التعرف على لغة الإشارة العربية في الوقت الحقيقي باستخدام التعلم العميق

دراسة أُعدّت لنيل درجة الماجستير في الهندسة المعلوماتيّة باختصاص هندسة البرمجيات ونظم المعلومات

إعداد
م. ناديا قسيس

إشراف:

أ.د. رانيا لطفي
مشرفاً أساسياً

د. إيفا حريقص
مشرفاً مشاركاً

مُلخَّص البحْث

تُعتبر لغة الإشارة الطريقة الأكثر تعبيراً وفاعليةً لتمكين ضعاف السمع من الانغماس في الأنشطة الاجتماعية وتحسين قدرات التواصل الاجتماعي لديهم، إلا أن أنظمة التعرف على لغة الإشارة لاسيما لغة الإشارة العربية تواجه عدة تحديات تجعل دقة التعرف الخاصة بها غير جيدة.

يهدف هذا البحث إلى تطوير نظام للتعرف على لغة الإشارة العربية باستخدام تقنيات التعلم العميق، وعلى وجه التحديد بنية الشبكات العصبية التلافيفية CNN. يهدف النظام إلى سد

فجوة التواصل للأشخاص ضعاف السمع مع محيطهم من خلال ترجمة إيماءات لغة الإشارة العربية تلقائياً إلى نص مكتوب، حيث تم اختبار أداء أربعة بنى معمارية من نماذج CNN وهي: نموذج CNN مخصص، VGG-16، MobileNet V1، MobileNet V2 على مجموعتي بيانات للغة الإشارة العربية إحداهما تم بناؤها يدوياً ومعترف بها من قبل جمعية رعاية الصم والبكم في مدينة حمص. أظهرت النتائج أن نموذج MobileNet V2 حقق أعلى دقة بلغت 96% على مجموعة البيانات المبنية، مع إمكانية تضمين هذا النموذج ضمن تطبيق على منصات الويب والأجهزة المحمولة لترجمة لغة الإشارة العربية لنص في الوقت الحقيقي.

الكلمات المفتاحيّة: التعرف على لغة الإشارة العربية، التعلم العميق، التعلم بالنقل، الشبكات العصبية التلافيفية (CNNs).

Real-Time Arabic Sign Language Recognition Using Deep Learning

Abstract

Sign language is considered the most expressive and effective way for enabling the hearing impaired to engage in social activities and improve their social communication skills. However, sign language recognition systems, especially Arabic sign language, face several challenges that hinder their recognition accuracy.

This research aims to develop a system for recognizing Arabic sign language using deep learning techniques, specifically convolutional

neural network (CNN) architecture. The system aims to bridge the communication gap between hearing impaired people and their surroundings by automatically translating Arabic sign language gestures into written text. The performance of four CNN architectures was tested: a custom CNN model, VGG-16, MobileNet V1, and MobileNet V2 on two datasets for Arabic sign language, one of which was manually built and authenticated by the Deaf and Dumb Care Association in Homs. The results showed that the MobileNet V2 model achieved the highest accuracy of 96% on the manually constructed dataset, with the potential to integrate this model into web and mobile applications for real-time translation of Arabic sign language into written text.

Keywords: Arabic Sign Language Recognition (ASLR), Deep Learning, Transfer Learning, Convolutional Neural Networks (CNN).

مقدّمة

تعد اللغة المنطوقة أحد وسائل التواصل بين البشر، لكن بالنسبة للأشخاص الذين يعانون من مشكلة في السمع والنطق فقد تم تطوير نظام تواصل بصري يسمى لغة الإشارة والذي يستخدم إيماءات محددة لترجمة التعابير إلى لغة لفظية. وفقاً لمنظمة الصحة العالمية عام 2020، هناك حوالي 466 مليون شخص يعانون من مشاكل فقدان السمع، 34 مليون منهم من الأطفال [1]، لذا تُعتبر لغة الإشارة وسيلة هامة للتواصل بين الصم والبكم أنفسهم ومع الناس العاديين للتعبير عن أفكارهم ومشاعرهم.

مع وجود هذه الطريقة البديلة للتواصل بين الصم والبكم يبقى من الصعب جداً تواصلهم مع عامة الناس، لأنه على الرغم من إتقان العديد من الأشخاص ضعاف السمع للغة الإشارة، إلاّ أنّ القليل جداً من الأشخاص العاديين بإمكانهم فهمها واستخدامها كونها تحتاج إلى الكثير من التدريب والممارسة. نتيجة لذلك، فإن هذه المشكلة تخلق الكثير من العزلة

الاجتماعية بين الأشخاص ذوي الإعاقة السمعية والنطقية في المجتمع الذي يعيشون فيه، ومن هنا ظهرت الحاجة لنظام يسمح بترجمة لغة الإشارة تلقائياً إلى نص والعكس بالعكس.

لغة الإشارة ليست ترجمة مرئية للغة المنطوقة، إنما لديها القواعد الخاصة بها. وهي من اللغات التي تستخدم فيها حركة اليدين بإشارات محددة تؤثر على معنى الإيماءات، وتكون الإشارات تبعاً لكل حرف من الحروف الأبجدية، ومن خلالها يمكن تكوين جمل. كما أنه لا يوجد توحيد بشأن لغة الإشارة التي يجب اتباعها على مستوى العالم، حيث أنها تختلف باختلاف الأشخاص والمنطقة التي ينتمون إليها.

تم تقديم لغة الإشارة العربية (اللغة المستخدمة في المناطق العربية) ArSL (Arabic Sign Language) رسمياً عام 2001 من قبل الاتحاد العربي للصم AFOD (Arab Federation of the Deaf)، وهي تتضمن مفردات وقواعد خاصة بها، مما يجعلها نظاماً مستقلاً للتواصل. كما يوجد العديد من اللهجات المختلفة لـ ArSL والتي تختلف من بلد إلى آخر، لكنها بالعموم مكونة من 28 حرفاً تتفق عليها اللهجات المختلفة.

هناك نوعان رئيسيان من لغات الإشارة التي يستخدمها الصم والبكم:

- **لغة الإشارة التي تعتمد على الكلمات (الكلمات والإيماءات الكاملة):** يعتمد هذه النوع على إشارات يدوية وحركات تعبر عن كلمات أو مفاهيم كاملة، مثل لغة الإشارة الأمريكية ASL (American Sign Language) أو لغة الإشارة البريطانية BSL (British Sign Language) أو لغة الإشارة العربية ArSL.
- **لغة الإشارة التي تعتمد على الأحرف (أبجدية الأصابع):** يستخدم هذا النوع حركات يدوية لتمثيل الحروف الأبجدية، حيث يتم تهجئة الكلمات حرفاً بحرف، مثل أبجدية الأصابع الإنجليزية (Fingerspelling) المستخدمة في لغة الإشارة الأمريكية أو أي لغة إشارة أخرى.

بما أن قاموس الاشارات المستخدمة يومياً ضخم وقد يصل إلى الآلاف من الكلمات فإن المهمة في التعرف على الإيماءات محدودة في عدد قليل من الكلمات، مما يشكل تحدي كبير في قابلية التوسع في التعرف على لغة الإشارة [2]. لذلك تم الاعتماد في هذا البحث على الاحرف الأبجدية للغة العربية فقط والتي من خلالها يمكن تشكيل أي جملة.

في الوقت الحالي، هناك العديد من الدراسات المتعلقة بالتعرف على لغة الإشارة (SLR) التي تهدف إلى تقليص الفجوة بين الأشخاص الصم والأشخاص العاديين، حيث يمكن لهذه التقنيات أن تحل محل الحاجة إلى المترجمين. ومع ذلك، تواجه أنظمة التعرف على الإشارات تحديات عديدة، مثل الدقة المنخفضة، الإيماءات المعقدة، الضوضاء العالية، والقدرة على العمل في ظروف متغيرة مع الحفاظ على التعميم أو التقييد ضمن هذه الظروف، ونظراً لأن كل لغة لها إشاراتها الخاصة، فإن تغطية جميع إشارات اللغات يُعد تحدياً كبيراً. يمثل الانتشار الواسع للهواتف المحمولة فرصة للأشخاص ذوي الإعاقة السمعية للمشاركة بشكل أكبر في مجتمعاتهم، لذا فإن تصميم وتنفيذ نظام للتعرف على لغة الإشارة العربية (ArSL) يمكن أن يؤثر بشكل كبير على جودة حياتهم.

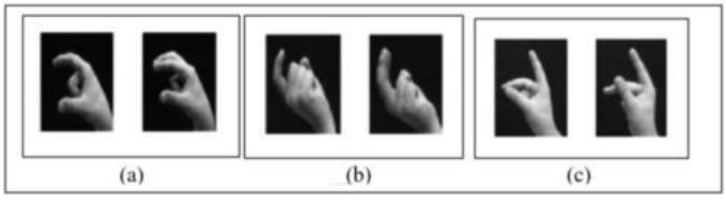
1. مشكلة البحث

هناك الكثير من الأبحاث والدراسات التي تهتم بالتعرف التلقائي على لغة الإشارة من أجل اللغة الانكليزية، في حين أن هناك ندرة بالأبحاث المتعلقة بالتعرف على لغة الإشارة العربية، كما تكمن صعوبة هذا البحث في ناحيتين:

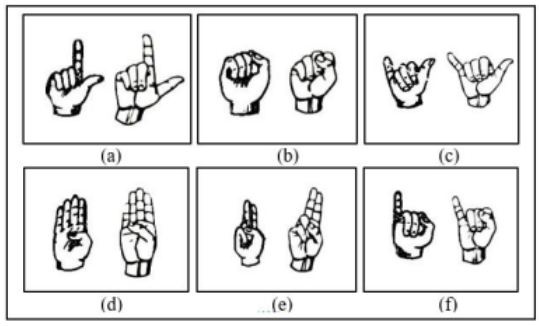
- تنوع طرق الحصول على البيانات
- تنوع الطرق المستخدمة للتعرف على لغة الإشارة

هناك العديد من الأسباب التي تجعل التعرف على لغة الإشارة مجال بحثي صعب أهمها:

- يعد تفسير مجموعة المعلومات المرئية على أنها جمل لغة طبيعية أحد التحديات الصعبة في حل مشكلة ترجمة لغة الإشارة، لأنه من أجل تفسير جملة واحدة علينا فهم الاختلاف الدقيق في شكل وحركة اليدين والجسد وأحياناً تعابير الوجه.
- يعد تجزئة المنطقة التي تشير إلى الايماءة من الصورة بأكملها مشكلة أيضاً بسبب البيئات المختلفة التي يتم فيها التقاط الصور، والإضاءة، حجم اليد، وتعابير الوجه، والاتجاه وما إلى ذلك. حيث ستؤدي التغييرات الديناميكية في الإضاءة والخلفية ومواضع الكاميرا إلى صعوبة استخراج المناطق المهمة من الإطار بأكمله كون الحركة في الخلفية لا يمكن التحكم بها، كما أن الكشف المعتمد على الجلد ليس أسلوباً دقيقاً لأنه لا يمكن استخراج منطقة الجلد بناءً على عتبة غير معروفة بسبب اختلاف ألوان البشرة.
- تتكون كل لغة إشارة من آلاف الاشارات التي تختلف عن بعضها باختلافات طفيفة في شكل اليد وحركتها وموضعها، وهذه المشكلة موجودة بشكل كبير في لغة الإشارة العربية حيث يوجد عدة أحرف تتشابه فيهم حركة اليد مع اختلاف بسيط في الأصابع مما يخلق غموضاً في مهمة تصنيف الصور كما هو موضح في الشكل(1) الذي يمثل إيماءات متماثلة في ArSL. كما يمكن أن تتشابه أيضاً إشارات أحرف اللغة العربية مع إشارات أحرف اللغة الأميركية كما هو موضح في الشكل(2).



الشكل (1): إيماءات متماثلة في ArSL – (a) التشابه بين حرفي (د) و(ذ)، (b) التشابه بين حرفي (ز) و(ر)، (c) التشابه بين حرفي (ظ) و(ط)



الشكل (2): إيماءات متماثلة بين ASL وArSL – (a) "ل" و "L" (b) "ص" و "S" (c) "ي" و "y" (d) "ك" و "B" (e) "ت" و "H" (f) "م" و "I"

- يتمثل التحدي الرئيسي في SLR في تصميم واستخراج الواصفات المرئية التي يمكنها تحديد الإشارات وتصنيفها بشكل موثوق، واختلاف نتائج الأنظمة باختلاف مجال الدراسة، وضبط البارمترات الفائقة الخاصة بهذه الأنظمة بما يتناسب مع حجم وطبيعة البيانات المستخدمة.

لاتزال الأساليب الحالية المتبعة للتنبؤ بلغة الإشارة للأحرف العربية تعاني من بعض الصعوبات والقصور في عدة جوانب، لذا لابد من العمل على محاولة إيجاد حل للتحديات السابقة من أجل بناء نموذج للتعرف على لغة الإشارة العربية بدقة عالية بحيث يساهم في تحسين عملية التواصل بين الصم والبكم والأشخاص العاديين.

2. الهدف من البحث وأهمّيته

يهدف البحث بشكل أساسي إلى تطوير نموذج للتعرف على لغة الإشارة العربية من خلال استخدام تقنيات التعلم العميق ونقل التعلّم للوصول إلى أفضل دقة ممكنة ضمن المجال المدروس، وفي النهاية تضمين هذا النموذج في تطبيق ويب يساعد على تحويل الإشارات إلى نص باللغة العربية لمساعدة الصم والبكم في التواصل مع الآخرين. ويتم ذلك من خلال تحقيق الآتي:

- دراسة تحليلية للأبحاث الموجودة التي تعمل على التعرف على لغة الإشارة العربية.
- بناء مجموعة بيانات خاصة بلغة الإشارة العربية.
- بناء نموذج قادر على التنبؤ بأحرف لغة الإشارة العربية بأفضل دقة ممكنة بالاستفادة من تقنيات التعلم العميق واستخدام النماذج المدربة مسبقاً في تحقيق هذا النموذج.
- مقارنة نتائج هذه الدراسة مع دراسات مشابهة.
- تطوير واجهة تخاطبية للنموذج الذي تم التوصل له لتوضيح إمكانية استثماره في التطبيقات الحقيقية في تحويل لغة الإشارة العربية إلى نص مكتوب، بحيث يتمكن الأشخاص الصم من فتح الكاميرا وإدخال حركات الأحرف بالوقت الحقيقي ليتم تحويلها إلى أحرف مكتوبة وتشكيل كلمات وجمل منها يمكن قراءتها من قبل المستخدم العادي.

تكمن أهمية هذا البحث من حيث كونه يُعنى بأحرف لغة الإشارة العربية من جهة مع توفير مجموعة بيانات خاصة بها معترف بها من قبل جمعية رعاية الصم والبكم في مدينة حمص يمكن الاستفادة منها في العديد من الدراسات والأبحاث المستقبلية ضمن هذا المجال، وبكونه يستخدم أحدث تقنيات التعلم العميق واستثمار النماذج المدربة مسبقاً في بناء نموذج قادر على التنبؤ بأحرف لغة الإشارة العربية بأفضل دقة ممكن، واستخدام هذا النموذج ضمن تطبيق ويب والذي من خلاله يمكن توفير وسائل لتدريس لغة الإشارة العربية القياسية في مراكز الصم والبكم واعتماده كوسيلة لتواصل الصم والبكم مع محيطهم في العديد من المجالات مثل المطاعم والمستشفيات ووسائل النقل وغيرها. يمكن للأشخاص العاديين أيضاً من خلال التطبيق تعلم لغة الإشارة العربية القياسية بسهولة أكبر واستخدام ماتعلموه للتواصل مع الصم والبكم، مما يجعل اندماجهم أفضل مع كافة نواحي المجتمع.

3. دراسات ذات صلة

قام الباحثون في الدراسة [3] بتطوير نظام للتعرف التلقائي على لغة الإشارة العربية يستخدم نموذج CNN للتعرف على 28 حرفاً عربياً باستخدام مجموعة بيانات تسمى ArSL2018 تتضمن أحرف لغة الإشارة العربية على شكل صور ملونة RGB. تتناول الدراسة مشكلة التواصل المحدود بين مجتمع الصم والبكم ومجتمع الأشخاص الطبيعين بسبب نقص المترجمين والاستخدام المحدود للغة الإشارة في مجتمع الصم، مما يؤدي إلى انخفاض فرص التعليم والعمل للأفراد الصم، لذا تم اقتراح إطار عمل جديد قائم على نموذج CNN، تم اختباره على (10810 صورة) والتي تمثل حوالي 20% من مجموعة البيانات، وقياس دقة النموذج وكذلك معدل الخطأ خلال مراحل التدريب والاختبار، وحقق النموذج دقة تعرف بلغت 92.9%.

قام الباحثون في الدراسة [4] من أجل تسهيل التواصل للأفراد ذوي الإعاقات السمعية بتطوير نظام آلي للتعرف على لغة الإشارة العربية وتحويلها إلى نص باللغة العربية. تم بناء نظام التعرف باستخدام نموذج التعلم العميق CNN المستخدم على نطاق واسع في مهام التعرف على الصور، حيث يقوم النظام بالتقاط صور لإشارات اليد، والتعرف على أحرف معينة من أبجدية لغة الإشارة العربية، ثم توليد نص مطابق باللغة العربية. لضمان قوة النموذج، شملت مجموعة البيانات تنوعات في ظروف الإضاءة، الزوايا، والمسافات لتغطي 31 حرفاً من لغة الإشارة العربية. حققت الدراسة معدل دقة تعرّف بنسبة 90%، وتم اقتراح كعمل مستقبلي دمج أجهزة متقدمة للتعرف على الإيماءات، مثل Leap Motion أو Xbox Kinect وتوسيع مجموعة البيانات.

اقترح الباحثون في الدراسة [5] نظاماً يعتمد على تقنية التعلم الآلي باستخدام الشبكة العصبية الالتفافية (CNN)، والتي تستفيد من أجهزة استشعار قابلة للارتداء للتعرف على 30 حرفاً من لغة الإشارة العربية. يتميز النموذج المقترح بقدرته على التكيف مع جميع الإيماءات المحلية العربية التي يستخدمها أفراد المجتمع العربي من ضعاف السمع. تم

استخدام نهج CNN لأغراض التصنيف، حيث تكون إيماءات اليد الخاصة بلغة الإشارة العربية هي المدخلات، بينما يكون الصوت المنطوق هو المخرج للنظام المقترح. في البداية، تم بناء شبكة CNN لاستخراج الميزات من البيانات التي تم جمعها بواسطة قفازات DG5-V المزودة بأجهزة استشعار قابلة للارتداء لالتقاط حركات اليد في مجموعة البيانات، من ثم تم معالجة البيانات المجمعة من هذه المستشعرات لتنسيق الصور، وتقسيمها إلى 80% مجموعة تدريب و20% مجموعة اختبار. حقق النموذج المقترح دقة تعرّف 90% وكان فعالاً للغاية في ترجمة إيماءات اليد الخاصة بلغة الإشارة العربية إلى كلام وكتابة.

سعى الباحثون في الدراسة [6] إلى تطوير نظام فعال للتعرف على لغة الإشارة العربية (ArSL) يجمع بين أداء التصنيف العالي وكفاءة التكلفة الحسابية، مما يجعله مناسباً للاستخدام على الأجهزة المحمولة. لتحقيق هذا الهدف، اعتمد الباحثون على بنية خفيفة تعتمد على نماذج EfficientNet، المعروفة بكفاءتها من حيث عدد المعلمات وتكلفة الحساب، كما استخدموا مجموعة بيانات واقعية تحتوي على صور لإيماءات اليد تمثل ثلاثين حرفاً مختلفاً من الأبجدية العربية، وقاموا بتطبيق وتقييم نماذج EfficientNet المتنوعة وتقنيات المعالجة المسبقة بهدف تحسين الأداء. أسفرت التجارب عن تحقيق أفضل النتائج باستخدام بنية EfficientNet-Lite 0 مع تقنية Label Smooth كدالة خسارة (Loss Function)، مما أدى إلى دقة بلغت 94%. كما أثبت هذا النهج فعاليته في التعامل مع الخلفيات المتنوعة. تناولت منهجية الدراسة أربعة سيناريوهات تجريبية، تضمنت دراسة تأثير تقسيم الصورة على أداء النموذج، ومقارنات بين إصدارات مختلفة من نماذج EfficientNet Lite، وتقنيات التحسين. أشارت النتائج إلى أن تقسيم الصورة كان له دور محوري في تحسين دقة النموذج، حيث حقق النموذج الأخف أفضل أداء. خلصت الدراسة إلى أن النماذج الحالية للتعرف على ArSL غالباً ما تعتمد على تقنيات معقدة ومرهقة من حيث الموارد، بينما يظهر النظام المقترح، المبني على EfficientNet،

تحسينات كبيرة من حيث الكفاءة والدقة. وتقترح الدراسة في المستقبل استكشاف استخدام نماذج قائمة على المحوّل Transformer لمهام الرؤية وتوسيع النظام ليشمل التعرف على الكلمات والتعابير الشائعة في لغة الإشارة العربية.

قام الباحثون في [7] بإنشاء نظام للتعرف على لغة الإشارة العربية ((ArSLR) يركز على التعرف على الحروف العربية. تم تطوير نظام ArSLR باستخدام طريقة خاصة لمعالجة الصور لاستخراج الموقع الدقيق لليد، بالإضافة إلى اقتراح هيكلية لشبكة عصبية تلافيفية عميقة (CNN). يهدف هذا النظام إلى تسهيل تفاعل الأشخاص الذين يعانون من مشاكل سمعية مع الأشخاص العاديين. بناءً على مدخلات المستخدم، يمكن للنظام اكتشاف والتعرف تلقائيًا على حروف الأبجدية العربية التي تُعبّر بالإشارة. حقق النظام دقة تبلغ 97.07% في التعرف على حروف لغة الإشارة العربية. أشارت النتائج إلى أن الطريقة المقترحة تتفوق على الدراسات المماثلة في التعرف على الإشارات الثابتة باستخدام نفس مجموعة البيانات. واقترحوا كعمل مستقبلي توسيع مجموعة البيانات وتحسين تقنيات تحويل النصوص باستخدام معالجة اللغة الطبيعية.

في الدراسة [8]، سعى الباحثون إلى تطوير نماذج قوية للتعلم بالنقل، مدربة على مجموعة بيانات تحتوي على 54,049 صورة تمثل 32 حرفاً من الأبجدية العربية بلغة الإشارة، بهدف تصنيف هذه الصور بدقة عالية إلى الحروف العربية المقابلة. اعتمدت الدراسة على نهجين: الأول تمثل في استخدام التعلم بالنقل، حيث تم تطبيق عدة نماذج مدربة مسبقاً مثل MobileNet، Xception، Inception، InceptionResNet، DenseNet وBiT، إضافة إلى محولي رؤية هما ViT و Swin. وقد تم تقييم إصدارات مختلفة من هذه النماذج باستخدام أوزان مهيأة من مجموعة بيانات ImageNet أو بشكل عشوائي. النهج الثاني تمثل في استخدام التعلم العميق عبر الشبكات العصبية الالتفافية (CNN)، حيث تم تدريب عدة نماذج من الصفر ومقارنتها بنتائج منهجية التعلم بالنقل. تم تقييم

النماذج باستخدام معايير متعددة تشمل الدقة (Accuracy)، المساحة تحت المنحني (AUC)، الدقة (Precision)، الاسترجاع (Recall)، F1−score، والخسارة. أظهرت النتائج أن منهجية التعلم بالنقل حققت أداءً ممتازاً على مجموعة بيانات ArSL، متفوقة على نماذج CNN الأخرى، حيث حققت نماذج ResNet و InceptionResNet نسبة دقة بلغت 98%. كما أظهر استخدام محولي ViT و Swin، من خلال دمج بنية المحولات مع التدريب المسبق، فعالية كبيرة في تقليل عدد المعلمات المطلوبة للتدريب، مما جعلها أكثر كفاءة واستقراراً مقارنة بالنماذج الأخرى والدراسات السابقة في تصنيف ArSL. يبرز هذا البحث فعالية وقوة استخدام التعلم بالنقل مع محولات الرؤية للتعرف على لغة الإشارة في اللغات التي تعاني من نقص في الموارد.

في الدراسة [9] تم تطوير نظام للتعرف على لغة الإشارة العربية باستخدام تقنيات التعلم العميق، وتم استخدام إحدى شبكات الـ CNN المدربة مسبقاً وخفيفة الوزن تدعى MobileNet V2 للتنبؤ بـ 32 حرف مختلف من أحرف اللغة العربية. تم تدريب النموذج على مجموعة البيانات تدعى ArSL21L حيث تم تقسيم الصور إلى 9955 صورة للتدريب و4247 صورة للاختبار وبحجم صور 416*416*3 لذلك تم تعديل حجم الصور في مجموعة البيانات قبل تدريب النموذج عليها لتصبح 224*224*3 حتى تتوافق مع دخل النموذج، وتم التدريب على 60 epochs . أعطى النموذج دقة تعرف جيدة حيث كانت قيمة مقياس F1−score تساوي 80%و قيمة مقياس Precision تساوي 80% ومقياس Recall تساوي 81%.

4. مواد وطرائق البحث

1.5 مجموعات البيانات (datasets) المستخدمة

قمنا باستخدام مجموعتي بيانات لأحرف لغة الإشارة العربية إحداها تم جمعها يدوياً باستخدام كاميرا موبايل نتيجة صعوبة إيجاد مجموعة بيانات باللغة العربية ملائمة لمتطلبات جميع

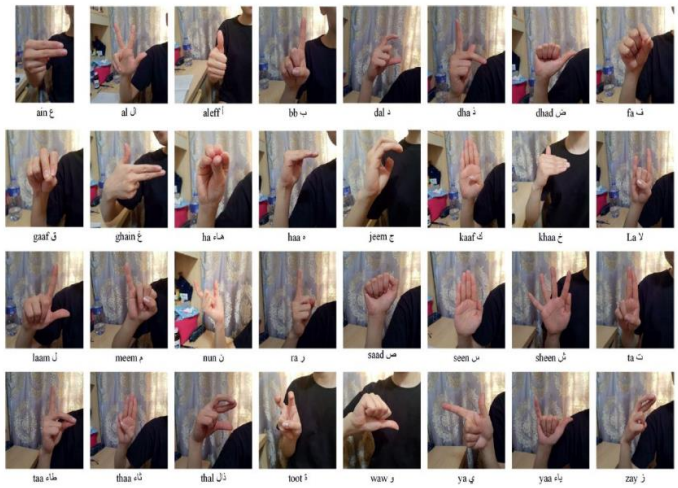
النماذج التي تم العمل عليها وضبطها ضمن التجارب العملية لهذه الدراسة، وقد تم الحصول على تأكيد جمعية رعاية الصم والبكم في مدينة حمص حول صحة مجموعة البيانات هذه بحيث يمكن استخدامها في الدراسة.

تتألف مجموعة البيانات الأولى من صور لـ 28 حرف من أبجدية لغة الإشارة العربية، بالاستعانة بـ 10 أشخاص لتمثيل صور الإشارات لكل حرف، حيث تم جمع 500 صورة مختلفة لكل حرف بحجم 3*224*224 ولاحقة 'jpg' تراعي اختلاف حجم اليد وموضعها، ظروف تغيير الإضاءة والخلفيات وذلك للحصول على مجموعة بيانات أكثر شمولية. هذا نتج عنه حوالي 13440 صورة ككل. تم تقسيم مجموعة البيانات هذه بحيث يكون 80% من البيانات للتدريب و20% للاختبار. يوضح الشكل (3) مثالاً عن الصور التي تم جمعها لكل حرف من أحرف لغة الإشارة العربية في مجموعة البيانات الأولى.

الشكل (3): عينة من مجموعة البيانات الأولى

تتألف مجموعة البيانات الثانية من صور لـ 32 حرف من أبجدية لغة الإشارة العربية وهي مستخدمة في الدراسة [9]، حيث تم جمع 500 صورة مختلفة لكل حرف بحجم 3*416*416 ولاحقة 'jpg' تراعي اختلاف حجم اليد وموضعها، ظروف تغيير الإضاءة. هذا نتج عنه حوالي 17448 صورة ككل. تم تقسيم مجموعة البيانات هذه أيضاً

بحيث يكون 80% من البيانات للتدريب و20% للاختبار. يوضح الشكل (4) مثالاً عن الصور لكل حرف من أحرف لغة الإشارة العربية في مجموعة البيانات الثانية [9].



الشكل (4): عينة من مجموعة البيانات الثانية [9]

2.5 مقاييس الأداء المستخدمة

تم الاعتماد على 4 مقاييس لتقييم أداء النماذج في التجارب العملية التي تم القيام بها في هذا البحث والتي يتم تحديدها عموماً بالاستعانة بمصفوفة الارتباك (Confusion matrix) لمسألة تصنيف ثنائية. هذه المقاييس هي [10]:

- **مقياس الدقة Accuracy:** يمثل هذا المقياس نسبة التوقعات الصحيحة (كل من التوقعات الإيجابية والسلبية) من إجمالي التوقعات ويعطى وفق الصيغة التالية:

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

- **مقياس الإحكام Precision:** يمثل هذا المقياس نسبة التوقعات الإيجابية الصحيحة من إجمالي التوقعات الإيجابية ويعطى وفق الصيغة التالية:

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP})$$

- **مقياس الاسترجاع Recall:** يمثل هذا المقياس نسبة التوقعات الإيجابية الصحيحة من إجمالي الحالات الإيجابية الفعلية ويعطى وفق الصيغة التالية:

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

- **مقياس F1−score:** هو المقياس الذي يجمع بين الإحكام والاسترجاع في مقياس واحد عن طريق أخذ الوسط التوافقي بينهما ويعطى وفق الصيغة التالية:

$$F1\ score = 2[(Precision \times Recall)/(Precision + Recall)]$$

3.5 مرحلة المعالجة المسبقة للبيانات

تعد المعالجة المسبقة للصور وتحضيرها العملية الأساسية في كشف إيماءات اليد بهدف تحسين دقة النماذج المستخدمة، لذلك تم تطبيق مجموعة من العمليات على الصور قبل إدخالها للنماذج وهي:

- زيادة عدد الصور من خلال تغيير اتجاه الصور وتكبيرها وتدويرها.
- تطبيق توابع لتجزئة الصور واستخراج الميزات المهمة منها.
- تطبيق توابع لتحديد الحواف المهمة في الصور والتي تحدد مكان اليد في الصورة.

4.5 التجارب والنتائج

تم تقييم أداء أربع نماذج تعلم عميق على مجموعتي البيانات الأولى والثانية وهي: CNN، VGG−16، MobileNet V1، MobileNet V2.

1.4.5 التجارب العملية للنماذج المدروسة على مجموعة البيانات الأولى

- **نموذج CNN**

الكتلة الأساسية للشبكة العصبية التلافيفية CNN هي الطبقة التلافيفية. تتميز شبكات CNN بالتعلم الهرمي للميزات، حيث تلتقط الطبقات المبكرة الميزات منخفضة المستوى

(مثل الحواف)، وتلتقط الطبقات العميقة أنماطاً أكثر تعقيداً (مثل الأشكال والأشياء). يعد هذا الهيكل الهرمي أمراً بالغ الأهمية لمعالجة البيانات المرئية بكفاءة.

يوضح الجدول (1) بنية طبقات CNN المستخدمة في التجربة، حيث اعتمدنا على 3 طبقات تلافيفية الأولى عدد فلاترها 32 بينما الثانية 64 والثالثة 128. كما يوضح الجدول (2) الإعدادات التي تم بها ضبط بارمترات شبكة CNN أثناء تدريبها، حيث تم ضبط عدد الـ epochs المستخدمة في التجربة لتكون 25 لأنها أعطت النتائج الأفضل من بين القيم الأخرى التي تم تجريبها على مجموعة البيانات.

الجدول (1): بنية طبقات نموذج CNN على مجموعة البيانات الأولى

طبقة الالتفاف الأولى:	
عدد الفلاتر	32
أبعاد الفلتر kernel size	3
تابع التنشيط activation	ReLU
طبقة التجميع max pooling	
طبقة الالتفاف الثانية:	
عدد الفلاتر	64
أبعاد الفلتر kernel size	3
تابع التنشيط activation	ReLU
طبقة التجميع Max pooling	
طبقة الالتفاف الثالثة:	
عدد الفلاتر	128
أبعاد الفلتر kernel size	3
تابع التنشيط activation	ReLU

الجدول (2): إعدادات نموذج CNN على مجموعة البيانات الأولى

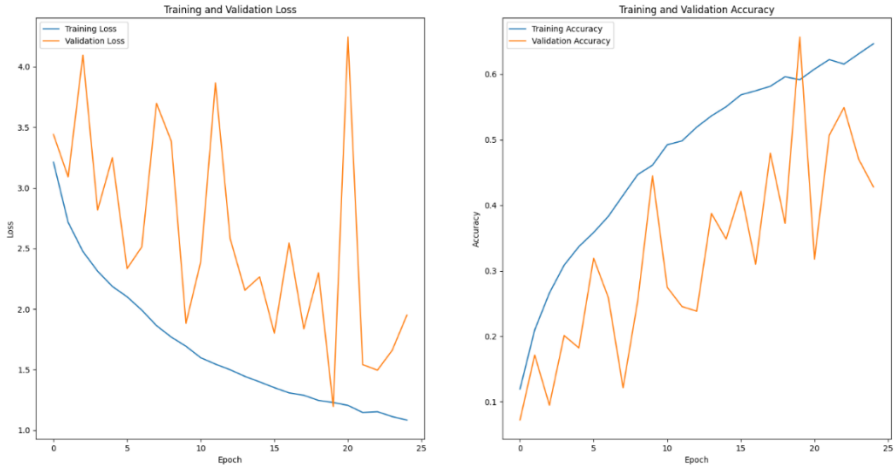
Batch size	**32**, 64, 128
Pooling	2
Epoch	10, 15, 20, **25**
Optimizer	Adam
Learning rate	0.001

الجدول (3): نتائج أداء نموذج CNN على مجموعة البيانات الاولى

Accuracy	Precision	Recall	F1-score
64%	63%	63%	63%

طبقة متصلة بالكامل Fully connected layer
طبقة الإخراج الأخيرة (طبقة التصنيف): تابع softmax

يوضح الجدول (3) نتائج أداء نموذج CNN على مجموعة البيانات الأولى باستخدام توليفة من أفضل قيم البارامترات الموضحة سابقاً. بينما يوضح الشكل (5) مقدار تناقص دالة الخسارة خلال عملية التدريب لنموذج CNN مع زيادة عدد الـ epochs (المخطط على اليسار)، ومقدار زيادة دقة النموذج المقابلة والتي وصلت لـ 64% (المخطط على اليمين).



الشكل (5): دالة الخسارة والدقة لنموذج CNN على مجموعة البيانات الأولى

- **نموذج VGG-16**

نظراً للتقدم الكبير الذي تم إحرازه في تصنيف الصور، تم تقديم العديد من بنى CNN المدربة مسبقاً والتي أثبتت فعاليتها عند تدريبها والتحقق منها على مجموعة بيانات واسعة النطاق مثل ImageNet التي تتكون من 1.2 مليون صورة طبيعية مصنفة لأكثر من 1000 فئة. بالتالي بدلاً من إنشاء نموذج CNN من الصفر، يمكن الاستفادة من هذه

البنى لتقليل الوقت المطلوب لإنتاج النموذج المطلوب. يُعتبر نموذج VGG16 واحداً من أكثر الشبكات العصبية التلافيفية العميقة تأثيراً، والتي طورتها مجموعة الهندسة البصرية (Visual Geometry Group) VGG في جامعة أكسفورد.

يوضح الجدول (4) بنية طبقات نموذج VGG16 المستخدمة، والتي تم اعتمادها بعد عدة تجارب للوصول إلى ضبط البارامترات الأفضل لهذا النموذج، كما يوضح الجدول (5) الإعدادات التي تم بها ضبط بارمترات نموذج VGG16 أثناء تدريبه، حيث تم ضبط قيمة الـ epochs المستخدمة في التجربة لتكون 20 مع إمكانية إيقاف التدريب عند الوصول الى أفضل نتيجة تدريب ممكنة لتجنب مشكلة فرط التلاؤم overfitting، لذا تم إيقاف التدريب عند الحقبة 17 كونها أعطت أفضل النتائج من بين القيم الأخرى التي تم تجريبها على مجموعة البيانات، كما اعتمدنا ايضاً على Adam optimizer.

الجدول(4): بنية طبقات نموذج VGG-16 على مجموعة البيانات الأولى

نموذج VGG16 المدرب مسبقاً	
طبقة التجميع GlobalAveragePooling2D	
طبقة متصلة بالكامل Fully connected layer	
طبقة Dropout	0.5
طبقة الإخراج الأخيرة (طبقة التصنيف): تابع softmax	

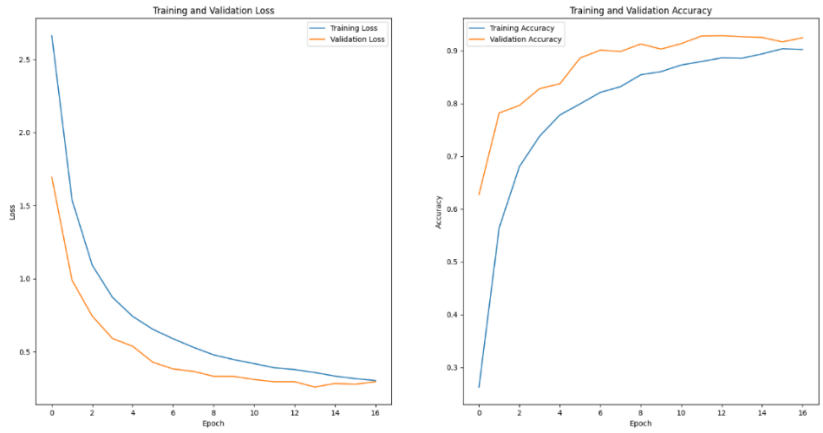
الجدول (5): إعدادات نموذج VGG-16 على مجموعة البيانات الأولى

Batch size	**32**, 64, 128
Epoch	17
Optimizer	Adam
Learning rate	0.001

يوضح الجدول (6) نتائج أداء نموذج VGG16 على مجموعة البيانات الأولى باستخدام توليفة من أفضل قيم البارامترات الموضحة سابقاً، بينما يوضح الشكل (6) مقدار تناقص دالة الخسارة خلال عملية التدريب للنموذج مع زيادة عدد الـ epochs ومقدار زيادة دقة النموذج المقابلة والتي وصلت لـ 90% تقريباً.

الجدول(6): نتائج أداء نموذج VGG-16 على مجموعة البيانات الاولى

F1-score	Recall	Precision	Accuracy
93%	94%	94%	90%



الشكل (6): دالة الخسارة والدقة لنموذج VGG-16 على مجموعة البيانات الأولى

- **نموذج MobileNet V1**

يتم استخدام MobileNet لأنها بنية عمل فعالة مصممة للتشغيل على أي منصة محدودة حسابياً مثل الأجهزة المحمولة فهي تعد منصات غير قوية مثل أجهزة الكمبيوتر ذات الأغراض العامة، ولكنها تتطلب دقة تصنيف قوية، وذلك لأنّ الـ MobileNet شبكة عصبية تلافيفية عميقة خفيفة الأوزان، أصغر حجماً، وأسرع في الأداء مقارنةً مع العديد من نماذج الشبكات العصبية الشائعة. يتم في شبكة MobileNet تغيير حجم كل صورة ملونة إلى 224 × 224 بكسل ثم يتم تغذيتها كمدخل لشبكة التعلم، والتي يمكنها معالجتها بنفس سرعة معالجة الصورة الغير ملونة أي ذات قناة واحدة.

يوضح الجدول (7) بنية طبقات نموذج MobileNetV1 المستخدمة، والتي تم اعتمادها بعد عدة تجارب للوصول إلى ضبط البارامترات الأفضل لهذا النموذج، كما يوضح الجدول

(8) الإعدادات التي تم بها ضبط بارمترات نموذج MobileNet V1 أثناء تدريبه. في حين يوضح الجدول (9) نتائج أداء نموذج MobileNet V1 على مجموعة البيانات الاولى باستخدام توليفة من أفضل قيم البارامترات الموضحة سابقاً. ويبين الشكل (7) مقدار تناقص دالة الخسارة خلال عملية التدريب للنموذج مع زيادة عدد الـ epochs، ومقدار زيادة دقة النموذج المقابلة والتي وصلت لـ 94% تقريباً.

الجدول(7): بنية طبقات نموذج MobileNet V1 على مجموعة البيانات الأولى

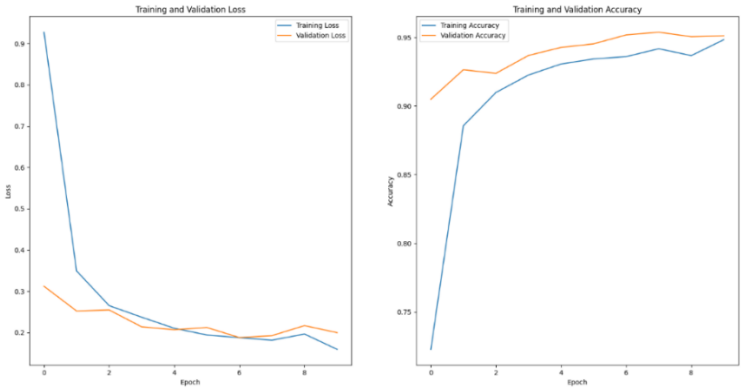
نموذج MobileNet V1 المدرب مسبقاً	
طبقة التجميع GlobalAveragePooling2D	
طبقة متصلة بالكامل Fully connected layer	
طبقة Dropout	0.5
طبقة الإخراج الأخيرة (طبقة التصنيف): تابع softmax	

الجدول(8): إعدادات نموذج MobileNet V1 على مجموعة البيانات الأولى

Batch size	**32**، 64، 128
Epoch	**11**، 20
Optimizer	Adam
Learning rate	0.001

الجدول(9): نتائج أداء نموذج MobileNet V1 على مجموعة البيانات الأولى

F1-score	Recall	Precision	Accuracy
95%	95%	96%	95%



الشكل (7): دالة الخسارة والدقة لنموذج MobileNet V1 على مجموعة البيانات الأولى

- **نموذج MobileNet V2**

إنّ نموذج MobileNetV2 هو نموذج محسن من MobileNetV1، وهو يقلل بشكل كبير من عدد المعلمات والتكلفة الحسابية مقارنة بالهندسة المعمارية التقليدية (مثل VGG-16 أو ResNet) مع الحفاظ على الدقة العالية. كما أنه أسرع بحوالي 30-40٪ من MobileNetV1 نظراً لتقليل العملية الحسابية للنصف والتقليل من البارمترات بنسبة 30%.

يوضح الجدول (10) بنية طبقات نموذج MobileNet V2 المستخدمة، والتي تم اعتمادها بعد عدة تجارب للوصول إلى ضبط البارامترات الأفضل لهذا النموذج. كما يوضح الجدول (11) الإعدادات التي تم بها ضبط بارمترات نموذج MobileNet V2 أثناء تدريبه. في حين يوضح الجدول (12) نتائج أداء نموذج MobileNet V2 على مجموعة البيانات الأولى باستخدام توليفة من أفضل قيم البارامترات الموضحة سابقاً، ويبين الشكل(8) مقدار تناقص دالة الخسارة خلال عملية التدريب للنموذج مع زيادة عدد الـ epochs، ومقدار زيادة دقة النموذج المقابلة والتي وصلت لـ 96% تقريباً.

الجدول(10): بنية طبقات نموذج MobileNet V2 على مجموعة البيانات الأولى

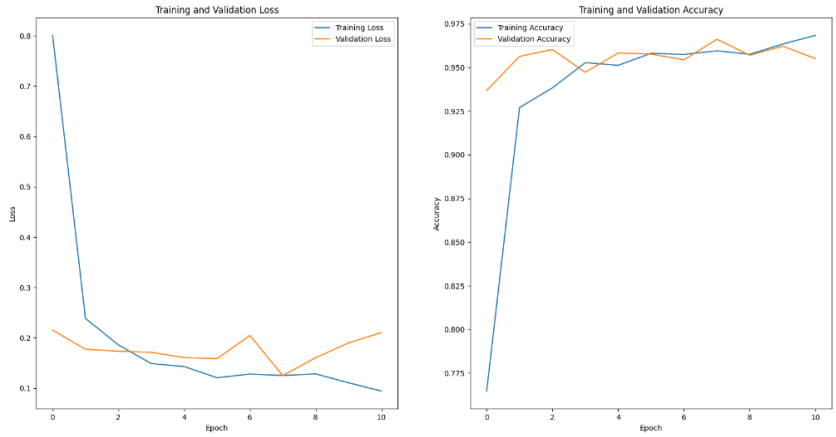
نموذج MobileNet V2 المدرب مسبقاً	
طبقة التجميع GlobalAveragePooling2D	
طبقة متصلة بالكامل Fully connected layer	
طبقة Dropout	0.5
طبقة الإخراج الأخيرة (طبقة التصنيف): تابع softmax	

الجدول(11): إعدادت نموذج MobileNet V2 على مجموعة البيانات الأولى

Batch size	**32**، 64، 128
Epoch	**10**، 20
Optimizer	Adam
Learning rate	0.001

الجدول(12): نتائج أداء نموذج MobileNet V2 على مجموعة البيانات الأولى

F1-score	Recall	Precision	Accuracy
96%	97%	97%	96%



الشكل (8): دالة الخسارة والدقة لنموذج MobileNet V2 على مجموعة البيانات الأولى

2.4.5 التجارب العملية للنماذج المدروسة على مجموعة البيانات الثانية

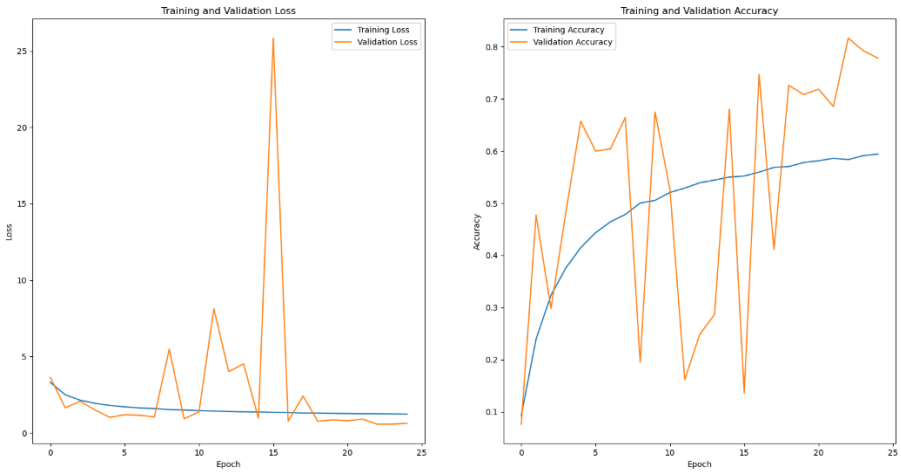
- **نموذج CNN**

تم تطبيق نموذج CNN على مجموعة البيانات الثانية بنفس بنية طبقاته الموضحة في الجدول (1) ونفس إعدادات بارمتراته الموضحة في الجدول (2) وتم التدريب على epoch=25، والحصول على النتائج الموضحة في الجدول (13) التي تمثل أداء نموذج CNN على مجموعة البيانات الثانية.

الجدول (13): نتائج أداء نموذج CNN على مجموعة البيانات الثانية

F1-score	Recall	Precision	Accuracy
58%	60%	60%	59%

يوضح الشكل (9) مقدار تناقص دالة الخسارة خلال عملية التدريب لنموذج CNN مع زيادة عدد الـ epochs، ومقدار زيادة دقة النموذج المقابلة والتي وصلت لـ 59% تقريباً.



الشكل (9): دالة الخسارة والدقة لنموذج CNN على مجموعة البيانات الثانية

- **نموذج VGG-16**

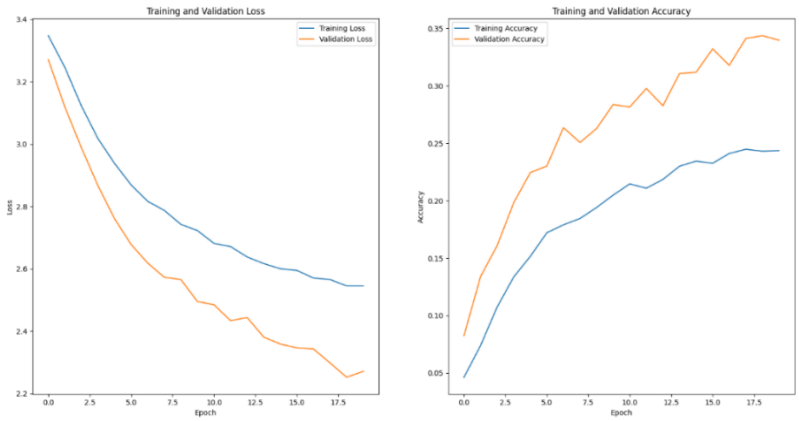
تم تطبيق نموذج VGG-16 على مجموعة البيانات الثانية بنفس بنية طبقاته الموضحة في الجدول (4) ونفس إعدادات بارمتراته الموضحة في الجدول (5)، ولكن تم التدريب على epoch=20 وتم الحصول على النتائج الموضحة في الجدول (14) التي تمثل أداء نموذج VGG-16 على مجموعة البيانات الثانية.

الجدول (14): نتائج أداء نموذج VGG-16 على مجموعة البيانات الثانية

F1-score	Recall	Precision	Accuracy
23%	35%	38%	24%

يوضح الشكل (10) مقدار تناقص دالة الخسارة خلال عملية التدريب لنموذج VGG-16 مع زيادة عدد الـ epochs، ومقدار زيادة دقة النموذج المقابلة والتي وصلت لـ 24%

تقريباً.



الشكل (10): دالة الخسارة والدقة لنموذج VGG-16 على مجموعة البيانات الثانية

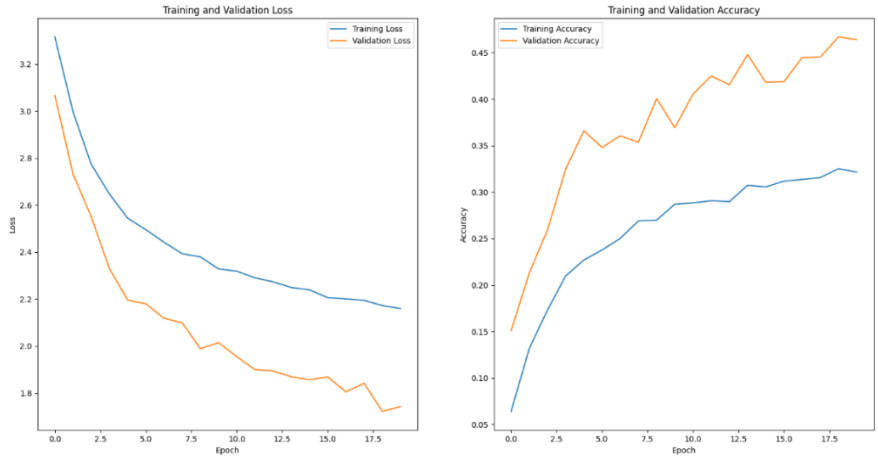
- **نموذج MobileNet V1**

تم تطبيق نموذج MobileNet V1 على مجموعة البيانات الثانية بنفس بنية طبقاته الموضحة في الجدول (7) ونفس إعدادات بارمتراته الموضحة في الجدول (8) ولكن تم التدريب على epoch=20، وتم الحصول على النتائج الموضحة في الجدول (15) التي تمثل أداء نموذج MobileNet V1 على مجموعة البيانات الثانية.

الجدول (15): نتائج أداء نموذج MobileNet V1 على مجموعة البيانات الثانية

F1-score	Recall	Precision	Accuracy
30%	40%	46%	32%

يوضح الشكل (11) مقدار تناقص دالة الخسارة خلال عملية التدريب للنموذج مع زيادة عدد الـ epochs، ومقدار زيادة دقة النموذج المقابلة والتي وصلت لـ 32% تقريباً.



الشكل (11): دالة الخسارة والدقة لنموذج MobileNet V1 على مجموعة البيانات الثانية

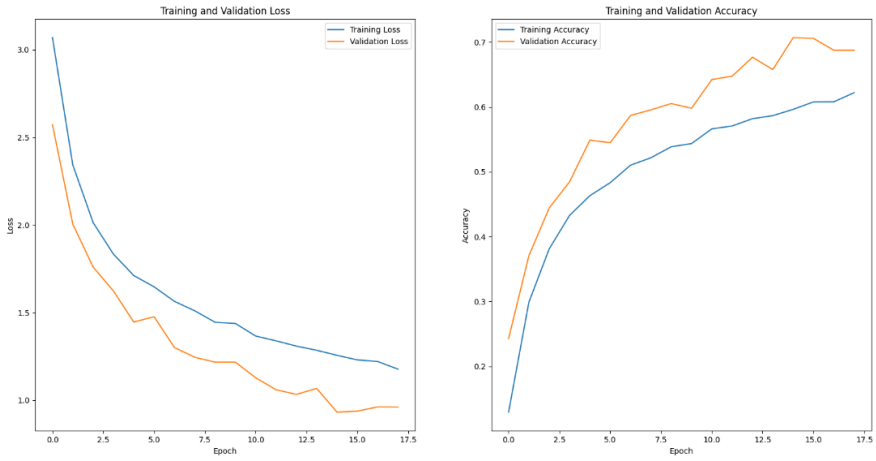
- **نموذج MobileNet V2**

تم تطبيق نموذج MobileNet V2 على مجموعة البيانات الثانية بنفس بنية طبقاته الموضحة في الجدول (10) ونفس إعدادات بارمتراته الموضحة في الجدول (11) ولكن تم إيقاف التدريب عند epoch=18، وتم الحصول على النتائج الموضحة في الجدول (16) التي تمثل أداء نموذج MobileNet V2 على مجموعة البيانات الثانية.

الجدول (16): نتائج أداء نموذج MobileNet V2 على مجموعة البيانات الثانية

F1-score	Recall	Precision	Accuracy
60%	71%	76%	62%

يوضح الشكل (12) مقدار تناقص دالة الخسارة خلال عملية التدريب للنموذج مع زيادة عدد الـ epochs، ومقدار زيادة دقة النموذج المقابلة والتي وصلت لـ 62% تقريباً.



الشكل (12): دالة الخسارة والدقة لنموذج MobileNet V2 على مجموعة البيانات الثانية

3.4.5 تحليل ومناقشة النتائج

يوضح الجدول (17) ملخص نتائج النماذج الأربعة المدروسة على مجموعتي البيانات الأولى والثانية. نلاحظ أن النتائج على مجموعة البيانات الثانية عموماً ليست جيدة نتيجة أن مجموعة البيانات هذه تحتوي على تنوع أكبر في الصور أو خصائص بصرية معقدة مثل الإضاءة، الخلفيات، الألوان، والأنماط المختلفة للأحرف أو الإشارات، وبالتالي فإن النماذج تجد صعوبة في التكيف مع هذه الاختلافات. مثلاً نموذج VGG16 يعتمد بشكل كبير على الأشكال والأنماط الأساسية (low-level features) في المراحل الأولى من الشبكة، وعندما كانت مجموعة البيانات تحتوي على خصائص بصرية متداخلة أو معقدة، أصبح النموذج غير قادراً على استخراج الميزات المطلوبة بشكل فعال. كما أنّ مجموعة البيانات لم تكن كبيرة بما يكفي لتمثيل الأنماط المختلفة بشكل جيد، ممّا أدى إلى تدهور الدقة بشكل كبير في هذا النموذج. بالإضافة الى أنّ VGG16 ليس جيداً بما فيه الكفاية في التمييز بين الخلفية والموضوع الرئيسي في الصورة عند التدريب على بيانات ذات

خلفيات متنوعة، وذلك كون الطبقات التلافيفية في النموذج تتعلم خصائص منخفضة المستوى (مثل الحواف والخطوط)، بالتالي فإنّ اختلاف الخلفيات قد يؤدي إلى تشويش النموذج وجعله أقل قدرة على التعلم الفعال من البيانات. في حين أنّ نماذج MobileNetV1 أو MobileNetV2 تمتاز بأنها أخف وتتمتع بعدد أقل من المعاملات وقابلة للتكيف مع الصور ذات الأبعاد الكبيرة، ومع ذلك أعطت هذه النماذج دقة سيئة على مجموعة البيانات الثانية وذلك لأنها تحتاج إلى عدد خطوات كبير للتدريب، الأمر المكلف جداً حسابياً وكأننا لم نستفد من مزايا النماذج المدربة مسبقاً مقارنة مع النماذج التقليدية.

الجدول (17): ملخص نتائج النماذج على مجموعتي البيانات الأولى والثانية

النماذج المدروسة	مجموعة البيانات الأولى				مجموعة البيانات الثانية			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
نموذج CNN	64%	63%	63%	63%	59%	60%	60%	58%
نموذج VGG-16	90%	94%	94%	93%	24%	38%	35%	23%
نموذج MobileNet V1	95%	96%	95%	95%	32%	46%	40%	30%
نموذج MobileNet V2	**96%**	97%	97%	96%	**62%**	76%	71%	60%

نستنتج أن طبيعة مجموعة البيانات تلعب دوراً مهماً في دقة النماذج وطريقة عملها عموماً، ومن خلال تجاربنا تبين أن مجموعة البيانات التي قمنا بجمعها تتناسب أكثر مع نماذج التعلم العميق المدروسة وساعدت على إعطاء أفضل نتيجة بأقل وقت تدريب ممكن، وتتناسب أكثر مع هدف هذا البحث في تسهيل عملية تواصل الصم والبكم مع مجتمعهم كونه سيتم استثمار هذا النموذج في النهاية في تطبيق يستطيع التعرف على الإشارات من حركات اليد أمام كاميرا الموبايل وترجمتها لأحرف مكتوبة.

بالمقارنة بين النماذج التي تم التدريب عليهم نجد أن نموذجي MobileNet V2 و MobileNet V1 قد أعطوا نتيجة جيدة وسرعة في التنفيذ مقارنة مع النماذج الأخرى وكان نموذج MobileNet V2 افضل من MobileNet V1 بنسبة 1% بسبب بنيته التي تخفف من الأعباء الحسابية في طبقاته، حيث أنه بمعالجة بسيطة للصور مثل تجزئة الصور وتحديد الحواف فيها استطاع النموذج اكتشاف اليد والتمييز بين الأحرف المتشابهة بأفضل شكل ممكن مع نسبة خطأ قليلة نسبياً.

5. مقارنة مع دراسات مشابهة

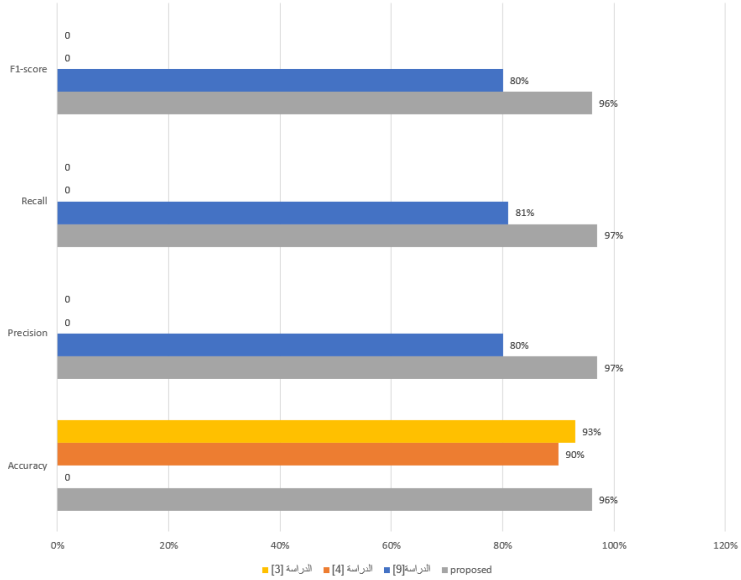
تم بناء صور مجموعة البيانات حسب ايماءات محددة ومعرّفة عالمياً بالنسبة للغة الإشارة العربية لذلك يمكن القيام ببعض المقارنات بين نتائجنا ونتائج ما سبق من أعمال على مجموعة البيانات للغة الإشارة العربية مع مقارنة النماذج المستخدمة في كل دراسة.

يظهر الشكل (13) والجدول (18) تفوق نتائج الدراسة الحالية على ما سبق من أعمال وتقنيات من ناحية الدقة وسرعة التدريب، حيث حققت دقة قدرها 96% مع استخدام المعاملات التالية: حجم الدفعات Batch Size يساوي 32، عدد الحقبات epoch تساوي 10، خوارزمية التحسين Adam، ومعدل التسريب 0.5.

الجدول (18): مقارنة نتائج الدراسة مع الدراسات المشابهة

الدراسات المشابهة	مجموعة البيانات المستخدمة	النموذج المستخدم	عدد الـ epochs	Accuracy	Precision	Recall	F1-score
الدراسة [9]	ArSL21L	MobileNet V2	60	-	%80	%81	%80
الدراسة [4]	غير متاحة	CNN	100	90%	-	-	-
الدراسة [3]	ArSL2018	CNN	300	92.9%	-	-	-

الدراسة الحالية	مجموعة البيانات التي تم تجميعها يدوياً من أجل هذه الدراسة	MobileNet V2	10	**96%**	97%	97%	96%

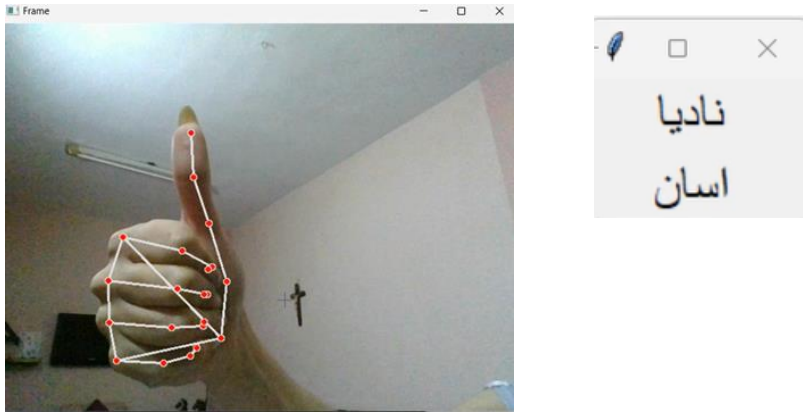


الشكل (13): مقارنة نتائج جميع المقاييس مع الدراسات المشابهة

6. بناء تطبيق للتعرف على لغة الإشارة اعتماداً على النموذج الأفضل

بالاستفادة من نتائج الدراسة الحالية، تم بناء تطبيق على أجهزة الحاسب باستخدام نموذج MobileNet V2 يقوم بالتقاط شكل اليد عبر الكاميرا ضمن مربع محدد عليها، وكتابة الحرف المقابل لحركة اليد التي يتم التنبؤ بها وتكوين جملة. عندما تكون اليد داخل المربع المحدد على الواجهة الاولى بالشكل الصحيح يتم ظهور مجموعة من النقاط التي تحدد شكل اليد وتميزه عن باقي أجزاء الجسم ويجب الضغط على زر (r) ليتم كتابة الحرف المطابق لشكل اليد. وعند الانتهاء من جميع الاحرف للكلمة الوحدة يجب الضغط على الزر (w) لكي يتم كتابة الكلمة كاملة التي تم التنبؤ بها. في حال أردنا إضافة مسافة بين

الكلمتين يجب الضغط على زر(s)، وفي حال أردنا الحذف يجب الضغط على زر(d). لقد تم اتباع هذه الطريقة في القراءة وكتابة التنبؤ لضبط سرعة الكتابة وضمان كتابة الحرف الصحيح للصورة التي يتم التقاطها في وقتها الصحيح. يوضح الشكل (14) مثال عن آلية عمل التطبيق.



الشكل (14): مثال عن التقاط اليد باستخدام كاميرا الحاسب وكتابة الكلمة المطلوبة

كما تم بناء تطبيق تعليمي يحتوي على إيماءات الأحرف على شكل أزرار بحيث يتم كتابة الكلمة المطلوبة من خلال الضغط على صورة الحرف ثم الضغط على زر عرض الكلمة ليتم كتابتها بشكل كامل، مع إمكانية حذف حرف او إضافة مسافة بين الكلمات. وبالعكس يمكن كتابة الكلمة المطلوبة ومن ثم الضغط على زر عرض الصور ليتم عرض صور للأحرف المكتوبة كما هو موضح في الشكل (15)، بالتالي يمكن للأشخاص الصم والبكم التواصل مع محيطهم بسهولة بالإضافة الى إمكانية تعلم ايماءات لغة الإشارة العربية بسهولة ايضاً للراغبين بذلك.

الشكل (15): مثال عن واجهة البرنامج التعليمي

7. الاستنتاجات والتوصيات

تم في هذا البحث تطوير نموذج للتعرف على لغة الإشارة العربية من خلال استخدام تقنيات التعلم العميق ونقل التعلّم مع بناء مجموعة بيانات خاصة بلغة الإشارة العربية معترف بها من قبل جمعية رعاية الصم والبكم في مدينة حمص يمكن الاستفادة منها في العديد من الدراسات والأبحاث المستقبلية ضمن هذا المجال، واستخدام هذا النموذج ضمن تطبيق ويب يساعد على تحويل الإشارات إلى نص باللغة العربية واعتماده كوسيلة لتواصل الصم والبكم مع محيطهم في العديد من المجالات.

تناولت الدراسة تقييم أداء 4 نماذج تعلم عميق وهي CNN، VGG-16، MobileNet V1، MobileNet V2 من أجل مهمة التعرف على لغة الإشارة العربية، وتوصلت إلى أنّ نموذج MobileNet V2 حقق أفضل النتائج من حيث حيث الدقة التي بلغت 96% وسرعة التدريب على مجموعة البيانات الأولى التي تم جمعها يدوياً، كما تبيّن أنّ طبيعة مجموعة البيانات تلعب دوراً هاماً في دقة النماذج المدروسة عموماً.

وبناءً على الدراسة الحالية والنتائج التي تم الحصول عليها، يمكن اقتراح بعض الأعمال المستقبلية لتطوير هذا العمل بشكل أفضل، ومنها:

- **تحسين بيانات التدريب:** يمكن توسيع مجموعة البيانات لتشمل تنوعاً أكبر في الإشارات والحركات لضمان تحسين دقة النماذج مثل إضافة بعض الايماءات لكلمات مستخدمة بشكل يومي.
- **الدمج بين النماذج:** تجربة دمج النماذج المختلفة أو إنشاء شبكة هجينة تجمع بين ميزات كل نموذج للحصول على أداء أقوى.
- **استخدام تقنيات تعلم محسنة:** تطبيق تقنيات مثل التعلم العميق المتقدم أو النقل المعرفي لتحسين أداء النماذج على مجموعات البيانات الصغيرة.
- **استخدام تقنيات تحسين الأداء:** مثل التنقيح المستمر للنماذج وضبط المعلمات بشكل أفضل للحصول على نتائج دقيقة بشكل أكبر.

المراجع الأجنبية

[1] Deafness and hearing loss. [online] Available at: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss [Accessed 29 Oct. 2020].

[2] Li, D., Rodriguez, C., Yu, X. and Li, H., 2020. **Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison**. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 1459-1469).

[3] Althagafi, A., Alsubait, G.T. and Alqurash, T., 2020. **ASLR: Arabic sign language recognition using convolutional neural networks**. IJCSNS International Journal of Computer Science and Network Security, 20(7), pp.124-129.

[4] Kamruzzaman, M.M., 2020. **Arabic sign language recognition and generating Arabic speech using convolutional neural network**. Wireless Communications and Mobile Computing, 2020(1), p.3685614.

[5] Ritonga, M., Abd El-Aziz, R.M., Alazzam, M.B., Alassery, F., NagaJyothi, A., Abd Algani, Y.M. and Balaji, S., 2021. **Machine Learning Approach for Gesture Based Arabic Sign Language Recognition for Impaired People**.

[6] AlKhuraym, B.Y., Ismail, M.M.B. and Bchir, O., 2022. **Arabic sign language recognition using lightweight cnn-based architecture**. International Journal of Advanced Computer Science and Applications, 13(4).

[7] Hdioud, B. and Tirari, M.E.H., 2023. **A deep learning based approach for recognition of Arabic sign language letters**. International Journal of Advanced Computer Science and Applications, 14(4).

[8] Alharthi, N.M. and Alzahrani, S.M., 2023. **Vision transformers and transfer learning approaches for arabic sign language recognition**. Applied Sciences, 13(21), p.11625.

[9] Mohamed Hamdy Abdallah Eissa, Alaa Ismaeel, Sherimon P.C., Vinu Sherimon, Remya Revi K.(2022). **Arabic Sign Language Letter Recognition using MobileNet-v2 Deep Neural Network**.

[10]Lum, K.Y., Goh, Y.H. and Lee, Y.B., 2020. **American sign language recognition based on MobileNetV2**. Advances in Science, Technology and Engineering Systems Journal, 5(6), pp.481-488.