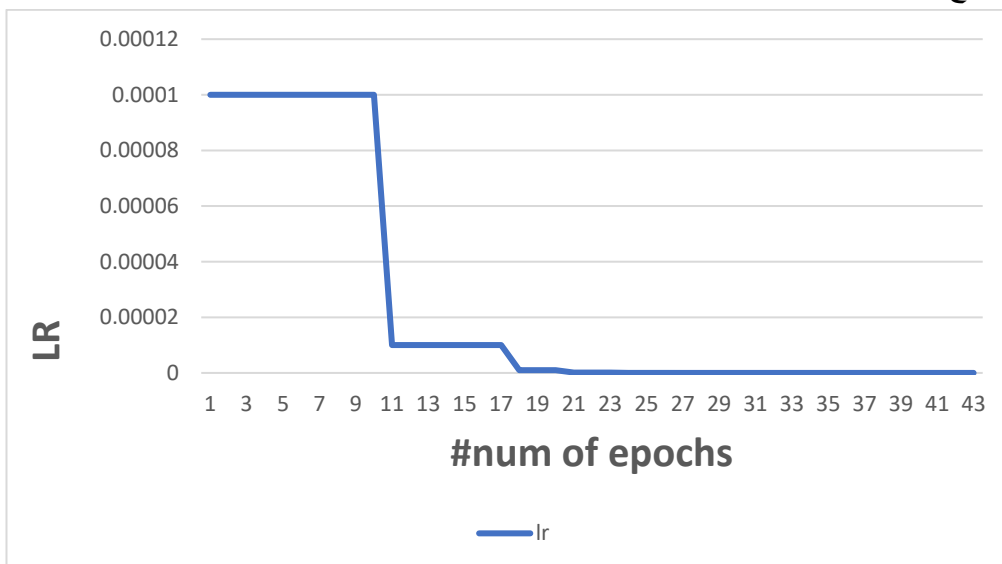
لمقارنة أداء الطرق المختلفة، قمنا بتعيين نفس البارامترات لكل الطرق المستخدمة، البارامترات هي: معدل التعلم الأولي 0.0001 ويتناقص تدريجياً ليصل إلى قيمة 0.000000001، حجم الـ Batch هو 32، وعدد الفترات القصوى هو 43 فترة (Epochs=43).

11-دراسة النتائج العملية

- مرحلة التدريب:

تمت تهيئة الشبكة بأوزان من نماذج الشبكات المختارة والتي تم تدريبها مسبقاً على عينة البيانات imagenet وتم تدريب الشبكة باستخدام Adam Optimizer مع

البارامترات الافتراضية. استخدمنا حجم batch = 32 ومعدل التعلم 0.0001 يتناقص تدريجياً مع مراحل التكرار (الشكل (7))، وقمنا بتدريب الشبكات لكل أنواع الدراسة بشكل منفصل بحيث قمنا بتطبيق نموذج الشبكة المحددة على عينة البيانات المدروسة على كل نوع بشكل مستقل (أي التطبيق على فئة المعصم بشكل مستقل، على فئة الأصابع بشكل مستقل... الخ) وبعدها مرات 43 لكل الأنواع، وقمنا بتسجيل النتائج لكل الحالات المدروسة.



الشكل (7): توضيح لقيم معدل التعلم (Learning Rate) خلال العمل على كل النماذج المدروسة

- تقييم الأداء:

قمنا بتقييم أداء النماذج المدروسة على الاختبار الذي تم تعيينه لكل أنواع الدراسة ومقارنتها معاً كما تمت مقارنتها بنموذج المجموعة (القسم التالي). تم إنشاء نموذج المجموعة باستخدام النماذج الثلاثة لنوع الدراسة (/ humerus العضد). تمت مقارنة الأداء باستخدام مقياس الدقة ومقياس كايا. لاحظنا أن أداء النماذج الفردية المدربة على نوع الدراسة المحددة أفضل بكثير مقارنةً بنموذج المجموعة.

- مقياس كوهين كابا (Cohen's kappa statistic)

المقياس المستخدم في مسابقة MURA ليس هو الدقة الكلاسيكية، ولكن درجة kappa أو مقياس كابا. وهو مقياس أكثر قوة من الدقة البسيطة، لأنه يأخذ في الاعتبار إمكانية حدوث الاتفاقية عن طريق الصدفة بطرحها من الاتفاقية الملاحظة.

ما هو كوهين كابا؟

يقيس كوهين كابا الاتفاق بين اثنين من المقيمين يصنف كل منهما عناصر N إلى فئات حصرية متبادلة. هناك طريقة بسيطة للتفكير في ذلك وهي أنه مقياس كمي للاعتمادية لاثنتين من المقيمين الذين يصنفون نفس الشيء، ويتم تصحيحه وفقاً لعدد المرات التي قد يتفق فيها المقيّمون عن طريق الصدفة.

تقييم كوهين كابا:

يمكن أن تكون قيمة kappa أقل من 0 (سالبة). تعني الدرجة 0 أن هناك اتفاقاً عشوائياً بين المقيمين، في حين أن الدرجة 1 تعني أن هناك اتفاقاً تاماً بين المقيمين. لذلك، فإن الدرجة الأقل من 0 تعني أن هناك اتفاقاً أقل من فرصة عشوائية.

		Rater 2	
		Correct	Incorrect
Rater 1	Correct	A	B
	Incorrect	C	D

الشكل (8): شبكة $N \times N$ المستخدمة لتفسير نتائج المقيمين

في الشكل تقسم الشبكة إلى الأجزاء الآتية:

A: إجمالي عدد الحالات التي قال كلا المقيمين إنها صحيحة، أي أن المقيمين متفقون.

B: العدد الإجمالي للحالات التي قال Rater 2 إنها غير صحيحة، لكن Rater 1 قال إنها صحيحة. وهذه تمثل حالة خلاف.

C: العدد الإجمالي للحالات التي قال Rater 1 إنها غير صحيحة ، لكن Rater 2 قال إنها صحيحة. وهذه تمثل أيضاً حالة خلاف.

D: إجمالي عدد الحالات التي قال كلا المقيمين إنها غير صحيحة، أي أن المقيمين متفقون.

من أجل إيجاد قيمة kappa، نحتاج أولاً إلى معرفة احتمالية الاتفاق (وهذا يفسر سبب تسليط الضوء على قطر الاتفاقية). بحيث يتم اشتقاق هذه الصيغة عن طريق إضافة عدد الاختبارات التي يتفق فيها المقيمون ثم تقسيمها على العدد الإجمالي للاختبارات. وفقاً للشكل، فهذا يعني $(A + D) / (A + B + C + D)$.

$$P_o = \text{Number in Agreement} / \text{Total}$$

والخطوة التالية هي إيجاد احتمالية الاتفاق العشوائي وهي إجمالي عدد المرات التي قال Rater 1 أنها صحيحة مقسومة على العدد الإجمالي للحالات، مضروبة في إجمالي عدد المرات التي قال Rater 2 أنها صحيحة مقسومة على العدد الإجمالي للحالات، المضافة إلى إجمالي عدد المرات التي قال Rater 1 أنها غير صحيحة مضروباً في إجمالي عدد المرات التي قال Rater 2 إنها غير صحيحة. هناك الكثير من المعلومات التي يجب أخذها هناك، ونختصرها بالمعادلة الآتية:

$$P_{(\text{correct})} = (A + B / A + B + C + D) * (A + C / A + B + C + D)$$

$$P_{(\text{incorrect})} = (C + D / A + B + C + D) * (B + D / A + B + C + D)$$

$$P_e = P_{(\text{correct})} + P_{(\text{incorrect})}$$

معادلة احتمالية الاتفاق العشوائي

وفي النهاية فإن صيغة Cohen's Kappa هي احتمال الاتفاق الذي يزيد احتمال الاتفاق العشوائي مقسوماً على 1 ناقصاً احتمال الاتفاق العشوائي.

$$K = P_o - P_e / 1 - P_e$$

صيغة معامل كبا كوهين

ملاحظات حول النتائج:

نوضح في الجداول أدناه نتائج المقارنة (من ناحية دقة التصنيف، مقياس كابا)، التي قمنا بإجرائها بين الشبكات الثلاث التي عملنا عليها وتم ذكرها سابقاً لتصنيف صور عينة البيانات المدروسة.

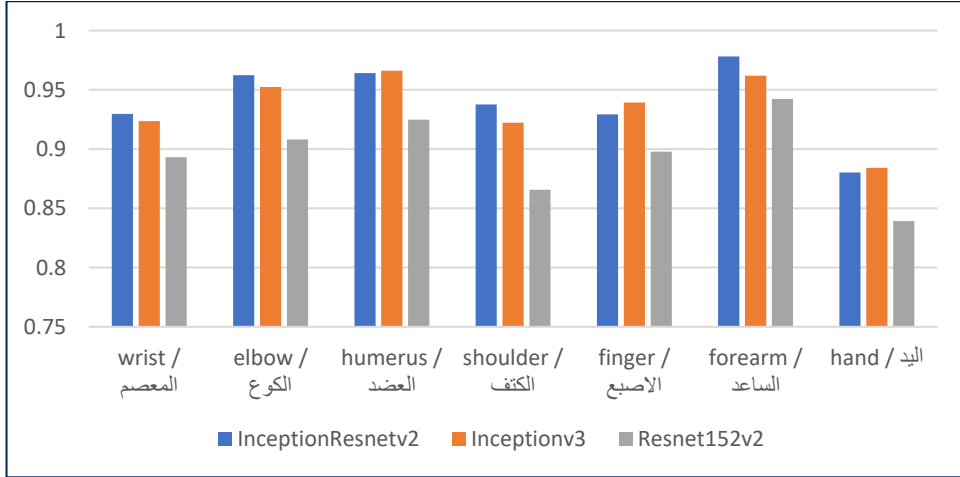
نلاحظ من الجدول الأول أننا حصلنا على أفضل دقة لأغلب (الكل ماعدا نوع المعصم) حالات الدراسة ضمن عينة البيانات من خلال نموذج الشبكة **Resnet152v2** بحيث وصلنا إلى دقة تصنيف تقريبا 88% في نوع الكوع والعضد وهي كانت الأفضل بين الأنواع الأخرى فقد وصلت دقة تصنيف نوع الكتف في هذا النموذج حوالي 78.5% ووصلت 79% في نوع الأصابع واقترب من 80% في تصنيف نوع الساعد وتجاوز هذه النسبة قليلاً في نوع اليد.

أما بالنسبة لنوع المعصم فقد تفوق أداء نموذج **Inceptionv3** على الأنواع الأخرى بدقة اقتربت من 86%.

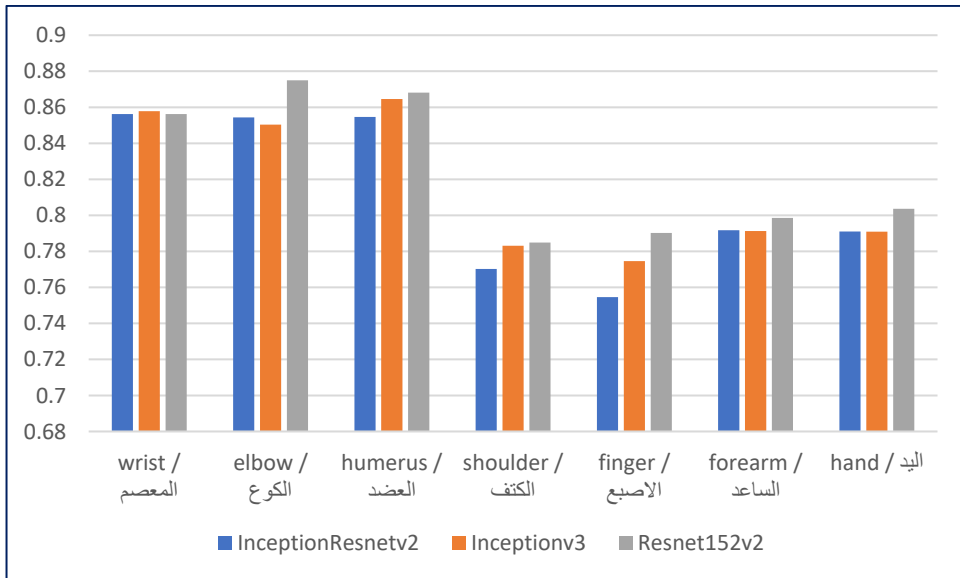
الجدول (1) نتائج مقارنة أداء الشبكات الثلاث المدروسة من ناحية دقة التصنيف

دقة التصنيف						نوع الدراسة
InceptionResnetv2		Inceptionv3		Resnet152v2		الشبكة المختارة
Train	test	train	Test	train	Test	
0.9296	0.8562	0.9236	0.8578	0.8931	0.8562	/ wrist المعصم
0.9624	0.8544	0.9524	0.8504	0.908	0.875	/ elbow الكوع
0.9641	0.8546	0.9661	0.8646	0.9248	0.8681	humerus / العضد
0.9376	0.7702	0.9222	0.7831	0.8656	0.7849	shoulder / الكتف
0.9292	0.7546	0.9392	0.7746	0.8978	0.7902	/ finger الاصبع
0.9782	0.7917	0.962	0.7913	0.9424	0.7986	/ forearm الساعد
0.8802	0.791	0.8842	0.7909	0.8392	0.8036	hand / اليد

كما نعرض في المخططين الآتيين توضيح لنتائج المقارنة من خلال المخططين التوضيحين لأداء النماذج الثلاثة على كل أنواع الدراسة لكل من التدريب والاختبار.



المخطط (1) مقارنة نتائج التدريب لأنواع الشبكات الثلاثة على كل أنواع الدراسة



المخطط (2) مقارنة نتائج الاختبار لأنواع الشبكات الثلاثة على كل أنواع الدراسة

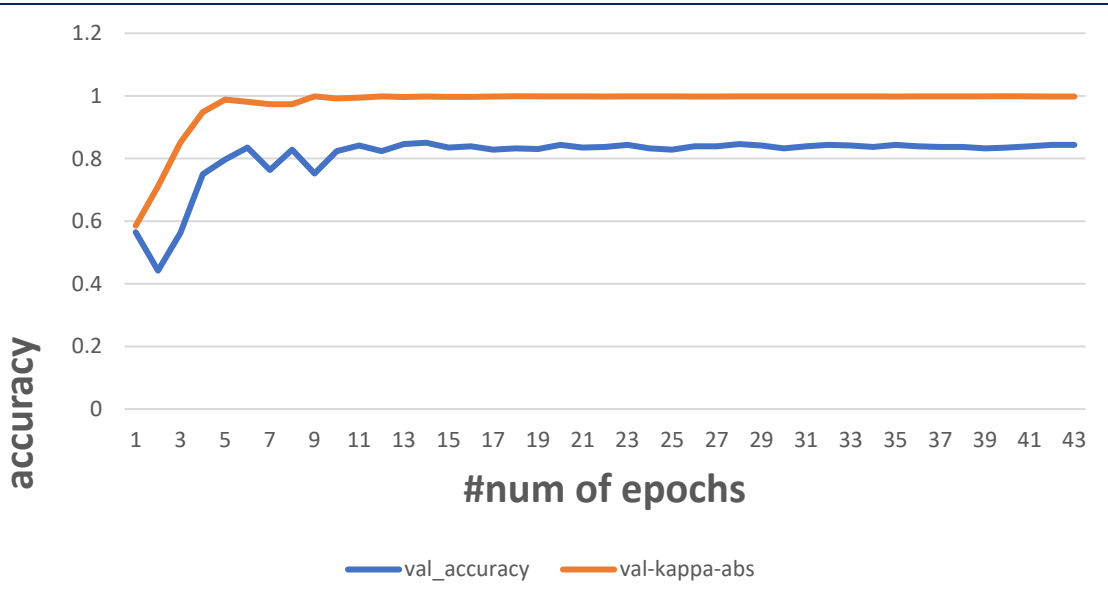
نعرض في الجدول الثاني نتائج مقياس كابا لحالات الدراسة ضمن عينة البيانات من خلال نماذج الشبكات الثلاثة.

Resnet152v2 بحيث حصلنا على قيم جيدة تتناسب مع دقة التصنيف المعروضة في الجدول الأول.

نجد من الجدول أنه بالنسبة لنوع المعصم والكوع فقد أعطت شبكة InceptionResnetv2 أعلى قيمة لمقياس كابا من ناحية الاختبار، بالنسبة للعضد كانت القيمة الأكبر لمعامل كابا لنموذج Inceptionv3 وتفوقت Resnet152v2 في باقي أنواع الدراسة بحصولها على أعلى قيم لهذا المقياس.

الجدول (2) نتائج مقارنة أداء الشبكات الثلاث المدروسة من ناحية مقياس كابا					
kappa static					
InceptionResnetv2		Inceptionv3		Resnet152v2	
Train	test	train	Test	train	Test
0.9545	0.9577	0.9373	0.9207	0.896	0.8497
0.9541	0.9488	0.9418	0.9099	0.9384	0.895
0.99578	0.9299	0.99621	0.948	0.997	0.942
0.9997	0.85	0.9993	0.889	0.9863	0.8627
0.9675	0.6649	0.9572	0.9598	0.9535	0.9649
0.9379	0.9138	0.9204	0.8894	0.9337	0.9194
0.841	0.8191	0.8649	0.8224	0.8965	0.8826

ونعرض في المخطط الآتي قيم الدقة وقيم مقياس كبا لإحدى حالات الدراسة ويظهر من خلاله التناسب الذي نتج في المقياسين (مقياس الدقة ومقياس كبا للتوافق).



المخطط (3) مخطط توضيحي لنتائج مقياسي الدقة وكبا لحالة دراسة الكوع لشبكة InceptionResnetv2

تطبيق طريقة التعلم الجماعي (Ensemble Method):

الخطوة الأخيرة في هذه الدراسة هي تطبيق تجميع (تم شرح التعلم الجماعي وآلياته سابقاً) للنماذج الثلاثة على واحدة فقط من حالات الدراسة وهي نوع العضد من خلال أكثر آلية تجميع شهيرة بين الآليات التي ذكرناها

الجدول (3) نتائج تطبيق التعلم الجماعي على نوع العضد (من ناحية دقة التصنيف)

دقة التصنيف				وع راسة
Avg	InceptionResnetv2	Inceptionv3	Resnet152v2	شبكة ختارة

single		ensemble		single		ensemble		single		ensemble		single		ensemble		
Test	train	Test	train	Test	train	Test	train	Test	train	test	train	Test	train	test		
0.85997																
0.98253																
0.8713																
0.9661																
0.8646																
0.9976																
0.8846																
0.9169																
0.8472																
0.979																
0.8572																
0.9248																
0.8681																
0.971																
0.8721																
																/ humerus

نلاحظ من الجدول الأخير التحسن في دقة نتائج التصنيف للنماذج الثلاثة للنوع المدروس (نوع العضد HUMERUS) فقد ارتفع أداء نموذج شبكة Resnet152v2 من ناحية دقة الاختبار من 0.8681 إلى 0.8721 وكذلك بالنسبة لنماذج شبكتي Inceptionv3 و InceptionResnetv2 فقد حصلنا أيضاً على زيادة في دقة التصنيف بحوالي 2% لكل من النموذجين من ناحية دقة الاختبار، في العمود الأخير للجدول قمنا بحساب متوسط الدقة الفردية للشبكات الثلاثة ومتوسط دقتها عند جمعها من خلال التكديس في نموذج جماعي لتوضيح التحسين النهائي الذي تم الحصول عليه.

12-النتائج والتوصيات:

- هناك العديد من التأثيرات البشرية في اتخاذ القرارات، بما في ذلك تحديد خيارات المعايير ووظائف التحسين. وبالتالي سنحتاج للخوارزميات للتفكير في طرق تدعم آليات التغذية الراجعة والتسجيل. كما تحتاج الحكومات والشركات والمؤسسات النظر في استخدام الخوارزميات في صنع القرار، والتشاور على نطاق واسع، والتأكد من وجود آليات لاكتشاف ومعالجة أي أخطاء أو عواقب غير مقصودة للقرارات المتخذة.

• تحديد نوع البيانات وكيفية استخدامها يمكن أن يساعد عملية صنع القرار في تقليل أي عدم يقين أو شكوك بشأن كيفية اتخاذ القرارات.

تسلط توصياتنا الضوء على الحالات التي يعمل فيها نموذج صنع قرار معين بشكل جيد (الحالة الخاصة التي تمت دراستها)، ولكن هذا النموذج قد لا يعمل مع جميع الحالات وفي جميع الظروف.

من الناحية البشرية فإنه مع نمو الفريق، يصبح اتخاذ القرارات معاً أكثر صعوبة. فعندما نضيف المزيد من الأشخاص، غالباً ما يصبح الوصول إلى الإجماع بطيئاً للغاية، وتصبح التنازلات التي يتعين القيام بها مؤلمة للغاية. رداً على ذلك، تضع بعض الفرق جميع مسؤوليات اتخاذ القرار على عاتق قائد الفريق، مما يثقل كاهلهم ويثبط عزيمة أي شخص آخر. ولكن هناك العديد من الطرق لاتخاذ القرار معاً، أكثر بكثير من مجرد إعطاء رأي لكل فرد أو إعطاء سلطة لشخص واحد فقط.

من الناحية الآلية (العمل مع الآلة) فإن نمو الفريق يمكن أن يدعم اتخاذ القرار ولا يكون هناك معاناة كما في العامل البشري والآراء المختلفة.

الأعمال المستقبلية:

يمكن توسيع هذا البحث ليشمل تطبيق التعلم الجماعي من خلال العمل على عدة شبكات أخرى وتكديس عدة نماذج من نماذج الشبكات المدروسة على عينة البيانات ومقارنة نتائجها مع نتائج البحث الحالي (في حال العمل على تقنية الجمع بالتكديس).

كما يمكن توسيع البحث بطريقة أخرى وهي تطبيق طرائق التعلم الجماعي المختلفة على عينة البيانات المحددة ودراسة نتائج كل طريقة وتحديد من الأفضل بينها حسب الحالة المدروسة.

13-المراجع:

- 1- John Z., Alistair K., James M., Colin G., 2019 – Algorithmic Decision-Making and the Control Problem, Springer, China.

- 2- Schmidt, J., Marques, M.R.G., Botti, S. *et al.* 2019 - Recent advances and applications of machine learning in solid-state materials science. npj Comput Mater.
- 3- Panagiotis P., Ioannis E. L., (2020) – Ensemble Algorithms and Their Applications, Department of Mathematics, University of Patras, Greece.
- 4- Sarker I.H., (2021) – Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI.
- 5- Kong Y., Yu T., (2018) - A Deep Neural Network Model using Random Forest to Extract Feature Representation for Gene Expression Data Classification. Sci Rep.
- 6- Jarrod C., Kane H., Rituparna D., Aviv S., (2019) - Failing & Falling (F&! F): Learning to Classify Accidents and Incidents in Aircraft Data, IEEE.
- 7- Kyoungok Kim, (2016) - A Hybrid classification algorithm by subspace partitioning through semi-supervised decision tree, Springer.
- 8- Congbo M.A., Wei Emma Z., Mingyu G., Hu W., Quan Z. Sh., (2020) - Multi-document Summarization via Deep Learning Techniques: A Survey, arXiv:2011.04843v2 [cs.CL].
- 9- Yanyang Gu., (2019) - Melanoma Detection Based on Color and Hyperspectral Imaging, School of Info & Comm Tech.
- 10- Rohit B., Aditya Kh., Mohammad Sh., Gaurav D., Parneet S., (2021) - Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning, Computational Intelligence and Neuroscience.
- 11- Wenqian Ch., Shuyu Ch., Hancui Z., Tianshu W., (2017) - A hybrid prediction model for type 2 diabetes using K-means and decision tree, 8th IEEE International Conference on Software Engineering and Service Science (ICSESS).
- 12- Xinli Y., David L., Xin X., Jianling S., (2017) - TLEL: A two-layer ensemble learning approach for just-in-time defect prediction, Information and Software Technology.

- 13- Peyman Sh. M., Sławomir N., Sepideh P., (2020) - Stacked Ensemble of Recurrent Neural Networks for Predicting Turbocharger Remaining Useful Life, Center for Applied Intelligent Systems Research (CAISR).
- 14- Li J., Li X., He D., (2019) - A Directed Acyclic Graph Network Combined with CNN and LSTM for Remaining Useful Life Prediction, IEEE.
- 15- Breiman L., (1996) - Bagging Predictors, Springer.
- 16- D H. Wolpert, (1992) - Stacked Generalization, SiteSeer.
- 17- www.usbji.org
- 18- [Musculoskeletal conditions \(who.int\)](http://www.who.int)
- 19- Pranav R., Jeremy I. , Aarti B. , Daisy D., Tony D., Hershel M., Brandon Y., Kaylie Z., Dillon L., Robyn L. B., Curtis L., Katie S., Matthew P. L., Andrew Y. Ng, 2018 - MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs, 1st Conference on Medical Imaging with Deep Learning.

