نظام التعرف على لغة الإشارة التونسية للإشارات الثابتة غير المتماثلة ثنائية اليدين باستخدام التعلم العميق الانتقالي

آمنة دقنو (1)، هيثم الهرميسي (2)، نبيل تبان (3)
(1,3) المدرسة العليا للاتصالات بتونس – (SUP'COM) تونس
(2) المعهد العالي لعلوم الكمبيوتر – تونس

emna.daknou@supcom.tn
haithem.hermessi@fst.utm.tn
nabil.tabbane@supcom.tn

***الملخص*:**

يستخدم الأشخاص الصم وضعاف السمع لغات الإشارة في التفاعل فيما بينهم ومع الأشخاص الآخرين. ويُعتبر التعرف التلقائي على الإشارات الثابتة غير المتماثلة ثنائية اليدين عمليةً معقدة، حيث يتطلب نظام معالجة متقدم لفهم الصورة. في هذه الورقة، قدمنا مجموعة بيانات مكونة من 2000صورة تشمل 12إشارة تونسية ثابتة غير متماثلة ثنائية اليدين كما استخدمنا آلية التعلم الانتقالي للتعرف التلقائي محققين درجة دقة بنسبة 98.29٪. وتثبت المحاكاة أن هذه القيمة المرتفعة للدقة قد تم الحصول عليها بواسطة نموذج (Xception) عند دمجه مع محسن (Adagrad) مما يشير إلى أن نهجنا يحقق نتائج عالية على الرغم من استخدام مجموعة بيانات صغيرة.

الكلمات المفتاحية: لغة الإشارة التونسية، التعلُّم الانتقالي، الإشارات غير المتماثلة ثنائية اليدين

1. المقدمة

ارتفع عدد الأشخاص الذين يعانون من فقدان السمع، وفقًا لمنظمة الصحة العالمية، إلى 466مليون شخص، أي ما يعادل 6٪ من سكان العالم. يواجه هؤلاء حواجز تواصل كبيرة، خصوصًا في مجالات الرعاية الصحية، والتعليم، والقوى العاملة، والنقل. ومع ذلك يحتاج الأشخاص الصم في كثير من الحالات إلى توافر دائم للمترجمين الذين يعملون كجسر اتصال للتعامل مع المجتمع القادر على الكلام والسمع [1].

إن هذه العملية ليست قابلة للتنفيذ عادةً كما أنها تحتاج توفر ميزانية عالية وخاصة في البلدان النامية التي تواجه مشكلة نقص حاد في خدمات الترجمة بسبب نقص التدريب لمترجمي لغة الإشارة ونظرًا للعدد الكبير من الأشخاص الصم، عمل الباحثونمن الأشخاص الصم فقد عمل الباحثون في جميع أنحاء العالم على التخفيف من فجوة الاتصال هذه من خلال إنشاء إطار عمل التعرف الآلي على لغة الإشارة [2].

يتم تصنيف إشارات اليد بشكل أساسي إلى ثلاثة أقسام رئيسية: 1) إشارات اليد الواحدة التي تستخدم يدًا واحدة. 2) إشارات متماثلة بكلتا اليدين (ثنائية اليدين) حيث تكون حركات وأشكال اليدين متطابقة. 3) إشارات غير متماثلة بكلتا اليدين (ثنائية اليدين) والتي تتم عبر تحريك اليد الأساسية والسماح لليد الأخرى التابعة بالعمل كقاعدة [3]. ويمكن تصنيف إيماءات اليد على أنها إما ثابتة أو ديناميكية. لقد كان هناك الكثير من الأبحاث في مجال التعرف على لغة الإشارة حول كل من الإيماءات الثابتة والمتحركة لتفسير لغات مختلفة مثل لغة الإشارة الأمريكية ولغة الإشارة الهندية ولغة الإشارة الصينية. ومع ذلك فعندما نتعمق في التعرف على الإشارات الثابتة نجد أن المؤلفين كانوا يتعاملون مع الحروف الأبجدية والأرقام التي يتم التعبير عنها من خلال إشارات اليد الواحدة [4]. فهم لم يتعاملوا على نطاق واسع مع كلمات الإشارة الثابتة غير المتماثلة ثنائية اليدين. يشكل التعرف التلقائي على هذه الإيماءات تحديًا كبيرًا بسبب التعقيد العالي في معالجة الصور.بسبب التعقيد الكبير لإدراك الصورة. كما يضيف عدم التماثل تعقيدًا في هذا المجال لأن النموذج يحتاج إلى مراعاة الأشكال المختلفة لكل يد.

يبدو أن لغة الإشارة التونسية (TnSL) هي اللغة الوطنية الرسمية للمواطنين الصم وضعاف السمع في تونس [5] مع اختلافها بشكل كبير عن لغات الإشارة الأخرى. وقد استخدمنا في هذا السياق شبكة العصبية الالتفافية (Convolutional Neural Network) جديدة يمكنها التعرف بشكل صحيح على كلمات لغة الإشارة التونسية الثابتة غير المتماثلة ثنائية اليدين. ويستفيد إطار عملنا بشكل خاص من أدوات التعلم الانتقالي (TL) من خلال ضبط نماذج الشبكة الحديثة المدربة مسبقًا على قاعدة بيانات (ImageNet) لأن التعلم الانتقالي [6] يمكنه التعامل بنجاح مع ندرة البيانات وتعزيز أداء التعرف على الإشارات. ونحن نهدف من خلال تجاربنا إلى إيجاد أفضل بنية نموذجية يمكنها التكيف مع مجموعة بيانات لغة الإشارة التونسية صغيرة الحجم المكونة من 2000 صورة والتعامل بكفاءة مع الإشارات الثابتة ثنائية اليدين.

2. مراجعة الأدبيات

تتكيف غالبية حلول تصنيف لغة الإشارة المذكورة في الأدبيات [7] مع مجموعات البيانات الكبيرة وهي ليست مستقرة عند تدريبها على مجموعات بيانات صغيرة الحجم. كما أن هناك تكلفة كبيرة لعملية جمع الصور وإنشاء أي مجموعة بيانات كبيرة إضافة إلى احتياجها جهدًا لوجستيًا كبيرًا. وبالتالي تثير مجموعات البيانات الصغيرة تساؤلًا حول ما إذا كان التعلم العميق قابلاً للتطبيق في البيئات ذات البيانات القليلة. وفي الواقع فإنه من النادر ومن الصعب على مجموعات البيانات ذات الأحجام الصغيرة الاستفادة من التعلم العميق بسبب مشكلة عدم الكفاءة عند التعامل مع البيانات الجديدة التي تحدث عند تنفيذ نماذج الشبكة العصبية الترشيحية (Convolutional Neural Network). ومن ثم فإننا نشير هنا إلى العديد من الأعمال التي تناولت تصنيف لغة الإشارة في مجموعات البيانات الصغيرة.

يطبق العمل في [8] نظامًا قائمًا على الرؤية لترجمة الحروف الأبجدية العربية إلى كلمات منطوقة باستخدام مجموعة بيانات مكونة من 3875 صورة. وبهدف تسهيل تعميم النموذج بشكل أفضل على البيانات غير المرئية قام المؤلفون باستخدام ميزة توليد البيانات الجديدة في عملية التدريب. وتحقق هذه الممارسات دقة بنسبة 90٪ مما يضمن كون هذا النظام موثوق وفعال للغاية. رغم النتائج الجيدة ضمن مجموعة بيانات صغيرة، إلا أن هذا النهج يركز فقط على الإشارات أحادية اليد.

قام المؤلفون في [9] بتنفيذ نظام التعرف على الشبكات العصبية الالتفافية. (CNN) لترجمة أبجدية لغة الإشارة البريطانية (BSL) بما يشمل مجموعة بيانات تضم حوالي 10000 صورة وتحتوي على 19 فئة. وتوجد في هذه الإشارات 12 إشارة غير متماثلة ثنائية اليدين. وتمر الصور بخطوات الترشيح التالية قبل التدريب: إزالة الخلفية والتحويل إلى تدرج الرمادي وتطبيق مرشح الضبابية (Gaussian blur filter) للاحتفاظ بميزات اليد الرئيسية. وعلى الرغم من أن هذا العمل قد ركز على الإيماءات ثنائية اليدين إلا أن معدل دقته أقل من 90٪ ولا يحقق نتائج مقبولة.

تم نشر ورقة بحثية حول نظام التعرف على لغة الإشارة البنغالية باستخدام شبكة (VGG-v16) المدربة مسبقًا لتصنيف 37 حرفًا من الأبجدية البنغالية ضمن مجموعة بيانات مكونة من 1147 صورة في [10]. ويتم التعبير عن هذه الحروف البنغالية من خلال إشارات ثنائية اليد وغير متماثلة. ومع ذلك فقد حصل هذا النموذج على دقة تحقق أقل من 90٪ مما يدل على أنه يتطلب المزيد من التحسينات للتكيف مع الميزات المعقدة.

وتقدم دراسة أخرى في [11] مصنفًا عميقًا قائمًا على الشبكة العصبية الالتفافية (CNN) يتعرف على كل من صور الحروف والأرقام في لغة الإشارة الأمريكية باستخدام مجموعة بيانات مكونة من 2515 صورة. وللتغلب على ندرة البيانات ومشكلة عدم الكفاءة عند التعامل مع البيانات الجديدة يستخدم هذا النموذج تقنيات توليد البيانات الجديدة في عملية التدريب. وقد حقق هذا النهج وفقًا لنتائج المحاكاة أداءً جيدًا دقة تصنيف تبلغ 94.34٪ ضمن مجموعة البيانات الصغيرة الحجم. ومع ذلك فإن جميع الإشارات المتضمنة يتم تنفيذها بيد واحدة فقط.

بناءً على هذه الملاحظات نلاحظ أن معظم النماذج المذكورة قد ركزت على الإشارات أحادية اليد ولم تتعامل بشكل فعال مع الإشارات غير المتماثلة ثنائية اليد. ونظرًا لأننا ندرك أن الإشارات غير المتماثلة ثنائية اليد تهيمن على معظم لغات الإشارة فقد قمنا بإنشاء مجموعة بيانات للغة الإشارة التونسية (TnSL)تتضمن 12فئة من الكلمات .التي يتم التعبير عنها جميعًا من خلال حركات ثنائية اليد. وللعثور على أفضل نموذج للتعرف على الإيماءات الثابتة في لغة الإشارة التونسية يستفيد نهجنا من أدوات التعلم الانتقالي من خلال استخدام بعض البنى البرمجية للشبكات الحديثة الشائعة المدربة مسبقًا على مجموعة بيانات (ImageNet) و[12] من خلال اختبار المحسنات المستخدمة بشكل شائع. ولذلك تقدم هذه الدراسة المقارنة رؤى تنفيذ نموذج الشبكة العصبية الالتفافية (CNN)الأنسب. للتعرف الثابت على لغة الإشارة التونسية .

3. المنهجية المقترحة

3.1 معالجة البيانات

من الضروري قبل مرحلة التدريب المرور بعملية إعداد البيانات لجعل مجموعة بيانات لغة الإشارة التونسية الخاصة بنا مدخلات متوافقة مع النماذج المختلفة.

3.1.1 جمع البيانات

نهدف إلى بناء مجموعة بيانات للغة الإشارة التونسية تتضمن 12فئة من كلمات الإشارة غير المتماثلة ثنائية اليد .تشمل فئات الكلمات" :قهوة"، "شاي"، "انتخابات"، "قانون"، "مساعدة"، "رقص"، "جمعية"، "سجن"، "علم النفس"، "وزارة"، "بلدية"، و"حكومة ."تلتقط صور الإيماءات الثابتة باستخدام كاميرا ويب تحت إضاءات مختلفة وخلفية ثابتة، بإجمالي 2000صورة، بحيث تحتوي كل فئة على أكثر من 160صورة، بتنسيق (RGB)وبدقة عالية، ومعدلة إلى قياس (224×224)بكسل .مضبوطة باستخدام وحدة معالجة الصور (OpenCV). يبلغ عدد الصور إجمالًا 2000صورة، بحيث تحتوي كل فئة على أكثر من 160صورة بتنسيق (RGB)وبدقة عالية، ومعدلة إلى قياس (224×224)بكسل.

3.1.2 إعادة تنظيم البيانات

ظرًا لاختلاف عدد الصور في كل فئة، قد يؤدي عدم التوازن بين الفئات إلى زعزعة استقرار عملية التدريب. ولذلك فإنه يجب أن يكون هناك عدد متساوٍ من الصور بين جميع الفئات الـ 12 للتخفيف من هذا التفاوت. في كل مرة يختار البرنامج عشوائيًا 54 صورة من كل مجلد ويخلطها ويزيل الباقي. ونظرًا لوجود 3 تكرارات للعملية فإن مجموعة البيانات النهائية تحتوي بالتالي على 1944 صورة ويحتوي كل مجلد على 162 عينة. ويعرض الشكل 1 بعض العينات لمجموعة بيانات لغة الإشارة التونسية.

الشكل 1. كلمات لغة الإشارة التونسية

3.1.3 تقسيم البيانات

تنقسم مجموعة بيانات لغة الإشارة التونسية الخاصة بنا إلى مجموعات تدريب وتحقق واختبار بنسبة 80% و10% و10% على التوالي. وتجعل هذه العملية مجموعة البيانات أكثر قوة حيث سيتم التدريب على نسبة تقسيم بيانات التدريب والتحقق.

3.1.4 تعزيز توليد البيانات الجديدة (**Data Augmentation**)

أخيرًا، نقوم بتعزيز بيانات التدريب من خلال توليد بيانات جديدة .ومع زيادة حجم مجموعة التدريب والتسلسل الأكثر تنوعًا من الصور، يمكننا إنشاء نماذج أكثر عمومية وكفاءة في التعامل مع البيانات الجديدة .تتضمن التعديلات المطبقة نطاق السطوع [1.2- 0.5]، ونطاق التكبير 1.0]، [1.2، ونطاق الدوران °10-]، [°10+، ونطاق التحول الرأسي بنسبة 10%، والتحول الأفقي بنسبة .%10بعد ذلك، يتم توحيد نمط جميع الصور في مجموعة البيانات من خلال إعادة قياس قيم البكسل إلى نطاق جديد .(0,1)3.2 التعلم الانتقالي

التعلم الانتقالي هو أحد مجالات التعلم العميق الذي يعيد استخدام نموذج مدرب مسبقًا على مجموعة بيانات كبيرة، وتطبيقه على مهمة جديدة غالبًا مع مجموعة بيانات صغيرة لزيادة الدقة .نعرض فيما يلي النماذج المدربة مسبقًا التي سيتم اختبارها في هذه الدراسة:

3.2.1 (InceptionV3)

(InceptionV3) [13] هو نموذج شهير للتعلم الانتقالي تم إصداره في عام 2015 ويأتي من عائلة (Inception) ذات بنية الشبكة العصبية الترشيحية (CNN). ونظرًا لكونه مناسبًا تمامًا للمواقف التي يكون فيها قيود على موارد الحوسبة فإن هذا النموذج يتفوق في عمليات محددة مثل اكتشاف الكائنات وتصنيف الصور. ويتكون (InceptionV3) من 48 طبقة ويقدم تحسينات على إصداراته السابقة بما في ذلك دمج ميزات (label smoothing) و(convolutions) (7 × 7).

3.2.2 (Xception)

[13] (Xception)هو نموذج شبكة عصبية التفاعلية أطلقه باحثو .(Google)مستوحى من بنية (Inception)، يتم استبدال طبقات (Inception)بطبقات التفافية منفصلة حسب العمق (Depth-wise Separate Convolution Layers)، مما يسرع عملية التصنيف ويحقق دقة أعلى مقارنة بنماذج (Inception)عند التدريب على مجموعة بيانات .(ImageNet)

3.2.3 (VGG-v16)

(VGG-16)، التي أُطلقت بواسطة مختبر Visual Geometry Groupبجامعة أكسفورد، هي واحدة من أشهر خوارزميات التعلم الانتقالي في مهام تصنيف الصور .[13]تتسم هذه الخوارزمية ببنية مرنة وبسيطة؛ إذ تحتوي على 16طبقة منها 13طبقة التفافية بحجم مرشح (3 × 3)، مما يسهل إدارة الشبكة ويحقق أداءً قويًا.

3.2.4 (VGG-v19)

ُعد (VGG-19)امتدادًا لنموذج [13](VGG-16)، حيث تحتوي على 19طبقة بدلًا من .16وتحتفظ بنفس بنية (VGG-16)مع طبقات إضافية من الالتفاف والتجميع الأعظمي .(Max-pooling)وتُظهر (VGG-19)دقة أعلى عند اختبارها على مجموعة بيانات (ImageNet)بفضل طبقاتها الإضافية

3.2.5 (MobileNetV2)

كما يتضح من اسمه، تم تصميم (MobileNetV2)للتطبيقات المحمولة .[13]ويُعد أول نموذج للرؤية الحاسوبية من (TensorFlow)مخصص للأجهزة المحمولة .يتميز (MobileNetV2)بأنه يتطلب طاقة حسابية ووقت تنفيذ أقل مقارنةً بالبنى البرمجية الأخرى.

3.3 الضبط الدقيق للنماذج المدربة مسبقًا

نقوم بضبط النماذج المدربة مسبقًا المذكورة أعلاه وإعادة تدريب كل منها على مجموعة بيانات لغة الإشارة التونسية عن طريق تثبيت الطبقات الأولى واستبدال الطبقات الأخيرة المتصلة بالكامل .وفيما يلي التعديلات الرئيسية التي ندمجها:

3.3.1 طبقة المدخلات

يتم تغيير حجم الصور إلى الشكل (3×224×224)قبل بدء عملية التدريب، ليتمكن مولد بيانات الصور (ImageDataGenerator)من إدخالها إلى الشبكة.

3.3.2 إضافة كتلة (**Block**)

في كل نموذج، نقوم بإزالة بعض الطبقات المتصلة بالكامل (Fully Connected - FC) من كل شبكة أساسية مرشحة لتناسب مجموعة البيانات الخاصة بنا ونضيف كتلة جديدة من 4 طبقات في أسفل البنية الجاهزة. حيث يجعل إدراج مثل هذه الكتلة نموذجنا بناءً ومناسباً للتنفيذ وفقًا لتعقيد وتنسيق مجموعة بيانات لغة الإشارة التونسية. وتتألف كتلة الطبقات الأربع الإضافية على وجه التحديد من: (GlobalAveragePooling2D Layer) و (FC1) من 1024 وحدة مع (Tanh) كدالة تنشيط (AF) و (FC2) من 1024 وحدة مع (Tanh) كدالة تنشيط (AF) و (FC3) من 512 وحدة مع (Tanh) كدالة تنشيط (AF). ويعزز استبدال الدالة (Relu) المستخدمة بشكل شائع في طبقات (FC) بدالة (Hyperbolic Tangent) (Tanh) من عملية تدريب النماذج ويجعلها أسرع دون التأثير على الأداء العام. ويمكن التعبير عن دالة (Tanh) في المعادلة 1 التالية:

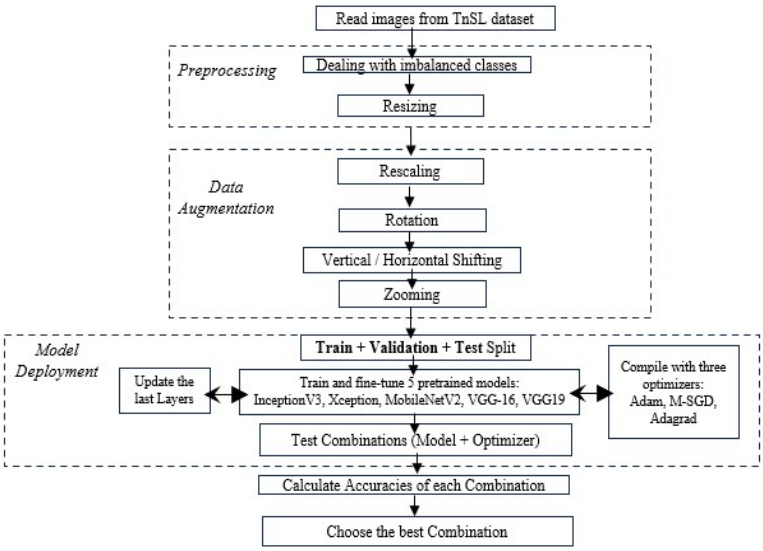
$$F_x = \frac{1 - \exp(2x)}{1 + \exp(2x)}$$

3.3.3 طبقة المخرجات

يتم ضبط طبقة المخرجات الأخيرة (Output Layer - OL)لتتوافق مع عدد الفئات المطلوب، وهو 12، حيث تستدعي طبقة المخرجات دالة (Softmax)للتمييز بين الإيماءات

3.4 المحسِّنات

يعد المحسِّن عنصرًا أساسيًا في إعداد النموذج لعملية التدريب .وبنفس المنطق المذكور أعلاه، نختار الطرق الشائعة في الأدبيات، وهي(Mini-batch Gradient Descent - M-SGD) :، و(Adam)، و (Adagrad) لتدريب كل من النماذج الخمسة المدرجة، ونختار الأنسب لحالتنا .يقدم الشكل 2مخطط التدفق العام للنهج المقترح.



الشكل 2. سير العمل المقترح للتعرف على لغة الإشارة التونسية

4. التجارب وعمليات التقييم

4.1 إعداد التجربة

يتم إجراء التجارب باستخدام منصة (Google Colaboratory) حيث نستخدم هذه الأطر الأساسية: (Keras) و (TensorFlow) و (Numpy) و(Matplotlib) إلخ. ونقوم أثناء مرحلة المحاكاة هذه بتقديم ثلاثة سيناريوهات اعتمادًا على نوع المحسِّن: السيناريو 1 والسيناريو 2 والسيناريو 3 والتي يقابل كل منها على التوالي كل من (M-SGD) و (Adam) و (Adagrad).

ونقوم في كل مجموعة تدريب باختيار نفس قيمة المعلمات الفائقة (hyper parameters). ونختار قيمة الدفعة 64 لضخ 64 عينة صورة في كل تكرار لعملية التدريب وذلك بعد تقييم السيناريوهات بأحجام دفعات مختلفة: 32 و64 و128. أما بالنسبة لمعدل التعلم فنختار قيمة 0.0001. ثم نضيف مدة انتظار الإيقاف المبكر في حال عدم تحسن النموذج (Early Stopping with Patience) لتكون 8 بعد تجربة قيم مختلفة (4 و5 و8) لتجنب مشكلة عدم الكفاءة عند التعامل مع البيانات الجديدة (over-fitting). ونستخدم بعض مقاييس التقييم كالدقة والتذكر والإرباك (Accuracy, Recall, F1-Score, Precision) (Confusion Matrix &) لقياس أداء النماذج المقترحة وتصور تأثير كل مجموعة من المعلمات (النموذج المدرب مسبقًا والمحسن) قبل اتخاذ القرار النهائي.

4.2 تقييم النموذج

تهدف التجارب في هذا القسم إلى ضبط الشبكة لتحقيق أعلى دقة اختبار تقيس قدرة النموذج على التعميم على البيانات غير المرئية .يتم تنفيذ ذلك في خطوتين :أولاً بتصور تأثير كل من المحسنات الثلاثة على النماذج المختلفة، وثانيًا بتحليل مختلف عمليات التنفيذ التي تمت بواسطة هياكل الشبكة الخمسة المدربة مسبقًا

4.2.1 إعداد مقارنة التعلم الانتقالي

نعتمد على مقاييس الدقة والفاقد لتحديد المحسن الذي يبدو أنه يحقق أفضل أداء على مجموعة بيانات التحقق. من الواضح في الشكل 3 (ب) والشكل 3 (ه) أن المحاولات التي تمت باستخدام المحسن (Adam) تقدم أداءً ضعيفًا لجميع النماذج المدربة مسبقًا. تُظهر التقلبات في عمليات المرور المحلية على كامل بيانات التدريب (epochs) أن (Adam) يواجه صعوبات في التحسين نحو حل جيد، ويتخذ خيارات مختلفة في نقاط مختلفة من عملية التعلم. وهذا يدل على مشكلة فرط التكيف (overfitting) عند التعامل مع البيانات الجديدة، حيث يقدم النموذج أداءً ضعيفًا على البيانات غير المرئية.

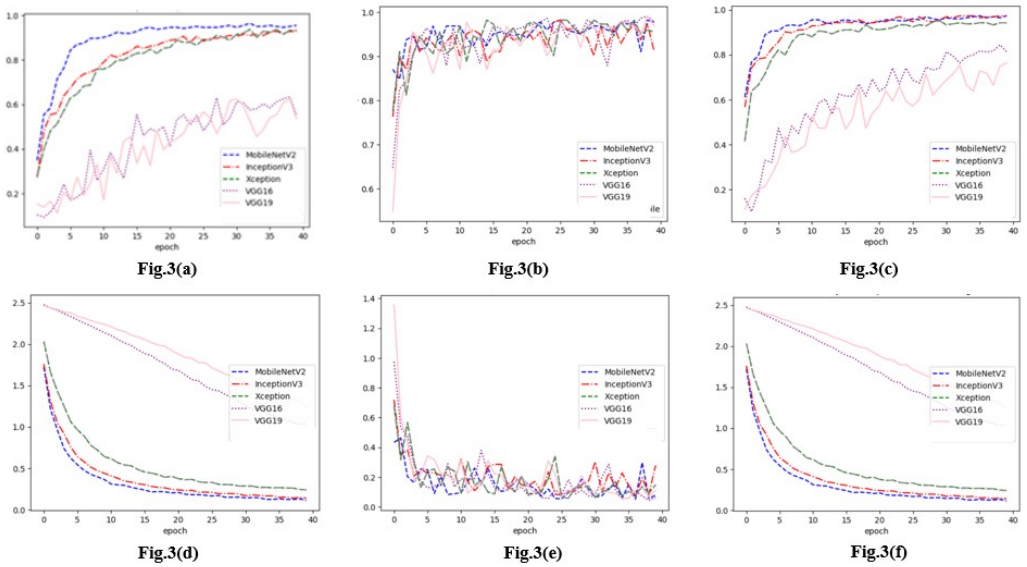


Fig.3(a) Fig.3(b) Fig.3(c)

Fig.3(d) Fig.3(e) Fig.3(f)

الشكل 3. (3.أ) دقة التحقق من الأداء (Validation Accuracy) باستخدام (M-SGD)، (3.ب) دقة التحقق من الأداء باستخدام (Adam)، (3.ج) دقة التحقق من الأداء باستخدام (Adagrad)، (3.د) فاقد التحقق (Validation Loss) باستخدام (M-SGD)، (3.هـ) فاقد التحقق باستخدام (Adam)، (3.و) فاقد التحقق باستخدام (Adagrad)

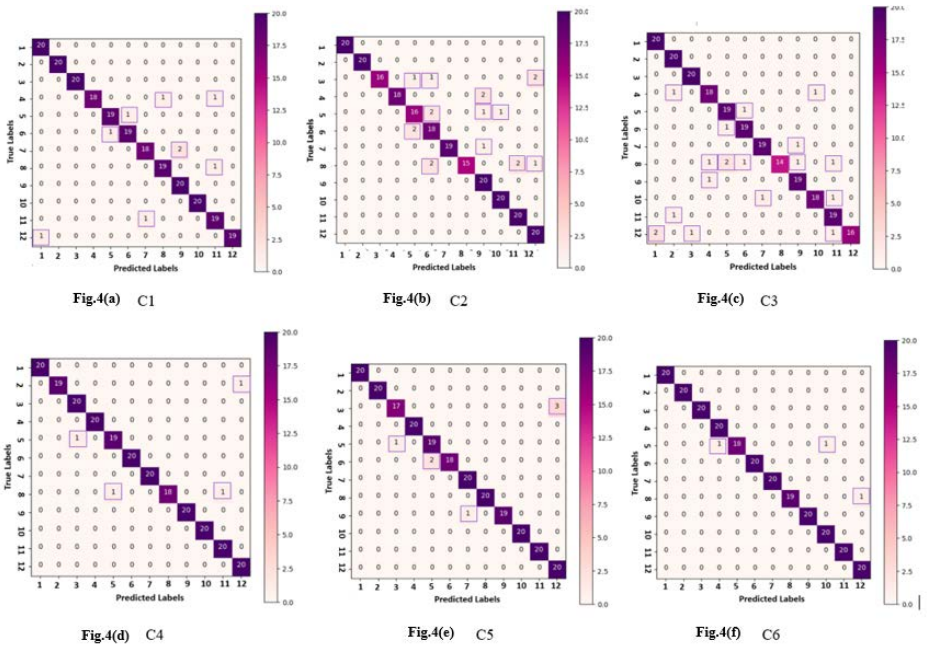


Fig.4(a) C1 Fig.4(b) C2 Fig.4(c) C3

Fig.4(d) C4 Fig.4(e) C5 Fig.4(f) C6

الشكل 4. مصفوفة الإرباك "Confusion Matrix" C1 و C2 و C3 و C4 و C5 و C6

من ناحية أخرى، يبدو أن التصنيف يسير بشكل أفضل مع محسنات (M-SGD) و (Adagrad) لأننا نلاحظ استمرارية في الاتجاه الصحيح على كل من منحنيات الدقة والفاقد (Accuracy and Loss). ومع ذلك، فإن (VGG-v16) و (VGG-v19) يقدمان نتائج أقل بكثير من النماذج المتبقية مع كل من (M-SGD) و(Adagrad). ولا تتحرك منحنيات الفاقد الخاصة بهما في الاتجاه الصحيح كما أنها تولد قيمًا عالية. وترجع هذه المشكلة إلى عدم الكفاءة عند التعامل مع البيانات التدريبية والبيانات الجديدة الأمر الذي يحدث عندما يكون الواقع أكثر تعقيدًا من النموذج. إن (VGG-v16) و (VGG-v19) بعيدان كل البعد عن تعلم البنية الأساسية للبيانات ولذلك فنحن نستبعد محسن (Adam) والنموذجين (VGG-v16) و (VGG-v19) من تحليلنا المستقبلي.

وبالتالي فإننا نأخذ في الاعتبار النماذج الثلاثة فقط: (MobileNetV2) و(InceptionV3) و (Xception) والمحسنين (M-SGD) و(Adagrad) لاختباراتنا القادمة حتى نختار أفضل حل من بين التركيبات الستة. وللتبسيط سيشار إليها باسم C1 وC2 وC3 وC4 وC5 وC6 لتتوافق على التوالي مع: (MobileNetV2 + M-SGD)، (InceptionV3 + M-SGD)، (Xception + M-SGD)، (MobileNetV2 + Adagrad)، (InceptionV3 + Adagrad) و(Xception + Adagrad).

4.2.2 مصفوفة الخطأ (Confusion Matrix)

نستخدم مصفوفة الخطأ لتحليل مساهمة تكوينات التعلم الانتقالي المذكورة أعلاه C1)، C2، C3، C4، C5، و (C6في التعرف على الكلمات الاثنتي عشرة في لغة الإشارة .وتوضح مصفوفة الخطأ مدى نجاح كل من هذه التكوينات الستة في تصنيف الإشارات .ونظرًا لأن لكل كلمة سماتها الخاصة، يمكن أن يُظهر تكوين معين أداءً أفضل من غيره في التعرف على بعض الإشارات، بينما يتكيف تكوين آخر بشكل أفضل مع إشارات أخرى.

نقوم بتقييم النماذج المختلفة باستخدام مقاييس أساسية مثل الدقة، والتذكر، ومعدل F1، كما هو موضح في الجدول .1تشير مصفوفة الخطأ إلى الكلمات الاثنتي عشرة التالية بأرقام من 1إلى 12:

"السجن"، "القهوة"، "القانون"، "البلدية"، "الانتخابات"، "الشاي"، "الجمعية"، "الرقص"، "المساعدة"، "الحكومة"، "الوزارة"، و"علم النفس ."يُشير مقياس التذكر (recall)إلى قدرة مصنِف معين على تحديد جميع التوقعات الصحيحة

Class	Jail	Coffee	Law	Municipality	Election	Tea	Association	Dance	Help	Governorate	Minist
Pre	0.95	1	1	1	0.95	0.95	0.95	0.95	0.91	1 1 1	0.90
Re	1	1	1	0.90	0.95	0.95	0.90	0.95	1		0.95
F1	0.98	1	1	0.95	0.95	0.95	0.92	0.95	0.95		0.93
Pre	1	1	1	1	0.84	0.78	1	1	0.83	0.95	0.91
Re	1	1	0.80	0.90	0.80	0.90	0.95	0.75	1	1	1
F1	1	1	0.89	0.95	0.82	0.84	0.97	0.86	0.91	0.98	0.95
Pre	0.91	0.91	0.95	0.90	0.86	0.90	0.95	1	0.90	0.95	0.86
Re	1	1	1	0.90	0.95	0.95	0.95	0.70	0.95	0.90	0.95
F1	0.95	0.95	0.98	0.90	0.90	0.93	0.95	0.82	0.93	0.92	0.90
Pre	1	1	0.95	1 1 1	0.95	1	1 1 1	1	1	1 1 1	0.95
Re	1	0.95	1		0.95	1		0.90	1		1
F1	1	0.97	0.98		0.95	1		0.95	1		0.98
Pre	1	1	0.94	1 1 1	0.90	1	0.95	1	1	1 1 1	1 1
Re	1	1	0.85		0.95	0.90	1	1	0.95		1
F1	1	1	0.89		0.93	0.95	0.98	1	0.97		
Pre	1	1	1	0.95	1	1	1 1 1	1	1	0.95	1 1
Re	1	1	1	1	0.90	1		0.95	1	1	1
F1	1	1	1	0.98	0.95	1		0.97	1	0.98	

الجدول 1: مقاييس الأداء لجميع التركيبات في مجموعة الاختبار

وفقًا للجدول 1، فإن النماذج المدربة باستخدام المُحسِّن (M-SGD)في التركيبات C1و C2وC3تنتج أخطاء تصنيف أكثر من تلك المدربة باستخدام المُحسِّن .(Adagrad)يبلغ إجمالي التصنيفات الخاطئة في التركيبات C1و C2وC3و C4و C5وC6على التوالي9 :، 21، 24، 4، 7، و .3من الواضح أن التركيبتين C3 (Xception + M-SGD)و (InceptionV3 + M-SGD) C2تقدمان أسوأ أداء مقارنةً بالتركيبات الأخرى،

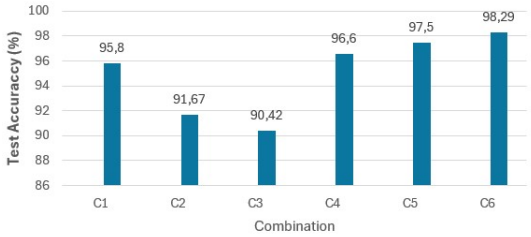
وخاصةً بالنسبة للإشارة "الرقص"، حيث تنخفض قيمة التذكر فيها إلى أقل من .0.75كما نلاحظ تدهورًا ملحوظًا في أداء تصنيف الإشارات "القانون"، "البلدية"، "الانتخابات"، و"الشاي "فيما يتعلق بعملية استخراج الميزات (Feature Extraction)باستخدام التركيبة .C2. تستمر المشكلة نفسها أثناء التدريب باستخدام (Xception + M-SGD) C3، مما يؤدي إلى العديد من التنبؤات غير الصحيحة للإشارات "البلدية"، "الحكومة"، و"علم النفس"، حيث وصلت قيم التذكر الخاصة بها إلى أقل من .0.95ووفقًا للشكل) 3أ(، تواجه المجموعة (MobileNetV2 + M-SGD) C1صعوبات في تصنيف الإشارتين "البلدية "و"الجمعية"، حيث تبلغ قيمة التذكر لهما .0.90

فيما يتعلق بالتركيبة (MobileNetV2 + Adagrad) C4، فهي تعمل بشكل جيد مع جميع الفئات باستثناء فئة "الرقص"، حيث يخطئ النموذج في تصنيفها مرتين كما هو موضح في الشكل) 3د .(بالإضافة إلى ذلك، تظهر نتائج التركيبة (InceptionV3 + Adagrad) C5تقاربًا مع C4من حيث إجمالي التصنيفات الخاطئة .وعلى الرغم من أن C4تقدم أداءً جيدًا في الشكل) 3هـ(، فإن النموذج يقدم تنبؤين غير صحيحين للإشارة "الشاي "وثلاثة تنبؤات غير صحيحة للإشارة "القانون"، حيث كانت قيم التذكر لهما 0.90و 0.85 على التوالي.

يبدو أن التركيبة C6تعمل بكفاءة أكبر، حيث تحتوي على أقل عدد من التنبؤات الخاطئة .ومع ذلك، فإن الإشارة "الانتخابات "ليست مصنفة جيدًا في التركيبة .C6ونظرًا لتعقيد الميزات، يمكن التعرف على كلمة "الانتخابات "بشكل أفضل بواسطة التركيبات C1، وC3، وC4، وC5

بناءً على التحليل أعلاه، قررنا استبعاد التركيبة C2وC3، والاحتفاظ بالتركيبات C1وC4وC5وC6لتصميم بنية الشبكة التالية .وللتحقق من صحة اختيارنا، نشير إلى الشكل 5الذي يوضح قيم دقة الاختبار (Test Accuracy)لكل تركيبة .تُظهر C1وC4وC5وC6قيمًا متقاربة لدقة الاختبار (95.8٪، 96.60٪، 97.5٪، و98.2٪ على التوالي(، مع اختلاف طفيف بينها .في المقابل، تُسجل C2وC3دقة اختبار تبلغ 91.67٪ و90.42٪ على التوالي، وهي بعيدة عن المتوسط .نظرًا لأنه يتعين علينا اختيار حل واحد من التركيبات الأربع المختارة، فإننا نناقش في القسم التالي أي نموذج وأي مُحسِّن يناسب الحالة المعطاة بشكل

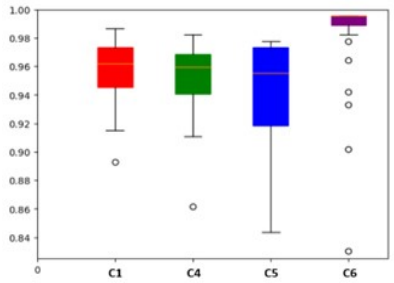
أفضل . ونظرًا لأنه يتعين علينا اختيار حل واحد من التركيبات الأربع المختارة فإننا نناقش في القسم التالي أي نموذج وأي مُحسِّن يناسب الحالة المعطاة بشكل أفضل.



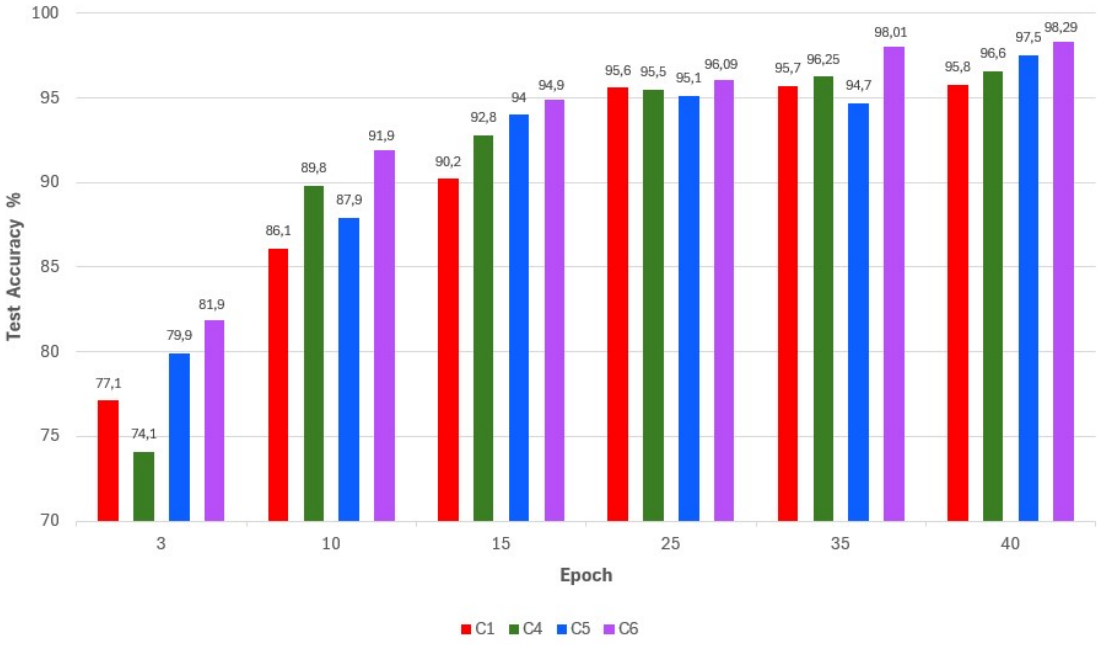
الشكل 5. دقة الاختبار لـ C1 وC2 وC3 وC4 وC5 وC6

4.3 اختيار النموذج

لأن الفروقات في الدقة والتنبؤات غير الصحيحة بين التركيبات الأربعة ليست كبيرة، فإننا بحاجة إلى مؤشرات إحصائية إضافية لإظهار أيهما أكثر أهلية لتصنيف لغة الإشارة التونسية وفي هذا السياق، يوضح . توزيع درجات دقة الاختبار لكل تركيبة 7و 6الشكلان الرسم البياني في الشكل . وبالرجوع إلى 6، نلاحظ أن انتشار درجات دقة الاختبار يضيق بشكل كبير عند التدريب باستخدام C6 (Xception + Adagrad). وعلى الرغم من أن C5 (InceptionV3 + Adagrad) يُظهر عددًا قريبًا من التصنيفات الخاطئة وقيمة دقة مماثلة لـ C1، إلا أنه يظهر تباينًا ملحوظًا في النتائج(MobileNetV2 + M-SGD).



الشكل 6: رسم بياني يوضح دقة الاختبار لـ C1 و C4 و C5 وC6



الشكل 7: الاختلافات في دقة الاختبار بين C1 وC4 وC5 وC6 عبر عمليات المرور المحلية على كامل بيانات التدريب (epochs)

ظهر نموذج InceptionV3 معدل اضطراب وعدم استقرار مرتفعين في التعرف على البيانات الجديدة، مما يجعله غير قادر على تعلم المشكلة بفعالية .تُظهر تركيبة C6 نطاق انتشار أقل مقارنةً بـ C1 وC4، على الرغم من وجود بعض القيم المتطرفة غير المؤثرة في خطها الرأسي .هذه القيم القليلة لا تؤثر بما يكفي لتُأخذ في الاعتبار ضمن عملية التقييم.

ويوضح الشكل 7 الرسم البياني الذي يعرض دقة الاختبار لكل تركيبة عند قيم مختلفة لدورات التدريب (epochs) (3، 10، 15، 35، 40)، حيث تُظهر C6 تفوقًا مستمرًا على التركيبات المتبقية (C1، C4، و C5) في كل تكرار للتدريب .تثبت المحاكاة في الشكلين 6 و7 أن نموذج Xception يعمل بشكل أفضل من النماذج الأخرى عند دمجه مع المحسن (Adagrad)، حيث يحقق أفضل معدل دقة، وهو حوالي 98.29٪.

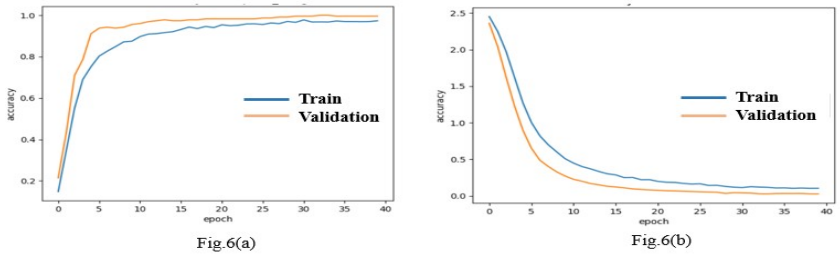


Fig.6(a) Fig.6(b)

الشكل 8. منحنيات دقة التدريب/التحقق وفاقد التدريب/التحقق في Xception

ولإثبات أن هذه القيمة ليست عشوائية، قمنا بتكرار عملية التدريب ست مرات، وجمعنا المقاييس ذات الصلة لكل خطوة في الجدول .2من الواضح أن هناك تقاربًا في القيم التي تم الحصول عليها، مما يوضح استقرار المجموعة C6أثناء عملية التنبؤ .في الوقت نفسه، يثبت الشكل 8، الذي يصور منحنيات الدقة والفاقد لمجموعات التدريب والتحقق، كفاءة هذا النموذج .(Xception + Adagrad)

ومع ذلك، يواجه هذا النموذج صعوبات في تفسير إشارة "الانتخابات"، وفقًا لمصفوفة الخطأ في الشكل (f) 4، حيث تختلط فئة "الانتخابات "مع فئتي "البلدية "و"الحكومة ."وقد يكون هذا نتيجة لتقنيات تعزيز البيانات المطبقة عبر الإنترنت أثناء عملية التدريب، مما يؤدي إلى التشابه بين التمثيلات المجردة والميزات التي تعلمها النموذج العصبي الالتفافي .(CNN)

المحاولة رقم	1	2	3	4	5	6
الدقة (%)	98.295	98.281	98.287	98.291	98.304	98.286

الجدول 2: الخطوات الستة لقياس أداء (Xception + Adagrad) C6 لتصنيف إشارات لغة الإشارة التونسية

5. الخاتمة

توضح هذه الدراسة الإمكانات الواعدة لاستخدام التعلم الانتقالي في التعرف على لغة الإشارة التونسية . تم تطبيق طريقتنا على مجموعة بيانات لغة الإشارة التونسية (TnSL)المكونة من 2000صورة والمجهزة بتقنية تعزيز البيانات .حقق نموذج Xceptionأفضل دقة اختبار بنسبة 98.29٪ عند دمجه مع المُحسِّن (Adagrad)في التعرف على الإشارات غير المتماثلة الثابتة باليدين ضمن مجموعة بيانات صغيرة الحجم . يمثل هذا البحث خطوة أساسية نحو تطوير نظام للتعرف على لغة الإشارة التونسية، يمكن أن يسهم في خدمة مجتمع الصم التونسي في المواقف اليومية، ويخفف من حواجز التواصل.

سيركز العمل المستقبلي على توسيع مجموعة البيانات وتطوير أنظمة للتعرف الديناميكي على الإشارات . ويجب توسيع مجموعة البيانات لتشمل المزيد من إشارات لغة الإشارة التونسية، مما يتيح التفسير الديناميكي للجمل

المراجع

[1] Othman, A., Dhouib, A., Chalghoumi, H., Elghoul, O., and Al-Mutawaa, A. (2024). The acceptance of culturally adapted signing avatars among deaf and hard-of-hearing individuals. IEEE Access.

[2] Rastgoo, R., Kiani, K., and Escalera, S. (2021). Sign language recognition: A deep survey. Expert Systems with Applications, 164:113794.

[3] Töngi, R. (2021). Application of transfer learning to sign language recognition using an inflated 3d deep convolutional neural network. arXiv preprint arXiv:2103.05111.

[4] Schmalz, V. J. (2022). Real-time Italian sign language recognition with deep learning. In CEUR Workshop Proceedings.

[5] Nefaa, A. (2023). Genetic relatedness of Tunisian sign language and french sign language. Frontiers in Communication.

[6] Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., and Azim, M. A. (2022). Transfer learning: a friendly introduction. Journal of Big Data.

[7] Chavan, A., Bane, J., Chokshi, V., and Ambawade, D. (2022). Indian sign language recognition using Mobilenet. In 2022 IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI).

[8] Zakariah, M., Alotaibi, Y. A., Koundal, D., Guo, Y., and Mamun Elahi, M. (2022). Sign language recognition for arabic alphabets using transfer learning technique. Computational Intelligence and Neuroscience, 2022(1):4567989.

[9] Buckley, N., Sherrett, L., and Secco, E. L. (2021). A CNN sign language recognition system with single & double-handed gestures. In 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), pages 1250–1253. IEEE.

[10] Hossen, M., Govindaiah, A., Sultana, S., and Bhuiyan, A. (2018). Bengali sign language recognition using deep convolutional neural network. In 2018 joint 7th international conference on informatics, electronics & vision (iciev) and 2018 2nd international conference on imaging, vision & pattern recognition (icIVPR).

[11] Das, P., Ahmed, T., and Ali, M. F. (2020). Static hand gesture recognition for American sign language using deep convolutional neural network. In 2020 IEEE region 10 symposium (TENSYMP), pages 1762–1765. IEEE.

[12] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255.

[13] Plested, J. and Gedeon, T. (2022). Deep transfer learning for image classification: a survey. arXiv preprint arXiv:2205.09904.