العدد 27
ديسمبر 2024

www.mada.org.qa

نفاذ

من مدى

معالجة لغة الإشارة

التعرف التلقائي على لغة الإشارة العربية استنادًا إلى الإيماءات: نهج التعلم الفيدرالي (التعلم الموزع) صفحة 34

ترجمة النص العربي إلى مصطلحات عربية للغة الإشارة صفحة 56

تطوير أنظمة الذكاء الاصطناعي لترجمة لغة الإشارة في الفيديو صفحة 68

ISSN 0278-9914

رؤساء التحرير

أماني علي التميمي
مركز مدى، قطر

أشرف عثمان
مركز مدى، قطر

هيئة التحرير

أميرة ذويب،
مركز مدى، قطر.

آمنة محمد المطوع،
مركز مدى، قطر.

هيئة المراجعة

أحلام أصيلة،
مركز الدراسات العليا الصناعية، رانس، فرنسا.

أحمد تليلي،
معهد التعلم الذكي بجامعة بكين للمعلمين، الصين

عليا جمال الكثيري،
مركز مدى، قطر.

الجازي الجبر،
مركز مدى، قطر.

آمنة محمد المطوع،
مركز مدى، قطر.

دينا آل ثاني،
جامعة حمد بن خليفة، قطر.

فخرية ألتيناي،
جامعة الشرق الأدنى، الجزء الشمالي من قبرص.

فتحي السالمي،
جامعة جدة، المملكة العربية السعودية

هيفاء بن الحاج،
جامعة قطر، قطر.

هاجر شلغومي،
المركز الكندي للتنوع والشمول، كندا.

هناء ربوش،
المعهد العالي للإدارة، سوسة، تونس.

محمد كثير خريبي،
مركز مدى، قطر.

أسامة الغول،
مركز مدى، قطر.

سامية كوكي،
كليات التقنية العليا، الإمارات العربية المتحدة.

توفيق الحضرمي،
جامعة نوتنجهام ترينت، المملكة المتحدة.

زياد بويدة،
جامعة كارلتون، أوتاوا، كندا.

نـــفـــاذ
من مدى

نوفمبر 27
ديسمبر 2024

الرقم الدولي الموحد للدوريات (النسخة الرقمية): 2789-9152
الرقم الدولي الموحد للدوريات (النسخة المطبوعة): 2789-9144

إعادة استخدام الحقوق وأذونات إعادة الطباعة
"نفاذ "هي مجلة متاحة للجميع. يُسمح بالاستخدام التعليمي أو الشخصي لهذه المواد بدون رسوم ، بشرط أن يكون هذا الاستخدام: 1) غير هادف للربح 2) يتضمن هذا الإشعار والاقتباس الكامل للعمل الأصلي في الصفحة الأولى من النسخة و 3) لا يلمح هذا الإشعار إلى مصادقة مركز مدى على أي من منتجات أو خدمات الطرف الثالث. يُسمح للمؤلفين وشركاتهم بنشر النسخة المقبولة من "نفاذ "على خوادم الويب الخاصة بهم دون إذن ، بشرط أن يظهر هذا الإشعار والاقتباس الكامل للعمل الأصلي على الصفحة الأولى من النسخة المنشورة. إن النسخة المقبول استخدامها هي النسخة التي تمت مراجعتها من قبل المؤلف لإضافة اقتراحاته بعد المراجعة، ولكن ليس النسخة المنشورة من قبل مركز مدى والتي قام المركز بتدقيقها وتحريرها وتنسيقها. لمزيد من المعلومات، يرجى زيارة: https://nafath.mada.org.qa. يجب الحصول من مركز مدى على إذن بإعادة طباعة / إعادة نشر هذه المواد لأغراض تجارية أو دعائية أو ترويجية أو لإنشاء أعمال جديدة لإعادة البيع أو إعادة التوزيع.

نفاذ© 2023 من مركز مدى برقم ترخيص CC BY-NC-ND 4.0

مركز "مدى"

مركز "مدى" – مركز التكنولوجيا المساعدة في قطر، مؤسسة خاصة ذات نفع عام تأسست عام ٢٠١٠ كمبادرة لتوطيد معاني الشمولية الرقمية وبناء مجتمع تكنولوجي قابل للنفاذ لذوي الإعاقة. وقد أصبح مدى اليوم مركز الامتياز في النفاذ الرقمي باللغة العربية في العالم.

يعمل المركز عبر شراكات استراتيجية ذكية على تمكين قطاع التعليم لضمان التعليم الشامل، وقطاع المجتمع ليصبح أكثر شمولاً من خلال تكنولوجيا المعلومات والاتصالات وقطاع التوظيف لتعزيز فرص التوظيف والتطوير المهني وريادة الأعمال للأشخاص ذوي الإعاقة.

ويحقق المركز أهدافه من خلال بناء قدرات الشركاء ودعم تطوير واعتماد المنصات الرقمية وفق المعايير العالمية للنفاذ الرقمي وتقديم الاستشارات ورفع الوعي وزيادة عدد حلول التكنولوجيا المساعدة باللغة العربية عبر برنامج مدى للابتكار، وذلك لتمكين تكافؤ الفرص لمشاركة الأشخاص ذوي الإعاقة في المجتمع الرقمي.

حول نفاذ

"نفاذ" هي دورية يصدرها مركز مدى باللغتين العربية والإنجليزية كل ثلاثة أشهر تهدف لتكون مصدر المعلومات الرئيسي حول أحدث التوجهات والابتكارات في مجال نفاذ تكنولوجيا المعلومات والاتصالات. وانطلاقاً من دورها كنافذة للمعلومات عبر العالم تسلط دورية نفاذ الضوء على العمل الرائد الذي تم في مجال تلبية الطلبات المتزايدة على حلول وخدمات نفاذ تكنولوجيا المعلومات والاتصالات والتكنولوجيا المساعدة في قطر والمنطقة العربية والعالم.

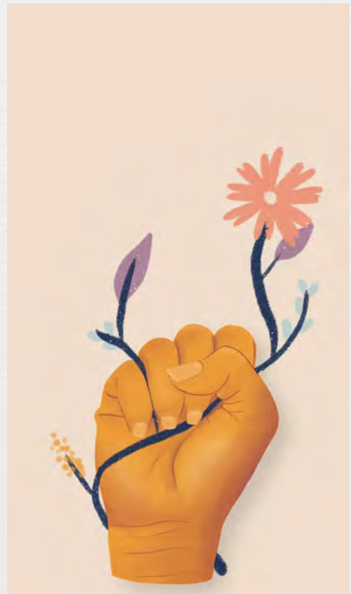
المحتويات

الصفحة 08
دراسة حالة حول استخدام الذكاء الاصطناعي في ترجمة لغة الإشارة التركية

الأستاذ المشارك أوزر سيليك
بينار رضا

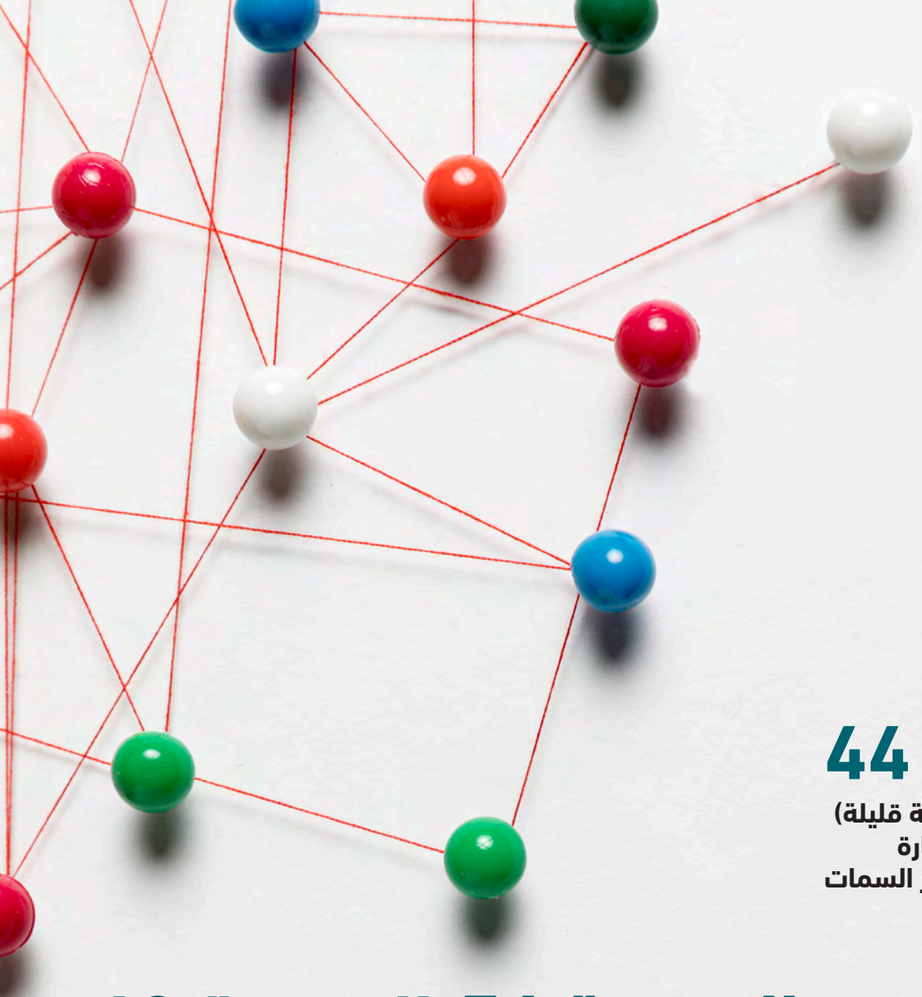
الصفحة 17
نظام التعرف على لغة الإشارة التونسية للإشارات غير المتماثلة الثابتة ثنائية اليدين باستخدام التعلم الانتقالي العميق

آمنة دقنو
هيثم الهرميسي
نبيل تبان



الصفحة 34
التعرف التلقائي على لغة الإشارة العربية استنادًا إلى الإيماءات:
نهج التعلم الفيدرالي (التعلم الموزع)

أحمد الزعبي
توفيق الحضرمي
أمجد البشايرة
لجين بني يونس



الصفحة 44
(التعلم باستخدام أمثلة قليلة) للتعرف على لغة الإشارة باستخدام تقنية انتشار السمات المضمنة

أمجد السلمي
خولة باجبع
حمزة لقمان
عصام لعراجي

الصفحة 56
ترجمة النص العربي إلى مصطلحات عربية للغة الإشارة

دعاء الغامدي
منصور السليمان
يوسف العوهلي
محمد أ. بن شريف
محمد الجابري
محمد أ. مختيش

الصفحة 68
تطوير أنظمة الذكاء الاصطناعي لترجمة لغة الإشارة في الفيديو

الأستاذ المشارك أوزر سيليك
أحمد أفجي أوغلو

Nafath

دعوة مفتوحة لتقديم المقالات

NAFATH
2024
Majlis | مجلس

دعوة مفتوحة لتقديم الأوراق

تعد "نفاذ" دورية تصدر بشكل ربع سنوي وفعالية تتضمن عدة ورش عمل تُعرف باسم "مجلس نفاذ".

يهدف مجلس نفاذ إلى عرض أحدث الأبحاث والتطورات وتبادل المعرفة في مجال الشمول الرقمي. وفي كل إصدار من نفاذ نقوم بتشجيع البحوث ووجهات النظر المبتكرة على المشاركة بالأوراق البحثية مما يعزز دورة الابتكار والتعاون في هذه المجالات. إن دورية "نفاذ" متاحة باللغتين العربية والإنجليزية. وهدفها دعم الحاجة المتزايدة لتكنولوجيا المعلومات والاتصالات القابلة للنفاذ والتكنولوجيا المساعدة في قطر والمنطقة العربية والعالم.

المواضيع الرئيسية

- التعاون متعدد التخصصات: التقدم الرائد في النفاذ الرقمي و التكنولوجيا المساعدة
- دور واجهات المستخدم من الجيل القادم في دعم الأشخاص ذوي الإعاقة
- المعالجة الآلية للغة الإشارة
- دور النفاذ الرقمي في تشكيل المدن الذكية المرنة والشاملة للجميع

التقديمات

ندعو لتقديم الأوراق البحثية باللغة الإنجليزية أو العربية فقط، كما يجب أن تكون منسقة وفقًا لإرشادات نموذج نفاذ (لمزيد من التفاصيل حول هذه التعليمات يرجى زيارة تعليمات المؤلفين - دورية نفاذ من مدى). ويمكن للمؤلفين تقديم أوراقهم من خلال بوابة التقديم الإلكترونية المتاحة على : nafath.mada.org.qa

لماذا تنشر ورقتك البحثية معنا؟

تُنشر جميع الأوراق البحثية المقبولة والمقدمة في دورية نفاذ تحت الرقم الدولي الموحد للدوريات (ISNN) على الورق وعلى منصة الدعم الرقمي (Digital Support). إن دورية نفاذ عضو في (http://www.crossref.org/) وكل ورقة بحثية في مكتبتنا الرقمية تعطى معرف للكائن الرقمي (DOI). وسيتم تقديم الأوراق للفهرسة في الباحث العلمي من جوجل.

دراسة حالة حول استخدام الذكاء الاصطناعي في ترجمة لغة الإشارة التركية

الأستاذ المشارك أوزر سيليك
ozer@ogu.edu.tr
جامعة إسكي شهير عثمان غازي
قسم الرياضيات وعلوم الكمبيوتر،
كلية العلوم، جامعة إسكي شهير
عثمان غازي، إسكي شهير، تركيا

بينار رضا
pinar.reza@signfordeaf.com
مبنى ETGB تكنوبارك رقم:
44/106 إسكي شهير/تركيا

الكلمات الرئيسية
ترجمة لغة الإشارة، الذكاء
الاصطناعي، البرمجة اللغوية العصبية،
النفاذ الرقمي

الملخص - يعاني نحو 3 ملايين الأشخاص في تركيا من ضعف السمع وقد أدى التحول إلى المنصات الرقمية أثناء الجائحة إلى تفاقم تحديات إمكانية النفاذ حيث تتجاهل مواقع الويب والتطبيقات غالبًا احتياجات هذه الفئة. وتُظهر الأبحاث أن 50% من الأشخاص الذين يعانون من ضعف السمع يجدون صعوبة في فهم النص المكتوب بسبب كون لغة الإشارة التركية لغتهم الأولى في حين تشكل التركية اللغة الثانية. كما تشكل الاختلافات في القواعد النحوية بين اللغة التركية ولغة الإشارة التركية إلى جانب مفردات لغة الإشارة المحدودة عائقاً أكبر أمام الفهم. وبهدف حل هذه المشكلة قمنا بتطوير أنظمة ترجمة بلغة الإشارة المدعومة بالذكاء الاصطناعي والتي تسمح للأشخاص الصم أو ضعاف السمع بالنفاذ إلى المحتوى الرقمي بلغة الإشارة التركية. ويترجم نظام (SignForDeaf) النص إلى مقاطع فيديو بلغة الإشارة باستخدام معالجة اللغة الطبيعية (NLP) ويولد مقاطع فيديو مرنة مع انتقال سلسة بين الكلمات. و يدعم النظام حاليًا لغة الإشارة التركية مع خطط مستقبلية لإضافة لغات أخرى مثل لغة الإشارة العربية والأمريكية والبريطانية والفنلندية. وقد تم تصميم هذا النظام بالتعاون مع خبراء لغة الإشارة لضمان الدقة والتطوير الشامل للبيئة الرقمية.

المقدمة

يعاني حوالي 3 ملايين فرد في تركيا من ضعف السمع كما أن التحول إلى المنصات الرقمية أثناء الجائحة قد أدى إلى زيادة تحديات إمكانية النفاذ. وغالبًا ما تفشل مواقع الويب والتطبيقات المحمولة في مراعاة احتياجات التواصل لمجتمعات الصم وضعاف السمع. تعد لغة الإشارة التركية (TSL) اللغة الأم للعديد من الأفراد في هذا المجتمع بينما تعتبر اللغة التركية لغتهم الثانية. وأظهرت الأبحاث أن ما يقرب من 50% من الأشخاص الذين يعانون من ضعف السمع في تركيا يواجهون صعوبات في فهم النص المكتوب مما يجعل من الصعب عليهم التنقل عبر المحتوى والخدمات الرقمية والتفاعل معها بشكل فعال [1].

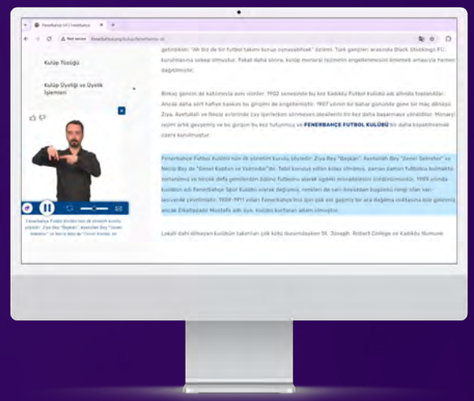
وتختلف البنية اللغوية للغة الإشارة التركية بشكل كبير عن اللغة التركية المنطوقة. ففي حين أن اللغة التركية هي لغة تراكمية ذات قواعد نحوية معقدة ولواحق فإن لغة الإشارة التركية هي شكل أبسط وأكثر مباشرة من أشكال الاتصال وهي عادة ما تستخدم أشكالًا أساسية من الكلمات. على سبيل المثال، بدلاً من قول "Ben ise gidiyorum" (والتي تُترجم إلى: "أنا ذاهب إلى العمل")، قد يقول الشخص الأصم "Ben is gitmek" ("أنا ذاهب إلى العمل"). ويخلق هذا الاختلاف في البنية حواجز أمام الأشخاص الذين يعانون من ضعف السمع أو الصم في مجال القراءة والكتابة باللغة التركية [2].

بالإضافة إلى ذلك فإن لدى لغة الإشارة التركية مفردات محدودة مقارنة بمجموعة المفردات الغنية والدقيقة للغة التركية المنطوقة. ويفرض تعقيد المرادفات والعبارات والأمثال التركية تحديات إضافية لأولئك الذين يعتمدون على لغة الإشارة. وتمتلك العديد من الكلمات في اللغة التركية عدة معاني وهكذا يصبح الفهم أكثر صعوبة بدون وجود قاموس لغة إشارة كافٍ للتعرف على هذه الفروق الدقيقة [3]. وتؤدي هذه الاختلافات اللغوية إلى فجوات كبيرة في التواصل والتي تتأثر بشكل أكبر بحقيقة أن العديد من الأشخاص الذين يعانون من ضعف السمع لم يكن لديهم سوى القليل من الوصول إلى التعليم الرسمي بلغة الإشارة مما أدى إلى تزايد الأمية في مجالات لغة الإشارة واللغة التركية المكتوبة [4].

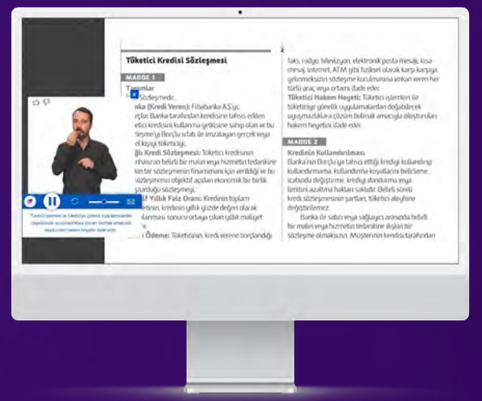
ويكمن حل هذه المشكلة في تطوير أنظمة ترجمة بلغة الإشارة تكون مدعومة بالذكاء الاصطناعي وتستخدم معالجة اللغة الطبيعية (NLP) لسد الفجوة بين اللغة التركية ولغة الإشارة التركية. ويوفر هذا النظام ترجمات مرنة في الوقت الفعلي للمحتوى الرقمي إلى مقاطع فيديو بلغة الإشارة التركية مما يوفر مستوى جديدًا من إمكانية النفاذ لمجتمعات الصم وضعاف السمع [5].

المنهجية

يتطلب تطوير نظام ترجمة لغة الإشارة التركية المدعوم بالذكاء الاصطناعي التكامل الدقيق بين تقنيات مختلفة بما في ذلك معالجة اللغة الطبيعية (NLP) والشبكات التوليدية التنافسية (GAN). حيث تبدأ العملية عندما يركز المستخدم المؤشر فوق نص ما على موقع ويب أو مستند PDF أو يحدده كما هو موضح في الشكلين 1 و2. يتم إرسال هذا النص إلى خدمتنا عبر واجهة برمجة التطبيقات حيث يخضع للتحليل الصرفي باستخدام تقنيات معالجة اللغة الطبيعية لتقسيم الجمل إلى أشكالها الجذرية. و تمثل هذه الخطوة أهمية، حيث إن لغة الإشارة التركية تعتمد على بنية جملة ابسط بكثير من اللغة التركية وتسمح هذه الخطوة بالتخلص من عناصر القواعد النحوية المعقدة مثل اللواحق ما يعد ضرورياً للحصول على ترجمة دقيقة.



الشكل 1. لقطة شاشة من موقع نادي فنربخشة الرياضي التركي والذي يستخدم البرنامج الإضافي للغة الإشارة على الويب من (SignForDeaf). وهنا تصبح الجمل المميزة قابلة للنقر ثم تُترجم على الفور إلى لغة الإشارة باستخدام الذكاء الاصطناعي.



الشكل 2. لقطة شاشة من عقد بنك تركي بصيغة PDF باستخدام البرنامج الإضافي للغة الإشارة من (SignForDeaf). وهنا تصبح الجمل المميزة قابلة للنقر ثم تُترجم على الفور إلى لغة الإشارة باستخدام الذكاء الاصطناعي.

تحليل النصوص والمعالجة المسبقة:

معالجة اللغة الطبيعية (NLP): يقوم النظام بتحليل نصوص الترجمة باستخدام تقنيات معالجة اللغة الطبيعية (NLP). وتشمل هذه المرحلة تقسيم النص إلى أجزاء وتحليل البنية النحوية واستخراج المعنى. وتمكن معالجة اللغة الطبيعية النظام من تقسيم الجمل إلى أشكالها الجذرية وإزالة العناصر النحوية مثل اللواحق غير الضرورية في لغة الإشارة التركية. وتضمن هذه العملية أن يكون المحتوى المترجم دقيقًا لغويًا وسهل الفهم في لغة الإشارة. كما تساعد معالجة اللغة الطبيعية النظام على التعامل مع المكونات الشائعة للغة المكتوبة مثل الكلمات المتجانسة والمرادفات والتعبيرات الاصطلاحية التي قد تربك عملية الترجمة

Conversion Modelنموذج التحويل: يتم إعداد بيانات النص بطريقة تمكن النموذج من التحويل من اللغة التركية إلى لغة الإشارة التركية. وتتضمن هذه العملية إعادة هيكلة النص وفقًا للغة الإشارة مع الحفاظ على معناه.

الشبكات التوليدية التنافسية (GANs): يعد ضمان سلاسة وطبيعية الانتقال بين الإشارات أحد التحديات في إنشاء فيديو سلس وسهل الفهم بلغة الإشارة. ولمعالجة هذه المشكلة نستخدم الشبكات التوليدية التنافسية (GANs) لتوليد إطارات وسيطة بين مقاطع لغة الإشارة المختلفة

وبمجرد أن يتم تحليل الجملة وتبسيطها يستخدم النظام قاعدة بيانات لمقاطع لغة الإشارة المسجلة مسبقًا لإنشاء ترجمة صحيحة نحويًا بلغة الإشارة التركية. ولضمان سلاسة وطبيعية الانتقال بين الإشارات يستخدم نظامنا الشبكات التوليدية التنافسية (GANs) لتوليد إطارات وسيطة بين العلامات مما يؤدي إلى إنتاج فيديو سلس. وتتلافى هذه العملية مظهر الرجل الآلي الذي تستخدمه العديد من أنظمة ترجمة لغة الإشارة القديمة مما يوفر نموذجًا قائمًا على الإنسان يقدم فيديو مفهوم للمستخدم النهائي. وتسمح هذه التقنية بانتقال سلسة للكلمات مما يجعل الفيديو أكثر تماسكًا وسهولة للمتابعة. وهكذا فإن استخدام الشبكات التوليدية التنافسية (GANs) يساعد في التغلب على المشاكل التقليدية المتمثلة في ظهور مقاطع فيديو لغة الإشارة غير المتصلة أو بمظهر الرجل الآلي مما يعزز تجربة المستخدم الإجمالية

نماذج لغويةوالتحليل الدلالي

نماذج لغوية: يستخدم النظام نماذج اللغة لفهم النص. ويتم تدريب هذه النماذج لفهم الاختلافات والسياقات الدلالية في اللغة التركية

التحليل الدلالي: يتم فصل معنى النص وفقًا لهياكل الجمل المختلفة. وتعد هذه عملية ضرورية لترجمة معنى اللغة بدقة إلى لغة الإشارة التركية (TID)

الاختلافات اللغوية والهياكل النحوية:

الهياكل النحوية المختلفة: قد تؤدي الاختلافات النحوية بين اللغة التركية ولغة الإشارة التركية إلى صعوبات في عملية الترجمة. ففي حين أن اللغة التركية تتبع نحويًا ترتيب - الفاعل - الفعل - المفعول به (SVO) فإن لغة الإشارة التركية عمومًا تتبع ترتيب – الفاعل – المفعول - الفعل (SOV). ويستخدم النظام خوارزميات تحويل مناسبة تأخذ هذه الاختلافات في الاعتبار

هياكل الجمل المعقدة: إنه لمن الصعب فهم هياكل الجمل المعقدة وترجمتها إلى لغة الإشارة. ولذلك يقوم النظام بإجراء تحليل السياق لتحليل معنى هذه الهياكل وتحويل الجمل إلى هياكل أبسط

قيود المفردات ولغة الإشارة:

مفردات لغة الإشارة المحدودة: قد لا تتطابق بعض الكلمات والمفاهيم في لغة الإشارة التركية بشكل مباشر مع المصطلحات الموجودة في اللغة التركية. وللتعامل مع هذه المواقف يستخدم النظام قاعدة بيانات كبيرة للغة الإشارة ويقوم بمطابقة الترجمة بناءً على أوجه التشابه الدلالية

إنتاج الفيديو وإشارات لغة الإشارة: يعد إنتاج إشارات لغة الإشارة بدقة في صيغة فيديو أمرًا مهمًا لضمان التواصل الطبيعي والسلس. ويحاكي النظام بدقة الحركات والانتقالات الطبيعية في لغة الإشارة أثناء إنتاج الفيديو

التعاون مع خبراء لغة الإشارة: بهدف ضمان أعلى مستوى من الدقة والأهمية الثقافية نقوم بالتعاون مع خبراء لغة الإشارة والمترجمين الفوريين وأبناء وبنات الصم. حيث تساعدنا رؤاهم في تحسين ترجمة النظام مما يضمن أن المحتوى الذي تم إنشاؤه بواسطة الذكاء الاصطناعي يتماشى مع تنوع وتعقيد وسياق لغة الإشارة التركية. كما يسمح لنا هذا التعاون بمعالجة التحديات المحددة داخل مجتمع الصم التركي مما يضمن تلبية نظامنا لاحتياجاتهم وتوقعاتهم الفريدة

التوسع في لغات الإشارة المتعددة: بينما ينصب التركيز الحالي على لغة الإشارة التركية فإن هذا النظام مصمم لدعم لغات الإشارة المتعددة. وتتضمن خطط التطوير المستقبلية توسيع لغات النظام لتشمل لغات الإشارة العربية والأمريكية والبريطانية والفنلندية. وسيسمح هذا النظام بتلبية احتياجات جمهور أوسع على المستوى الدولي وتوفير إمكانية نفاذ أكبر للأفراد من خلفيات لغوية وثقافية مختلفة

الخاتمة

بالنسبة لمجتمع الصم وضعاف السمع في تركيا فإن تطوير وتنفيذ أنظمة ترجمة لغة الإشارة المدعومة بالذكاء الاصطناعي يمثل تقدمًا كبيرًا في إمكانية النفاذ. وقد قمنا باستخدام معالجة اللغة الطبيعية (NLP) والشبكات التوليدية التنافسية (GANs) لإنشاء نظام يسمح بالترجمات الدقيقة في الوقت الفعلي من التركية إلى لغة الإشارة التركية [6]. و يساهم هذا النظام في سد الفجوة اللغوية التي طالما كانت عائقًا أمام النفاذ إلى المحتوى الرقمي للأفراد ضعاف السمع مما يمنحهم الفرصة للتفاعل مع المنصات الرقمية بلغتهم الأم [7]. ويتوفر البرنامج الإضافي للاستخدام في مقاطع الفيديو وملفات PDF وكذلك على مواقع الويب. كما تتوفر الترجمات للمواد المطبوعة من خلال رمز الاستجابة السريعة الذي يمكن مسحه ضوئيًا بواسطة الأجهزة.

لا يقتصر عمل أنظمة الترجمة الخاصة (SignForDeaf) على تحسين إمكانية النفاذ لثلاثة ملايين شخص من ضعاف السمع في تركيا بل إنها تمهد الطريق أيضًا ليكون بالإمكان استخدامها على مستوى العالم. فمع وجود خطط مستقبلية لإدراج لغات إشارة إضافية في النظام، يشكل (SignForDeaf) أداة لديها القدرة على إحداث ثورة في إمكانية النفاذ الرقمي للأشخاص الصم في جميع أنحاء العالم [8]. كما أن التعاون مع خبراء لغة الإشارة واستخدام تقنيات الذكاء الاصطناعي المتقدمة يضمنان بقاء نظامنا دقيقًا ومستدامًا مما يسمح بإجراء التحسينات والتحديثات المستمرة [9]. ويساهم نظام الترجمة المدعوم بالذكاء الاصطناعي في إنشاء بيئة رقمية أكثر شمولاً وكسر الحواجز أمام مجتمعات الصم وضعاف السمع وتزويدهم بفرصة متساوية للنفاذ إلى المعلومات والخدمات [10]. كما أننا نهدف إلى تحسين خوارزميات الذكاء الاصطناعي وتسريع الترجمةوالتعرف على الإيماءات الأكثر تعقيدًا في لغة الإشارة بشكل أكثر دقة مما سيجعل الترجمة أكثر دقة. وسيتم تحسين قدرات الترجمة في الوقت الفعلي باستخدام تقنيات معالجة البيانات المتقدمة والذكاء الاصطناعي. حيث سيمكن هذا الأمر للنظام بتوفير ترجمة فورية للغة الإشارة وجعل تفاعلات المستخدم أكثر سلاسة. وسيوفر نظامنا من خلال شمول هذه اللغات إمكانية نفاذ أكبر للجمهور من مختلف أنحاء العالم مما يجعل المحتوى الرقمي أكثر شمولاً في مناطق وثقافات مختلفة.

شكر وتقدير

نود أن نعرب عن امتناننا لخبراء لغة الإشارة والمترجمين الفوريين وأبناء وبنات الصم الذين ساهموا بمعرفتهم ورؤاهم القيمة طوال هذا المشروع. لقد كانت خبرتهم في لغة الإشارة التركية ضرورية لضمان دقة وأهمية نظام الترجمة المدعوم بالذكاء الاصطناعي

المراجع

1. Othman, A., Dhouib, A., Chalghoumi, H., El Ghoul, O., & Al-Mutawaa, A. (2024). The Acceptance of Culturally Adapted Signing Avatars Among Deaf and Hard-of-Hearing Individuals. IEEE Access, 12, 78624-78640. doi:10.1109/ACCESS.2024.3407128
2. Akın, E. (2020). Grammatical differences between Turkish and Turkish Sign Language. Journal of Language and Speech Research, 35(2), 122-134. doi:10.1234/jlsr.2020.35.2.122
3. Yıldırım, H. (2019). A Study on the Turkish Sign Language Dictionary. Hacettepe University Press, pp. 45-67.
4. Tuncer, F. (2021). Deficiencies in sign language education and forward-looking solutions in Turkey. Journal of Deaf Education in Turkey, 10(3), 201-215. doi:10.5678/jdet.2021.10.3.201
5. Özkan, Y. (2022). Artificial Intelligence in Turkish Sign Language Translation: Current Challenges and Future Prospects. Journal of Computational Linguistics and AI, 14(2), 150-172. doi:10.5555/jclai.2022.14.2.150
6. Kaya, M. (2020). The Role of Artificial Intelligence in Enhancing Accessibility for Deaf Communities. Middle East Technical University, Department of Computer Engineering, pp. 55-89.
7. Çetin, B., & Yılmaz, G. (2021). Challenges in developing AI-based Turkish Sign Language translation systems. Journal of Artificial Intelligence Research, 15(1), 78-92. doi:10.6789/jair.2021.15.1.78
8. Alkan, S. (2022). Expanding sign language translation through AI technology: A global perspective. International Journal of Deaf Studies, 23(4), 101-118. doi:10.9999/ijds.2022.23.4.101
9. Polat, E. (2023). Sustainability in AI-powered Sign Language Translation Systems. Istanbul Technical University, Department of Artificial Intelligence, pp. 35-68.
10. Demirel, Z. (2023). The Future of Digital Accessibility for Deaf Communities. Bosaziçi University Press, pp. 75-100.

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

آمنة دقنو
emna.daknou@supcom.tn
المدرسة العليا للاتصالات بتونس
(SUP'COM) – تونس

هيثم الهرميسي
haithem.hermessi@fst.utm.tn
المعهد العالي لعلوم الكمبيوتر –
تونس

نبيل تبان
nabil.tabbane@supcom.tn
المدرسة العليا للاتصالات بتونس
(SUP'COM) – تونس

الملخص - يستخدم الأشخاص الصم وضعاف السمع لغات الإشارة في التفاعل فيما بينهم ومع الأشخاص الآخرين. ويعد التعرف التلقائي على الإشارات غير المتماثلة الثابتة ثنائية اليدين عملية صعبة، حيث تتضمن تنفيذ نظام معالجة معقد لإدراك الصورة. وقد قدمنا في هذه الورقة مجموعة بيانات من 2000 صورة تحتوي على 12 إشارة تونسية 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

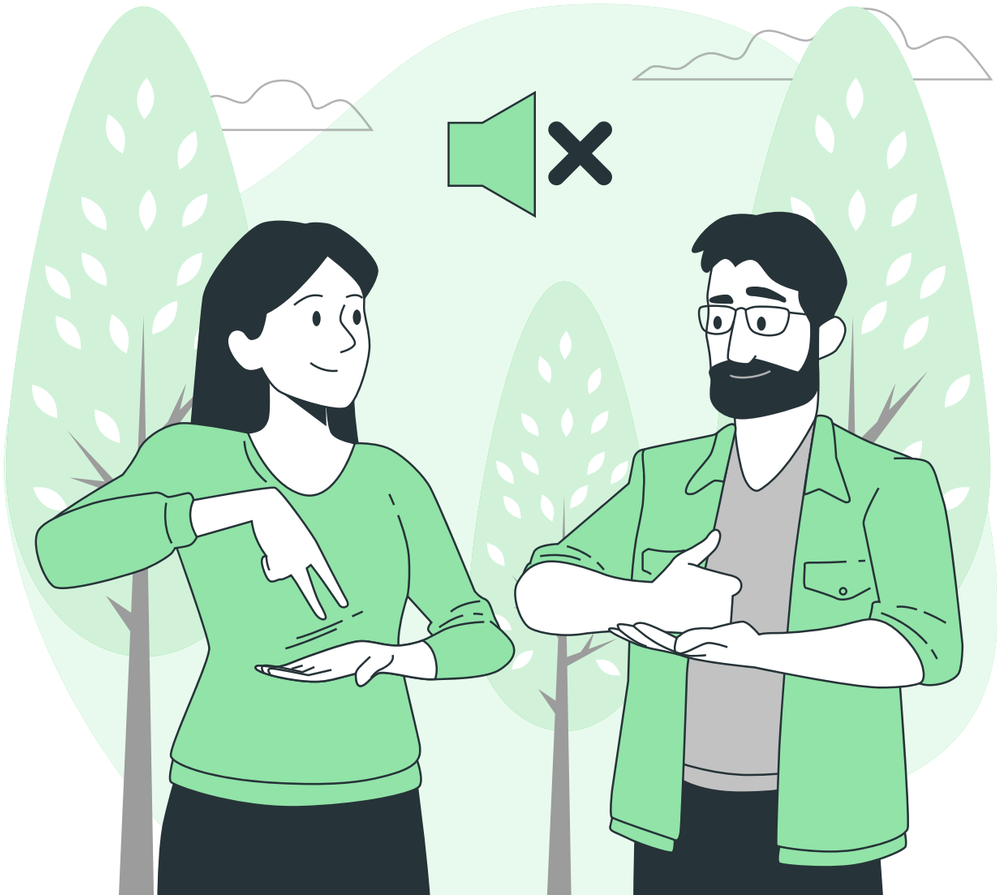
الكلمات الرئيسية
لغة الإشارة التونسية، التعلُّم الانتقالي، الإشارات غير المتماثلة ثنائية اليدين.

1. المقدمة

ارتفع عدد الأشخاص الذين يعانون من فقدان السمع وفقًا لمنظمة الصحة العالمية إلى 466 مليونًا ما يعادل 6٪ من سكان العالم. ويواجه هؤلاء حواجز تواصل كبيرة خاصة في الرعاية الصحية والتعليم والقوى العاملة والنقل. وتعد لغة الإشارة طريقتهم الوحيدة للتعبير والتواصل. 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 فقد عمل الباحثون في جميع أنحاء العالم على التخفيف من فجوة الاتصال هذه من خلال إنشاء إطار عمل التعرف الآلي على لغة الإشارة [2].

بشكل أساسي، يتم تصنيف كلمات الإشارة إلى ثلاثة أقسام على النحو التالي: 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. وقد شكل التعرف التلقائي على هذه الإيماءات مهمة صعبة 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 عصبية ترشيحية عميقة (Convolutional Neural Network) جديدة يمكنها التعرف بشكل صحيح على كلمات لغة الإشارة التونسية الثابتة غير المتماثلة ثنائية اليدين. ويستفيد إطار عملنا على وجه التحديد 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 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. وعلى الرغم من هذه النتائج الجيدة في ظل مجموعة البيانات الصغيرة فإن هذا النهج يركز على الإشارات التي يتم تنفيذها بيد واحدة فقط

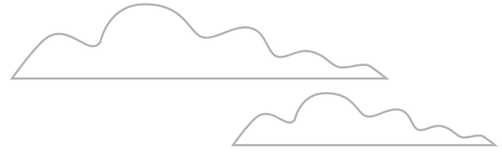
قام المؤلفون في [9] بتنفيذ نظام التعرف على الشبكة العصبية الترشيحية (CNN) لترجمة أبجدية لغة الإشارة البريطانية (BSL) بما يشمل مجموعة بيانات تضم حوالي 10000 صورة وتحتوي على 19 فئة. وتوجد في هذه الإشارات 12 إشارة غير متماثلة ثنائية اليدين. وتمر الصور بخطوات الترشيح التالية قبل التدريب: إزالة الخلفية والتحويل 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 غير متماثلة ثنائية اليد. 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 الترشيحية (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 بدقة تحقق تبلغ 94.34٪ ضمن مجموعة البيانات الصغيرة الحجم. ومع ذلك فإن جميع الإشارات المتضمنة يتم تنفيذها بيد واحدة فقط

الشكل 1. كلمات لغة الإشارة التونسية.

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 لغة الإشارة التونسية (TnSL) مع 12 فئة 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 حول تنفيذ نموذج الشبكة العصبية الترشيحية (CNN) الصحيح للتعرف الثابت على لغة الإشارة التونسية



3. المنهجية المقترحة

3.1 معالجة البيانات

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

3.1.1 جمع البيانات

نحاول بناء مجموعة بيانات للغة الإشارة التونسية تحتوي على 12 فئة من كلمات الإشارة غير المتماثلة ثنائية اليد. وفئات الكلمات هي: "قهوة"، "شاي"، "انتخابات"، "قانون"، "مساعدة"، "رقص"، "جمعية"، "سجن"، "علم النفس"، "وزارة"، "بلدية" و"حكومة". ونقوم بالتقاط الإيماءات الثابتة للصور من كاميرا ويب تحت إضاءات مختلفة وخلفية مضبوطة باستخدام وحدة معالجة الصور (OpenCV). ويوجد إجمالاً 2000 صورة حيث تحتوي كل فئة على أكثر من 160 صورة جميعها بتنسيق (RGB) بدقة عالية ومعدلة لتكون بقياس (224*224) بكسل

3.1.2 إعادة تنظيم البيانات

نظرًا لاختلاف عدد الصور في كل فئة فإن عدم التوازن بين الفئات قد يؤدي 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

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 توليد البيانات الجديدة في مجموعة التدريب. ومع زيادة حجم مجموعة التدريب والتسلسل الأكثر تنوعًا من الصور فإنه يمكن للعملية إنشاء نماذج أكثر عمومية ومهارة وتجنب مشكلة عدم الكفاءة عند التعامل مع البيانات الجديدة. وتتضمن التكوينات المطبقة: نطاق السطوع [0.5 1.2-] ونطاق التكبير [1.0، 1.2] ونطاق الدوران [°10-، °10+] ونطاق التحول الرأسي بنسبة 10% والتحول الأفقي بنسبة 10%. وبعد ذلك يتم توحيد نمط جميع الصور في مجموعة البيانات عن طريق إعادة قياس قيم البكسل هذه إلى نطاق جديد (0,1)

3.2 التعلم الانتقالي

التعلم الانتقالي هو مجال من مجالات التعلم العميق الذي يعيد استخدام نموذج تم تدريبه مسبقًا على مجموعة بيانات كبيرة وتطبيقه على موقف آخر عادةً بمجموعة بيانات صغيرة بهدف تحقيق دقة أعلى. ونعرض فيما يلي النماذج المدربة مسبقًا التي سيتم اختبارها في حالتنا:

3.2.1 (InceptionV3)

[13] (InceptionV3) 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(7 × 7) (convolutions)

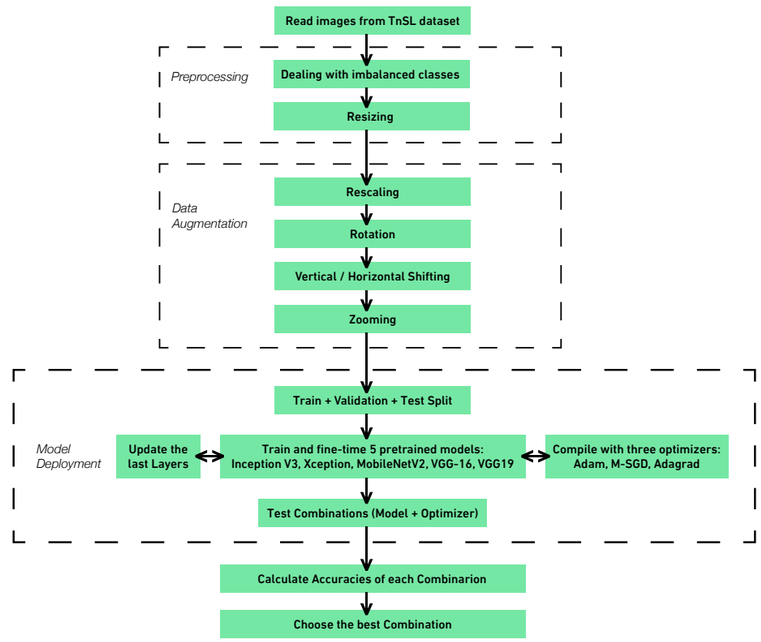
3.2.2 (Xception)

[13] (Xception) عبارة عن شبكة عصبية ترشيحية أطلقها باحثو (Google). ويعتبر نظام (Xception) مستوحى من بنية (Inception) حيث يتم استبدال (Inception) بطبقات ترشيح منفصلة حسب العمق (Depth-wise Separate Convolution Layers). ويعمل هذا الحل على تسريع عملية التصحيح ويحقق دقة أعلى بكثير مقارنة بنماذج (Inception) عند تدريبه على مجموعة بيانات (ImageNet)

3.2.3 (VGG-v16)
باعتبارها خوارزمية نقل التعلم الأكثر استخدامًا في مهام تصنيف الصور [13] تم إطلاق (VGG) بواسطة مختبر Visual Geometry Group بجامعة أكسفورد. وتعتبر هذه الخوارزمية خوارزمية ممتازة وفقًا لمعايير اليوم بفضل مرونة وبساطة بنيتها البرمجية. وهنا يجري استخدام 16 طبقة فقط بحيث يتم تنفيذ 13 طبقة ترشيح عبر مرشحات (3 × 3) مما يسمح لخوارزمية (VGG-v16) بجعل الشبكة سهلة الإدارة ويحقق أداءً قويًا

3.2.4 (VGG-v19)
باعتبارها امتدادًا لنموذج [13](VGG-v16] تحتوي (VGG-v19) على 19 طبقة بدلاً من 16. ولديها نفس بنية (VGG-v16) مع طبقات إضافية من الترشيح والتجميع الأعظمي (Max-pooling). وتعد (VGG-v19) أكثر دقة قليلاً من (VGG-v16) وفق مجموعة بيانات (ImageNet) بسبب طبقاتها الإضافية

3.2.5 (MobileNetV2)
As its name mentions, MobileNetV2 كما يشير الاسم فقد تم تصميم (MobileNetV2) للتطبيقات المحمولة [13] وهو أول نموذج رؤية حاسوبية محمول من (TensorFlow). إن ما يجعل (MobileNetV2) مميزًا هو أنه يتطلب طاقة حسابية أقل لتشغيله ووقت تنفيذ أقل مقارنة بالبنى البرمجية الأساسية الأخرى



الشكل 2. سير العمل المقترح للتعرف على لغة الإشارة التونسية.

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

3.3.1 طبقة المدخلات
يتم تغيير حجم الصور قبل تهيئة عملية التدريب إلى شكل (224,224,3) حتى يتمكن مولد بيانات الصور (ImageDataGenerator) من إدخالها إلى الشبكة

3.3.2 إضافة كتلة (Block)
نقوم في كل نموذج بإزالة بعض الطبقات المتصلة كلياً (FC-Fully Connected) 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 (Tanh) (Hyperbolic Tangent) 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 (OL) هذه وفقًا لعدد الفئات التي يجب ضبطها على 12. وتستدعي طبقة المخرجات الدالة (Softmax) للتمييز بين الإيماءات

3.4 المحسِّنات
إن المحسِّن هو وسيطة إلزامية مطلوبة لتجميع النموذج قبل عملية التدريب. وبنفس المنطق المذكور أعلاه نقوم باختيار الطرق المستخدمة بشكل شائع في الأدبيات وهي (Mini-batch Gradient Descent - M-SGD) و (Adam) و (Adagrad) لتدريب كل من النماذج الخمسة المدرجة ونختار أفضلها الذي يناسب حالتنا. ويلخص الشكل 2 مخطط التدفق العام لنهجنا المقترح

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)

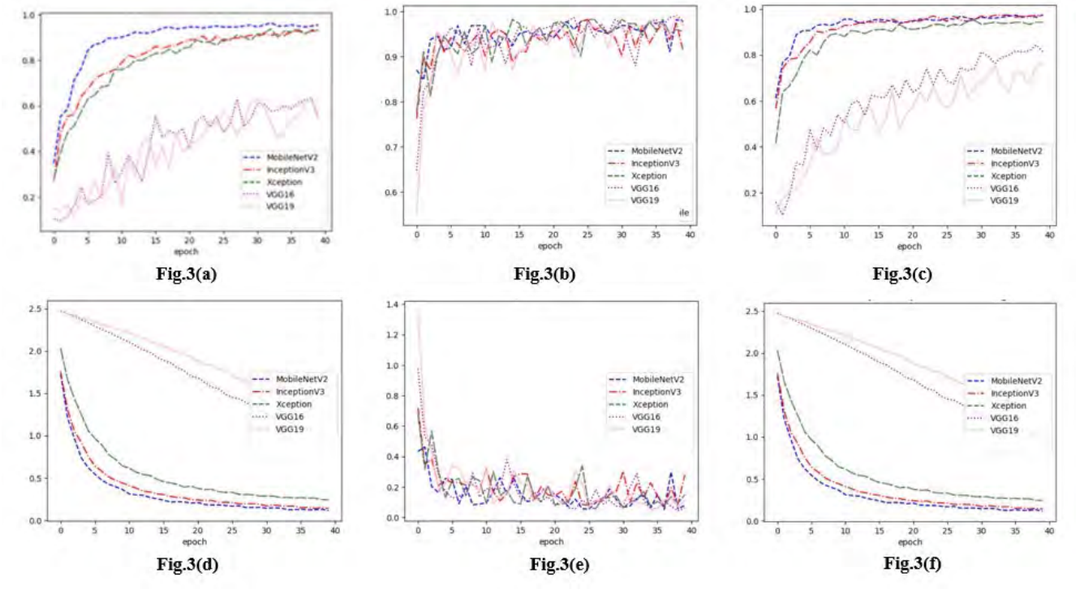
ونقوم في كل مجموعة تدريب باختيار نفس قيمة المعلمات الفائقة (hyper parameters). ونختار قيمة الدفعة 64 لضخ 64 عينة صورة في كل تكرار لعملية التدريب وذلك بعد تقييم السيناريوهات بأحجام دفعات مختلفة: 32 و64 و128. أما بالنسبة لمعدل التعلم فنختار قيمة 0.0001. ثم نضيف مدة انتظار الإيقاف المبكر في حال عدم تحسن 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 - F1 (Accuracy, Recall (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

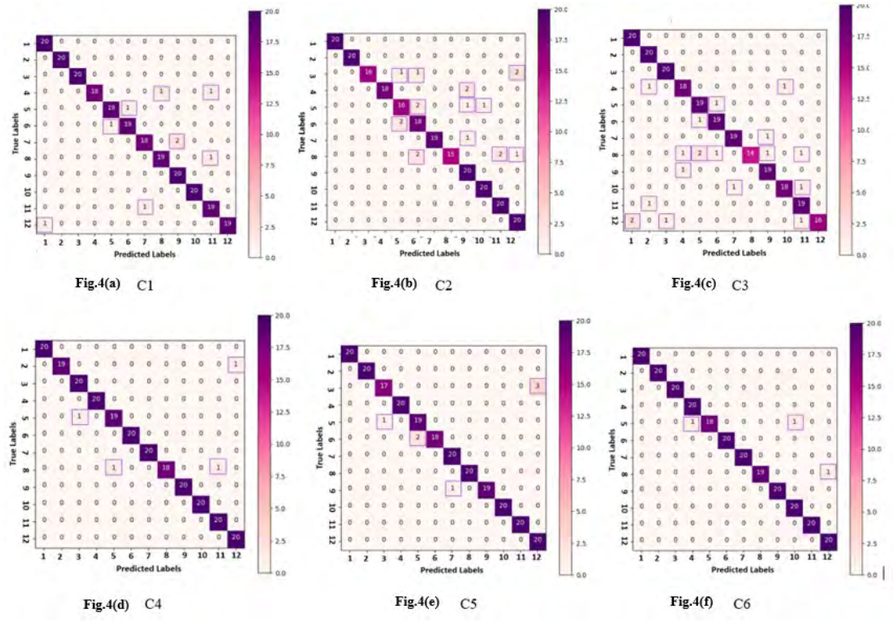
نهدف من خلال التجارب التي أجريناها في هذا القسم إلى ضبط الشبكة بأعلى دقة اختبار تقيس تعميم النموذج على البيانات غير المرئية. ويتم تنفيذ ذلك في خطوتين: أولاً بتصور تأثير المحسنات الثلاثة على النماذج المختلفة وثانيًا بتحليل مختلف عمليات التنفيذ التي تمت بواسطة هياكل الشبكة الخمسة المدربة مسبقًا

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 الصحة. ومن الواضح في الشكل 3 (ب) والشكل 3 (هـ) أن المحاولات التي تمت باستخدام المحسن (Adam) تقدم أداءً سيئًا بشكل ملحوظ لجميع النماذج المدربة مسبقًا. وتُظهر التقلبات في عمليات المرور المحلية على كامل بيانات التدريب (epochs) أن (Adam) يواجه صعوبات في التصحيح نحو حل جيد ويتخذ خيارات مختلفة في نقاط مختلفة في عملية التعلم. وهذه علامة على مشكلة عدم الكفاءة عند التعامل مع البيانات الجديدة (over-fitting) التي يحدث عندما يقدم النموذج أداءً سيئًا على البيانات غير المرئية



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



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 و C2 و C3 و C4 و C5 و C6 لتتوافق على التوالي مع: (MobileNetV2 + M-SGD)، (InceptionV3 + M-SGD)، (Xception + M-SGD)، (MobileNetV2 + Adagrad)، (InceptionV3 + Adagrad)) و(Xception + Adagrad)

4.2.2 مصفوفة الإرباك (Confusion Matrix)

نستخدم مصفوفة الإرباك لتحليل مساهمة تركيبات التعلم الانتقالي المذكورة أعلاه (C1، C2، C3، C4، C5 وC6) في التعرف على كلمات الإشارة الاثنتي عشرة في لغة الإشارة. وتوضح مصفوفة الإرباك إلى أي مدى ينجح كل من هذه تركيبات الستة في تصنيف الإشارات الاثنتي عشرة. ونظرًا لأن كل كلمة لها سماتها الخاصة فإنه يمكن أن يؤدي تركيب واحد أداءً أفضل من غيره في التعرف على عدد هذه الإشارات بينما يتكيف تركيب آخر بشكل أفضل مع الإشارات الأخرى. ونقوم بتقييم النماذج المختلفة باستخدام مقاييس أساسية مثل الدقة والتذكر ومعدل F1 كما هو موضح في الجدول 1. وتتم الإشارة في مصفوفة الإرباك إلى هذه الكلمات الاثنتي عشرة: "السجن" و"القهوة" و"القانون" و"البلدية" و"الانتخابات" و"الشاي" و"الجمعية" و"الرقص" و"المساعدة" و"الحكومة" و"الوزارة" و"علم النفس" على التوالي بأرقام من 1 إلى 12. ويتم الإشارة إلى قدرة مصنف معين على تحديد جميع التوقعات الصحيحة من خلال مقياس التذكر (recall)

وفقًا للجدول 1، فإن النماذج المدربة باستخدام مُحسِّن (M-SGD) وهي C1 و C2 وC3 تنتج أخطاء تصنيف أكثر من تلك المدربة باستخدام مُحسِّن (Adagrad). ويبلغ العدد الإجمالي للتصنيفات الخاطئة الناجمة عن C1 وC2 وC3 وC4 وC5 وC6 على التوالي 9 و21 و24 و4 و7 و3. ومن الواضح أن التركيبات + Xception (M-SGD وInceptionV3 + M-SGD) (C2

Class	Jail	Coffee	Law	Municipality	Election	Tea	Association	Dance	Help	Governorate	Ministry	Psychology	Model
Pre	0.95	1	1	1	0.95	0.95	0.95	0.95	0.91	1	0.90	1	
Re	1	1	1	0.90	0.95	0.95	0.90	0.95	1	1	0.95	0.95	C1
F1	0.98	1	1	0.95	0.95	0.95	0.92	0.95	0.95	1	0.93	0.97	
Pre	1	1	1	1	0.84	0.78	1	1	0.83	0.95	0.91	0.87	
Re	1	1	0.80	0.90	0.80	0.90	0.95	0.75	1	1	1	1	C2
F1	1	1	0.89	0.95	0.82	0.84	0.97	0.86	0.91	0.98	0.95	0.93	
Pre	0.91	0.91	0.95	0.90	0.86	0.90	0.95	1	0.90	0.95	0.86	1	
Re	1	1	1	0.90	0.95	0.95	0.95	0.70	0.95	0.90	0.95	0.80	C3
F1	0.95	0.95	0.98	0.90	0.90	0.93	0.95	0.82	0.93	0.92	0.90	0.89	
Pre	1	1	0.95	1	0.95	1	1	1	1	1	0.95	0.95	
Re	1	0.95	1	1	0.95	1	1	0.90	1	1	1	1	C4
F1	1	0.97	0.98	1	0.95	1	1	0.95	1	1	0.98	0.98	
Pre	1	1	0.94	1	0.90	1	0.95	1	1	1	1 1	0.87	
Re	1	1	0.85	1	0.95	0.90	1	1	0.95	1	1	1	C5
F1	1	1	0.89	1	0.93	0.95	0.98	1	0.97	1	1	0.93	
Pre	1	1	1	0.95	1	1	1	1	1	0.95	1	0.95	
Re	1	1	1	1	0.90	1	1	0.95	1	1	1	1	C6
F1	1	1	1	0.98	0.95	1	1	0.97	1	0.98	1	0.98	

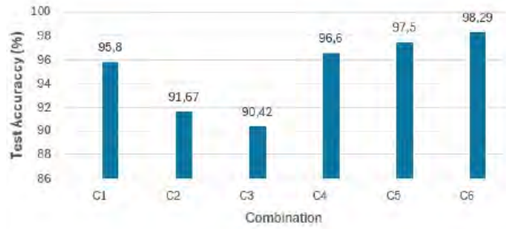
الجدول 1. مقاييس الأداء لجميع التركيبات في مجموعة الاختبار.

تقدم أسوأ أداء مقارنة بالتركيبات الأخرى وخاصة بالنسبة للإشارة "الرقص" التي تنخفض قيمة التذكر فيها إلى أقل من 0.75. كما نلاحظ في الوقت نفسه تدهورًا خطيرًا للإشارات "القانون" و"البلدية" و"الانتخابات" و"الشاي" فيما يتعلق باستخراج الميزات (feature extraction) استنادًا إلى C2. وتستمر نفس المشكلة في ظل التدريب باستخدام Xception) C3 M-SGD +) مما يؤدي إلى الكثير من التنبؤات غير الصحيحة للإشارات "البلدية" و"الحكومة" و"علم النفس" والتي وصلت قيم التذكر الخاصة بها إلى أقل من 0.95. ووفقًا للشكل 3 (أ) فإن المجموعة C1 (MobileNetV2 + M-SGD) تواجه صعوبات في تصنيف الإشارتين "البلدية" و"الجمعية" اللتين تبلغ قيمة التذكر الخاصة بهما 0.90.

وفيما يتعلق بالتركيبة C4 (MobileNetV2 + Adagrad) فهي تعمل بشكل جيد مع جميع الفئات باستثناء الفئة " الرقص" التي يخطئ فيها النموذج مرتين كما ثبت في الشكل 3 (د). وعلاوة على ذلك فإن نتائج التركيبة C5 (InceptionV3 + Adagrad) قريبة من تلك التي حصلت عليها C4 من حيث إجمالي التصنيفات الخاطئة. وعلى الرغم من أن C4 تقدم أداءً جيدًا في الشكل 3 (هـ) فإن النموذج يقدم تنبؤين غير صحيحين للإشارة "الشاي" وثلاثة تنبؤات غير صحيحة للإشارة "القانون" حيث كانت قيم التذكر الخاصة بهما هي 0.90 و0.85 على التوالي. ويبدو أن التركيبة C6 تعمل بكفاءة أكبر لأنها تحتوي على أقل عدد من التنبؤات الخاطئة. ومع ذلك فإن الإشارة "الانتخابات" غير مصنفة جيدًا في التركيبة C6. ونظرًا لوجود ميزات

معقد، فإنه يمكن التعرف على كلمة " الانتخابات " بشكل أفضل بواسطة C1 و C3 وC4 وC5.

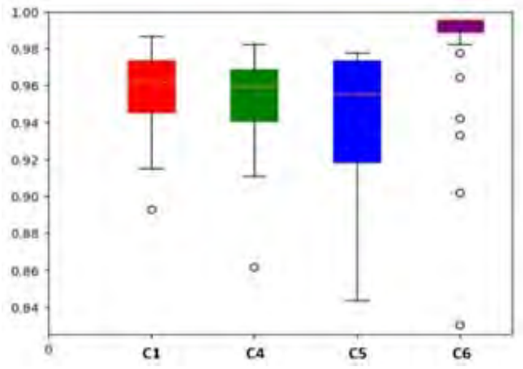
وبناءً على المنطق أعلاه فقد قررنا استبعاد التركيبة C2 وC3 والاحتفاظ بـ C1 وC4 وC5 وC6 لتصميم بنية الشبكة القادمة. وللتحقق من صحة اختيارنا نشير إلى الشكل 5 الذي يوضح قيم دقة الاختبار (Test Accuracy values) لكل تركيبة. حيث تقدم C1 وC4 وC5 وC6 قيمًا مماثلة لدقة الاختبار (95.8٪ و96.60٪ و97.5٪ و98.2٪ على التوالي) مع اختلاف طفيف بينها في حين أن C2 وC3 اللذان يبلغان 91.67٪ و90.42٪ كقيم دقة اختبار على التوالي بعيدان كل البعد عن القيمة المتوسطة. 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



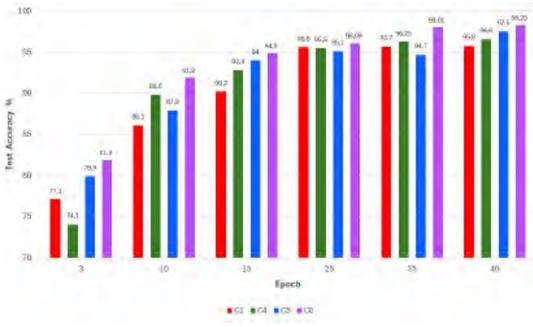
الشكل 5. دقة الاختبار لـ C1 وC2 وC3 وC4 وC5 وC6

4.3 اختيار النموذج

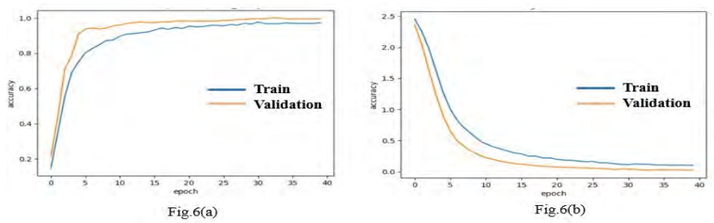
نظرًا لعدم وجود اختلاف كبير من حيث الدقة والتنبؤات غير الصحيحة بين التركيبات الأربع التي ناقشناها أعلاه فنحن بحاجة إلى تصورات إحصائية أخرى لإظهار أيهما أكثر أهلية لتصنيف لغة الإشارة التونسية. وفي هذا السياق يوضح الشكلان 6 و 7 توزيع درجات دقة الاختبار لكل تركيبة. وبالإشارة إلى الرسم البياني في الشكل 6 نرى أن انتشار درجات دقة الاختبار يضيق بشكل كبير عند التدريب باستخدام C6 (Xception + Adagrad)). وعلى الرغم من أن C5 (InceptionV2 + Adagrad) يُظهر عددًا قريبًا قليلاً من التصنيفات الخاطئة وقيمة الدقة مثل C1 (MobileNetV2 + M-SGD)، إلا أنه يظهر تبايناً كبيراً في النتائج



الشكل 6. رسم بياني يوضح دقة الاختبار لـ C1 وC4 وC5 وC6



الشكل 7. الاختلافات في دقة الاختبار بين C1 وC4 و C5 وC6 عبر عمليات المرور المحلية على كامل بيانات التدريب (epochs).



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

يظهر نموذج InceptionV3 معدلًا مرتفعًا من الاضطرابات وعدم الاستقرار في التعرف على البيانات الجديدة من مجموعة الاختبار لذلك لا يمكنه تعلم المشكلة جيدًا. وتولد تركيبة C6 نطاق انتشار أقل من C1 وC4 على الرغم من بعض القيم المتطرفة غير ذات الصلة في خطها الرأسي. إن هذه القيم ليست كثيرة بحيث يتم أخذها في الاعتبار في عملية التقييم. أيضًا. وفي الشكل 7 يوضح الرسم البياني الذي يعرض دقة الاختبار لكل تركيبة عند قيم مختلفة لعمليات المرور المحلية على كامل بيانات التدريب (epochs) (3، 10، 15، 35، 40) أن C6 تظل متقدمةً على التركيبات المتبقية (C1 وC4 وC5) في كل تكرار للتدريب. وبالتالي تثبت المحاكاة الناتجة المعروضة في الشكل 6 والشكل 7 أن نموذج Xception يعمل بشكل أفضل من النماذج الأخرى عند دمجه مع محسن (Adagrad) لأنه يحصل على أفضل معدل دقة والذي يبلغ حوالي 98.29٪.

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 C6 (Xception + Adagrad) لتصنيف إشارات لغة الإشارة التونسية.

ولإثبات أن هذه القيمة ليست عشوائية نكرر عملية التدريب ست مرات ونجمع المقاييس ذات الصلة لكل خطوة في الجدول 2. ومن الواضح أننا نلاحظ بعض التقارب في القيم التي تم الحصول عليها مما يوضح استقرار المجموعة C6 أثناء عملية التنبؤ. وفي الوقت نفسه يثبت الشكل 8 الذي يصور كل من منحنيات الدقة والفاقد المتعلقة بمجموعات التدريب والتحقق كفاءة هذا النموذج (Xception + Adagrad)

ومع ذلك يواجه هذا النموذج صعوبات في تفسير إشارة "الانتخابات" وفقًا لمصفوفة الإرباك في الشكل 4 (f) حيث يتم الخلط بين فئة "الانتخابات" وفئتي "البلدية" و"الحكومة". وقد يكون هذا نتيجة لتقنيات تعزيز توليد البيانات الجديدة عبر الإنترنت المطبقة أثناء عملية التدريب مما يؤدي إلى التشابه بين التمثيلات المجردة والميزات التي تعلمتها الشبكة العصبية الترشيحية (CNN)

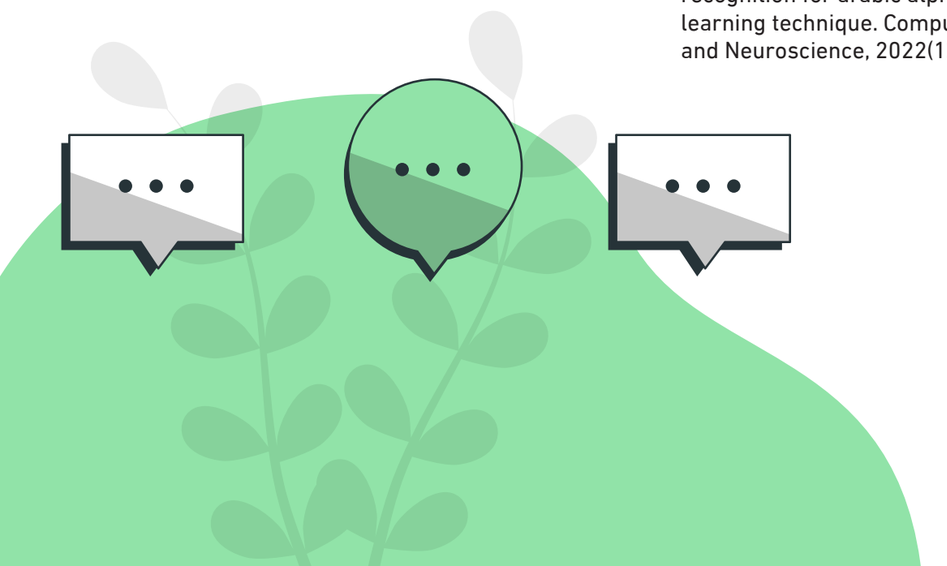
5. الخاتمة

توضح هذه الدراسة الإمكانات الكامنة لاستخدام التعلم الانتقالي للتعرف على لغة الإشارة التونسية. ويتم تطبيق طريقتنا على مجموعة بيانات لغة الإشارة التونسية (TnSL) المكونة من 2000 صورة والمجهزة بتقنية تعزيز توليد البيانات الجديدة. وقد قدم نموذج Xception أفضل قيمة دقة اختبار بنسبة 98.29% عند دمجه مع محسن (Adagrad) للتعرف على الإشارات غير المتماثلة الثابتة باليدين تحت ظل مجموعة بيانات صغيرة الحجم. إن هذا البحث هو الخطوة الأساسية نحو تطوير نظام التعرف على لغة الإشارة التونسية (TnSL) الذي يمكن أن يخدم مجتمع الصم التونسي في المواقف اليومية ويخفف من حواجز التواصل.

وسيركز العمل المستقبلي على توسيع مجموعة البيانات وتطوير أنظمة للتعرف الديناميكي على الإشارات. حيث أنه يجب توسيع مجموعة البيانات بشكل أكبر لتشمل المزيد من إشارات لغة الإشارة التونسية والنجاح في 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.



التعرف التلقائي على لغة الإشارة العربية استنادًا إلى الإيماءات:
نهج التعلم الفيدرالي (التعلم الموزع)

أحمد الزعبي
agalzubi@just.edu.jo
قسم علوم الحاسوب، جامعة العلوم والتكنولوجيا الأردنية، إربد، الأردن

أمجد البشايرة
amalbashayreh20@cit.just.edu.jo
قسم علوم الحاسوب، جامعة العلوم والتكنولوجيا الأردنية، إربد، الأردن

توفيق الحضرمي
tawfik.al-hadhrami@ntu.ac.uk
كلية العلوم والتكنولوجيا، جامعة نوتنجهام ترينت، نوتنغهام، المملكة المتحدة

لجين بني يونس
lhbaniyounis19@cit.just.edu.jo
قسم علوم الحاسوب، جامعة العلوم والتكنولوجيا الأردنية، إربد، الأردن

الملخص - يعد استخدام خوارزميات التعلم الآلي للتعرف على أنماط إيماءات اليد المخصصة للأشخاص ذوي الإعاقة اتجاهاً متنامياً في العيش المدعم. وتعالج هذه الورقة تحدي تفسير دلالات إيماءات اليد القائمة على الصور من خلال تقديم بنية التعلم العميق الموزع للتعرف على لغة الإشارة العربية. ويدير النموذج المقترح التعلم الموزع من خلال نموذج العميل والخادم حيث تتعاون العديد من العقد الطرفية لتعلم السمات التمييزية للبيانات السرية بشكل مشترك دون انتهاك خصوصيتها. وسيوفر هذا النموذج إمكانية نفاذ أكبر للأشخاص الذين يعانون من الصمم أو ضعف السمع باستخدام إيماءات الصور. ويعتمد إجراء التعلم الموزع في المقام الأول على الأساس العميق لـ(ResNet32) وآلية المتوسط التوزيعي. وتظهر النتائج التجريبية فعالية نموذج التعلم العميق المقترح حيث أنه حقق دقة بنسبة 98.30٪ بمتوسط 33 ثانية لكل عميل في جولة تدريبية واحدة. ويوضح هذا الأمر قدراته العالية في التعرف على لغة الإشارة العربية وتحسين تجربة التواصل للأشخاص ذوي الإعاقة

الكلمات الرئيسية
لغة الإشارة العربية، التعلم العميق الفيدرالي (الموزع)، التعرف على الصور، إمكانية النفاذ، صعوبات.

1.المقدمة

بما أن لغة الإشارة تُعَد الوسيلة الرئيسية للتواصل بين ملايين البشر على مستوى العالم فإن هناك حماس كبير حول الاستخدامات المحتملة لأدوات التعرف على لغة الإشارة المتقدمة (Al-Qurishi et al.,, Semreen, 2023) 2021). ونظراً للمجموعة المتنوعة من الفرص المتاحة فقد تمتد حلول التكنولوجيا المساعدة هذه إلى ما هو أبعد من مجرد الترجمة. فقد تسمح ببث نشرات بلغة الإشارة بسهولة وتعزز إنشاء أجهزة مستجيبة قادرة على تفسير أوامر لغة الإشارة بسلاسة بل وقد تقود تطوير أنظمة معقدة مصممة لمساعدة الأشخاص ذوي الإعاقة في إنجاز المهام اليومية بقدر أكبر من الاستقلالية (Othman et al., 2024).

يستخدم الأشخاص ذوو الإعاقة مثل الصم أو ضعاف السمع لغة الإشارة (SL) وهي طريقة اتصال بصرية تستخدم الإيماءات وتعبيرات الوجه وحركات الجسم. وتقوم خوارزميات التعلم العميق بالاستفادة من هياكل الشبكة العصبية العميقة في تحليل كميات هائلة من البيانات لتعلم الأنماط والسمات المعقدة المتأصلة في حركات اليد (Cui) (Rastgoo et al., 2021 et al., 2019) ومع ذلك تبقى هناك العديد من المشكلات المتعلقة بأنظمة التعرف على لغة الإشارة (SLR) القائمة على الصور وخاصة فيما يتعلق بتعقيدات تعلم السمات ومعالجة الصور وسرية المعلومات الخاصة وفعالية هذه الأنظمة في الواقع العملي. ونتيجة لذلك لا يزال من المهم للغاية الحفاظ على سرعة ودقة وموثوقية خوارزميات الترجمة (Elsheikh, (Cheok et al., 2019) (2023).

يعد التعلم العميق الموزع (FL) نموذجاً ناشئاً للتعلم الآلي مرتبط بالطرق اللامركزية، وقد أثبت أنه نهج فعال لتدريب النماذج الشاملة المشتركة (Wen et al., 2023). وتتضمن طرق التعلم الفيدرالي تنسيق تدريب نموذج مركزي من مجموعة من الأجهزة المشاركة. فعندما يتم الحصول على بيانات التدريب من تفاعلات المستخدم مع التطبيقات المحمولة على سبيل المثال يظهر سيناريو تطبيقي مهم للتعلم الفيدرالي (Lee et al., 2024). ويسمح التعلم الفيدرالي في هذا السياق للهواتف المحمولة بتعلم نموذج تنبؤ مشترك بشكل جماعي مع الاحتفاظ بجميع بيانات التدريب على الجهاز وإجراء العمليات الحسابية بشكل فعال على بياناتها المحلية لتحديث نموذج شامل. ويتجاوز هذا النهج مسألة استخدام النماذج المحلية من أجل تنبؤات الأجهزة المحمولة من خلال توفير تدريب النموذج على مستوى الجهاز. ويوفر هذا النهج في سياق التعلم الفيدرالي حلاً واعدًا لتحديات الحفاظ على الخصوصية وتنوع البيانات وقابلية تكييف النموذج(Krishnan and Manickam, 2024) (You et al., 2023).

تشتمل لغة الإشارة العربية (ArSL) على مفردات غنية وهياكل معقدة. ومثلها كمثل اللغات الأخرى فإنها تشمل مزيجاً من أشكال اليد والاتجاهات والحركة وتعبيرات الوجه لنقل معاني مختلفة (Zakariah et al., 2022). وفي حين أنه قد تم تطبيق خوارزميات التعلم العميق المختلفة للتعرف على لغة الإشارة العربية (Aldhahri et al., 2023) (Saleh and Issa et al., 2020) (Ahmed et al., 2021) (Kamruzzaman et al., 2020) (Alawwad et al., 2021) إلا أن الدراسات السابقة لم تستخدم هياكل التعلم الفيدرالي (الموزع). وقد دفعنا هذا إلى معالجة هذه الفجوة من خلال الاستفادة من نموذج التعلم العميق الفيدرالي للتعرف على لغة الإشارة العربية بشكل أكبر وضمان الخصوصية للأشخاص ذوي الإعاقة وتوفير أداء عالٍ في ظل تعقيد زمني منخفض. ويسمح هذا الأمر بتدريب النموذج محليًا على الأجهزة المحلية للمستخدمين أو الخوادم اللامركزية وحماية خصوصية المعلومات السرية. وبالتالي فإن هذا النهج يتيح التعرف على الإيماءات بشكل أكثر دقة وفي بيئات وظروف متنوعة

تم تنظيم بقية هذه المقالة على النحو التالي: يقدم القسم 2 عملية المعالجة المسبقة للصور. ويتم التعرف على المعمارية البرمجية المقترحة للنموذج القائم على التعلم الفيدرالي في القسم 3، ويقدم القسم 4 النتائج التجريبية. أما القسم 4 فيناقش إمكانية تطبيق النموذج وقابلية التوسع فيه والقضايا الأخلاقية ويختتم القسم 5 هذه الدراسة

2. إعداد صور أبجدية لغة الإشارة العربية (ARASL)

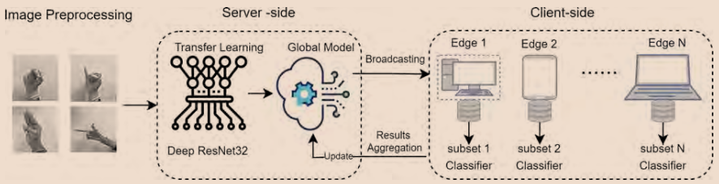
إن مجموعة البيانات المعيارية المستخدمة في هذه الدراسة هي مجموعة بيانات أبجدية لغة الإشارة العربية (ARASL) (Latif et al., 2019) والتي تتكون من 54049 صورة تصور إيماءات اليد التي تمثل الأبجدية العربية. وقد تم تصميم مجموعة البيانات هذه خصيصًا لمساعدة مجتمع الصم على فهم اللغة والتعبير عن أفكارهم وعواطفهم بحرية. وتشمل 32 فئة تمثل الحروف العربية وتحتوي كل فئة على عدد محدد من الصور. ويعرض الشكل 1 مجموعة مختارة من صور إيماءات اليد في أبجدية لغة الإشارة العربية (ARASL).

تم إجراء تغيير على بيانات أبجدية لغة الإشارة العربية بما يسمح بتغيير حجم الصورة وتحويل التينسر (tensor conversion) والتسوية ((normalization [0,1]]. وتم أخيراً تقسيم مجموعة الصور إلى 70% للتدريب و10% للتحقق و20% للاختبار. كما تم إنشاء مجموعات فرعية متعددة من صور التدريب والاختبار وهو أمر ضروري لمحاكاة عملاء لامركزيين مختلفين في إطار عمل التعلم الفيدرالي

الشكل 1. عينة من صور الإشارات العربية من مجموعة بيانات ARASL.

3. المنهجية

يوضح الشكل 2 الإطار العام لهندسة التعلم الفيدرالي المقترحة والتي تتضمن خادمًا مركزيًا يتفاعل مع عملاء متعددين يعملون كعقد حوسبة موزعة. ويستضيف الخادم نموذجًا شاملاً للتعلم العميق وهو نموذج مصمم للتدريب على البيانات المحلية للعملاء. وفي ما يخص العميل فإن كل عميل يحتفظ بمجموعة فرعية من صور إيماءات اليد العربية التي تحتوي على عينات تمت تسميتها. وبهدف الحفاظ على الخصوصية لا يشارك العملاء صور لغة الإشارة المحلية الخاصة بهم مع الخادم أو العملاء الآخرين. حيث يبث الخادم في البداية النموذج الشامل لجميع العملاء المشاركين باستخدام البيانات من كل عميل بشكل تعاوني. وتهدف هذه العملية إلى تحديد أوزان النموذج المثلى التي تقلل من معدل فاقد التصنيف لكل عميل. ويجمع الخادم نتائج التدريب على مدار عدة جولات تدريبية والتي تمثل تدرجات معلمات (معاملات) النموذج المحلي ويقوم بتحديث النموذج الشامل ثم يرسله مرة أخرى إلى العملاء



الشكل 2. مسار عملية التعلم الفيدرالي للتعرف على الإشارات العربية.

لقد استخدمت هذه الدراسة المتوسط الفيدرالي(McMahan) (FedAvg et al., 2017) لتجميع البيانات مع شبكة مكونة من خمسة عملاء باستخدام الانحدار المتدرج العشوائي الموزع (D-SGD). وتتضمن عملية التدريب 10 عمليات مرور محلية على كامل بيانات التدريب (epochs) و10 جولات شاملة مع تحديثات نموذجية متكررة. ويقوم هذا النهج بمزامنة المساهمات المحلية لكل عميل مما يؤدي إلى تحسين التصنيف الشامل لصورة إيماءة اليد. ويقوم الخادم بتحديث النموذج الشامل باستمرار بعد كل جولة وإعادة توزيع هذه التحديثات على النماذج المحلية على جانب العميل

وقد تم استخدام (ResNet32) (He et al., 2016)وهي بنية شبكة عصبية عميقة معروفة جيدًا في إطارنا الفيدرالي لتسهيل عمليات التدريب والتقييم عبر شبكة من أجهزة العملاء المشاركة. ويمكن هذا النهج تعلم انتقالي فعال من مجال عام إلى مجال لغة الإشارة العربية المحدد

4. النتائج التجريبية

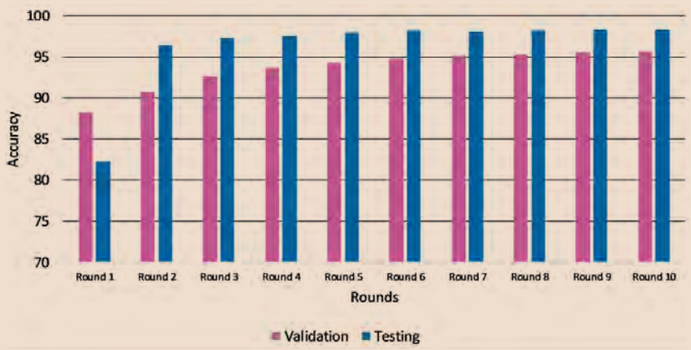
4.1. Experiments Setup

يتم إجراء العديد من التجارب بهدف تحديد المعلمات الفائقة (hyperparameters) المثلى لتقييم أداء نموذج التعلم الفيدرالي. وفي كل تجربة يقوم خمسة عملاء بإجراء عشرة عمليات مرور محلية على كامل بيانات التدريب (epochs). ويتم تجميع التدرجات باستخدام (FedAvg) من جهة الخادم بينما يتم تكوين البنية باستخدام دالة فقدان الإنتروبيا المتقاطعة الفئوية ووظيفة (SoftMax) لتصنيف الصور ومحسن (SGD) ومعدل تعلم 0.01. ويجري حساب دقة تصنيف التعرف على صور لغة الإشارة العربية. كما يتم استخدام القياسات الصحيحة والخاطئة (TP وTN وFP وFN) لحساب مقاييس التقييم القياسية مثل الدقة والتذكر ومعدل F1. وتعكس الدقة العالية فعالية النموذج في التعرف على حركات اليد المختلفة بشكل صحيح وتقليل أخطاء التصنيف

4.2. نتائج التعرف على لغة الإشارة العربية

يوضح الشكل 3 نتائج المتوسط الكلي لنموذج (FL-ResNet32) المقترح على مدار عشر جولات. ويُظهر (FL-ResNet32) أداءً عاليًا باستمرار في كل من عمليات الاختبار والتحقق حيث حقق دقة اختبار بنسبة 98.3%، ودقة بنسبة 98.28%، وتذكر بنسبة 98.26%، ودرجة F1 بنسبة 98.27%. ويتم استخدام مقاييس الدقة والمتوسط الكلي لتقييم أداء النموذج لأن مجموعة بيانات لغة الإشارة العربية غير متوازنة. ويعامل المتوسط الكلي جميع الفئات على قدم المساواة دون تفضيل الفئة المهيمنة

أما من حيث وقت التدريب فيتعرف (FL-ResNet32) بشكل فعال على لغة الإشارة العربية بدقة 98.3% في متوسط 33 ثانية على مدار 10 عمليات مرور محلية على كامل بيانات التدريب (epochs). كما يستغرق تدريب النموذج بالكامل عبر 10 جولات مع 5 عملاء موزعين (عقد طرفية) حوالي 28 دقيقة في المتوسط



الشكل 3. متوسط الدقة الكلية التي تم تحقيقها بواسطة (FL-ResNet32) على صور لغة الإشارة العربية

يقدم الجدول 1 مقارنة بين الأداء بين نموذج التعلم العميق الفيدرالي المقترح لدينا وأساليب التعرف على لغة الإشارة العربية الحالية التي تم تقييمها على مجموعة بيانات (ArASL2018). ويسلط الجدول الضوء على السمات الرئيسية ونتائج الأداء الموثقة أثناء الاختبار. ويتفوق (FL-ResNet32) كما هو موضح على الطرق الأخرى حيث يتعرف على صور لغة الإشارة العربية بشكل أكثر دقة ليحقق دقة بنسبة 98.3٪ في متوسط 33 ثانية على مدار 10 عمليات مرور محلية على كامل بيانات التدريب (epochs)
الجدول 1. مقارنة متوسط الدقة الكلية على (ArASL2018) مع الأعمال ذات الصلة

الجدول 1. مقارنة متوسط الدقة الكلية على (ArASL2018) مع الأعمال ذات الصلة

البحث المرجعي	النهج	دقة الاختبار (%)	عمليات المرور محلية على كامل بيانات التدريب
Kamruzzaman et al. (2020)	CNN	90.0	100
Aldhahri et al. (2023)	MobileNet	94.5	15
Zakariah et al. (2022)	EfficientNet-B4	95.0	30
This Work	FL-ResNet32	98.3	10

5. مناقشة

تؤكد هذه الدراسة على مدى أهمية وجود بيئات حوسبة فيدرالية (موزعة) لتمكين الاستفادة من المعلومات المتنوعة التي يمكن جمعها من أنواع مختلفة من طرفيات الحوسبة أو أجهزة العميل. إن هذه المعلومات حساسة للغاية وسرية لأنها تتعلق بالأشخاص ذوي الإعاقة. وغالبًا ما تتضمن تقنيات التعلم الآلي التقليدية تجميع البيانات على محطة عمل أو خادم واحد. ولكن نظرًا لأن الاتصال البشري حساس للغاية فيجب أن يتم التعامل مع مخاوف الخصوصية خاصة في بيئات إنترنت الأشياء (IoT). ومع ذلك فإن نقل هذه البيانات يتطلب اتصالاً بالشبكة بنطاق ترددي كافٍ لمجموعات البيانات الكبيرة وزمن انتقال منخفض لضمان التنبؤات في الوقت المناسب (Diaz et al., 2023). بالإضافة إلى ذلك تتطلب تبعية اتصالات الشبكة تقنيات تشفير متطورة لضمان خصوصية وأمان المعلومات الحساسة. ويمكن أيضًا استخدام تقنيات مثل ضغط البيانات لتعزيز كفاءة الاتصال وزيادة قابلية التوسع لأنظمة التعرف على لغة الإشارة العربية القائمة على التعلم الفيدرالي

يعد تطوير مترجم لغة إشارة قادر على تحويل لغة الإشارة إلى نص أو لغة منطوقة أمرًا بالغ الأهمية لتسهيل التفاعل بين الصم والمجتمع. ويمكن تطوير هذا المترجم من خلال مناهج تركز على الرؤية الحاسوبية الممكّنة في الأجهزة المحمولة (Talov،2022). ولا تزال هناك حاجة إلى مزيد من البحث في هذا المجال لتطوير نظام عملي وفعال لترجمة لغة الإشارة. وقد اتجهت الأبحاث والأنظمة الحديثة ذات الرؤية (Othman et al., 2024) (Othman and El Ghoul, 2022) (Bennbaia, 2022 نحو تطوير تقنيات المترجم الافتراضي للإشارة (الأفاتار) والتي تتناسب مع الثقافة المعنية. ويمكّن هذا الأمر الأفراد الصم وضعاف السمع من الانخراط في الحياة المجتمعية مما يؤدي إلى ظهور مناهج اتصال أكثر ديناميكية وقابلية للتكيف

تُعَد الشخصيات البشرية الافتراضية المعروفة أيضًا باسم (أفاتار) الترجمة أو (أفاتار) لغة الإشارة نوعًا من تقنيات المحادثة التي تستخدم تمثيلًا ثلاثي الأبعاد لشخص لإنتاج نص بأي لغة إشارة أو إشارة دولية. ويعد استخدام الشخصيات الافتراضية لترجمة لغة الإشارة أحد الحلول التفاعلية المتطورة لمشكلة النفاذ إلى محتوى لغة الإشارة. وستستفيد هذه التقنية القائمة على الشخصيات الافتراضية من التعلم الفيدرالي حيث يتماشى نموذج الاتصال في الأنظمة القائمة على لغة الإشارة بشكل جيد مع بيئة الخادم والعميل والتي تنطوي على أجهزة عميل تفاعلية مختلفة يمكنها تزويد الخادم ببيانات تدريب إضافية بتنسيقات متعددة مثل النص والصوت. إن هناك حاجة إلى مزيد من البحث للتحقيق في جدوى الحلول الذكية القائمة على الشخصيات الافتراضية للتعرف على لغة الإشارة وترجمتها في الشبكات اللامركزية واسعة النطاق. ويمكن لهذه التقنية المتقدمة أن تعزز الاتصال بشكل كبير في المدن الذكية المستقبلية

6. الخاتمة

تقدم هذه الدراسة نهجًا للتعلم العميق الفيدرالي للتعرف على لغة الإشارة العربية وتصنيفها باستخدام صور إيماءات اليد. وتتميز البنية المقترحة بأنها استراتيجية ناجحة لتحقيق دقة عالية مع الحفاظ على حماية خصوصية بيانات المرضى وهو الأمر الذي تفتقر إليه مقاربات أبجدية لغة الإشارة العربية الحالية. ويسمح نهج التعلم الموزع التعاوني هذا بتدريب النماذج بشكل فعال على الأجهزة البعيدة. وتسعى الجهود المستقبلية إلى تحسين تجربة المستخدم للتعرف على لغة الإشارة العربية من خلال واجهة مستخدم تفاعلية على الهواتف المحمولة. ويمكن أن يسهل هذا التعلم السياقي لتعبيرات الإشارة الخاصة بالأشخاص ذوي إعاقات التواصل.

المراجع

1. Ahmed, M., Zaidan, B., Zaidan, A., Salih, M. M., Al-Qaysi, Z., and Alamoodi, A. (2021). Based on wearable sensory device in 3d-printed humanoid: A new real-time sign language recognition system. Measurement, 168:108431.
2. Al-Qurishi, M., Khalid, T., and Souissi, R. (2021). Deep learning for sign language recognition: Current techniques, benchmarks, and open issues. IEEE Access, 9:126917– 126951.
3. Alawwad, R. A., Bchir, O., and Ismail, M. M. B. (2021). Arabic sign language recognition using faster r-cnn. International Journal of Advanced Computer Science and Applications, 12(3).
4. Aldhahri, E., Aljuhani, R., Alfaidi, A., Alshehri, B., Alwadei, H., Aljojo, N., Alshutayri, A., and Almazroi, A. (2023). Arabic sign language recognition using convolutional neural network and mobilenet. Arabian Journal for Science and Engineering, 48(2):2147– 2154.
5. Bennbaia, S. (2022). Toward an evaluation model for signing avatars. Nafath, 6(20).
6. Cheok, M. J., Omar, Z., and Jaward, M. H. (2019). A review of hand gesture and sign language recognition techniques. International Journal of Machine Learning and Cybernetics, 10:131–153.
7. Cui, R., Liu, H., and Zhang, C. (2019). A deep neural framework for continuous sign language recognition by iterative training. IEEE Transactions on Multimedia, 21(7):1880– 1891.
8. Diaz, J. S. P., & Garcia, A. L. (2023). Study of the performance and scalability of federated learning for medical imaging with intermittent clients. Neurocomputing, 518, 142-154.
9. Elsheikh, A. (2023). Enhancing the Efficacy of Assistive Technologies through Localization: A Comprehensive Analysis with a Focus on the Arab Region. Nafath, 9(24).
10. He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778.
11. Kamruzzaman, M. et al. (2020). Arabic sign language recognition and generating Arabic speech using convolutional neural network. Wireless Communications and Mobile Computing, 2020.
12. Krishnan, R., & Manickam, S. (2024). Enhancing Accessibility: Exploring the Impact of AI in Assistive Technologies for Disabled Persons. Nafath, 9(25).

13. Latif, G., Mohammad, N., Alghazo, J., AlKhalaf, R., and AlKhalaf, R. (2019). Arasl: Arabic alphabets sign language dataset. Data in brief, 23:103777.
14. Lee, J., Solat, F., Kim, T. Y., & Poor, H. V. (2024). Federated learning-empowered mobile network management for 5G and beyond networks: From access to core. IEEE Communications Surveys & Tutorials.
15. McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Artificial intelligence and statistics, pages 1273–1282. PMLR.
16. Othman, A., Dhouib, A., Chalghoumi, H., Elghoul, O., & Al-Mutawaa, A. (2024). The Acceptance of Culturally Adapted Signing Avatars Among Deaf and Hard-of-Hearing Individuals. IEEE Access.
17. Othman, A., & El Ghoul, O. (2022). BuHamad: The first Qatari virtual interpreter for Qatari Sign Language. Nafath, 6(20).
18. Rastgoo, R., Kiani, K., & Escalera, S. (2021). Sign language recognition: A deep survey. Expert Systems with Applications, 164, 113794.
19. Saleh, Y., & Issa, G. (2020). Arabic sign language recognition through deep neural networks fine-tuning. International Association of Online Engineering, 71-83.
20. Semreen, S. (2023). Sign languages and Deaf Communities. Nafath, 9(24).
21. Talov, M. C. (2022). SpeakLiz by Talov: Toward a Sign Language Recognition mobile application. Nafath, 7(20).
22. Wen, J., Zhang, Z., Lan, Y., Cui, Z., Cai, J., & Zhang, W. (2023). A survey on federated learning: challenges and applications. International Journal of Machine Learning and Cybernetics, 14(2), 513-535.
23. You, C., Guo, K., Yang, H. H., & Quek, T. Q. (2023). Hierarchical personalized federated learning over massive mobile edge computing networks. IEEE Transactions on Wireless Communications, 22(11), 8141-8157.
24. Zakariah, M., Alotaibi, Y. A., Koundal, D., Guo, Y., Mamun Elahi, M., et al. (2022). Sign language recognition for arabic alphabets using transfer learning technique. Computational Intelligence and Neuroscience, 2022.

أمجد السلمي
جامعة الملك فهد للبترول والمعادن

خولة باجبع
جامعة الملك فهد للبترول والمعادن

حمزة لقمان
جامعة الملك فهد للبترول والمعادن
مركز البحوث المشترك للذكاء الاصطناعي SDAIA وجامعة الملك فهد للبترول والمعادن، الظهران، المملكة العربية السعودية
{g20210113, g202115030,
hluqman}@kfupm.edu.sa

عصام لعراجي
issam.laradji@servicenow.com
ServiceNow

(التعلم باستخدام أمثلة قليلة) للتعرف على لغة الإشارة باستخدام تقنية انتشار السمات المضمنة

الملخص - إن لغة الإشارة هي القناة الأساسية للتواصل لمجتمع الصم وضعاف السمع. وتتكون لغة الإشارة من العديد من الإشارات المختلفة في أشكال اليد وأنماط الحركة ووضع اليدين والوجه وأجزاء الجسم. ويجعل هذا الأمر التعرف على لغة الإشارة (SLR) مجالًا صعبًا في أبحاث الرؤية الحاسوبية. وتعالج هذه الورقة مشكلة التعرف على لغة الإشارة باستخدام التعلم من أمثلة محدودة حيث يتم استخدام النماذج المدربة على فئات الإشارات المعروفة للتعرف على الإشارات غير المرئية باستخدام التعلم من أمثلة محدودة. فقط. ويتم في هذه الطريقة استخدام مشفر محول لتعلم السمات المكانية والزمانية لإيماءات الإشارة كما يتم استخدام تقنية انتشار السمات المضمنة (embedding propagation) لإسقاط هذه السمات في مساحة. ويتم بعد ذلك تطبيق طريقة التصنيفات (label propagation) لتشذيب التضمينات الناتجةوقد أظهرت النتائج التي تم الحصول عليها أن الجمع بين طريقتي انتشار التضمين وانتشار التسميات يعزز أداء نظام التعرف على لغة الإشارة (SLR) ويحقق دقة 76.6٪ وهو ما يتجاوز دقة الشبكة النموذجية التقليدية قليلة اللقطات والتي تبلغ 72.4٪

الكلمات الرئيسية
التعرف على لغة الإشارة،
ترجمة لغة الإشارة، التعلم قليل الأمثلة.

1 المقدمة

تمثل لغة الإشارة الوسيلة الرئيسية للأشخاص الصم أو ضعاف الصوت للتواصل وتبادل المعرفة والتعبير عن مشاعرهم وبناء علاقات اجتماعية مع الآخرين (1). ومع تقدم التكنولوجيا أصبح بإمكان الأشخاص الذين يعانون من ضعف السمع والصمم التواصل مع مجتمعهم بكفاءة أكبر من خلال ترجمة لغة الإشارة إلى لغات طبيعية والعكس (2)

يعد التعرف على لغة الإشارة (SLR) أحد أكبر المشكلات التي يتم تناولها في مجال الرؤية الحاسوبية [3]. فعلى الرغم من أن معظم الإشارات لها مظهر محدد بوضوح إلا أنها تختلف قليلاً عن بعضها البعض بصريًا (4؛ 5). وهكذا فإنه لكي يشكل التعرف على لغة الإشارة تقنية شاملة فإنه يتطلب تقدمًا أساسيًا في نمذجة وتحديد الأنماط المكانية الزمنية الدقيقة لحركات اليد [3]. وهناك أيضًا عوامل أخرى تؤثر على أداء مهمة التعرف على لغة الإشارة بما في ذلك الاختلافات في منظور الرؤية (6) وتطور لغات الإشارة بمرور الوقت [7]) والاختلافات الإقليمية في لغة الإشارة (8)

يمكن تصنيف تقنية التعرف على لغة الإشارة إلى نوعين معزول ومستمر. حيث تستهدف أنظمة التعرف المعزول على لغة الإشارة إشارات من مستوى الكلمة في حين تتعرف مقاربات التعرف المستمر على لغة الإشارة على جمل لغة الإشارة (9). وقد تمت دراسة تقنية التعرف المعزول على لغة الإشارة على نطاق واسع

في الأدبيات مقارنةً بتقنية التعرف المستمر (2). وكانت إحدى المشكلات الرئيسية في هذه الطرق هي الحاجة إلى عدد كبير من العينات الموضحة لكل إشارة (10) (11) (12). حيث يجب جمع عينات موضحة لجميع الإشارات في جميع اللغات ذات الاهتمام لتلبية هذه الحاجة. ويجب أن تتضمن هذه العينات إشارات يتم التعبير عنها عدة مرات من قبل أفراد في إعدادات تسجيل مختلفة. يتم التحدث بأكثر من 140 لغة إشارة على مستوى العالم جنبًا إلى جنب مع العديد من اللهجات (13). وبالتالي فإن الطلب على الأمثلة الخاضعة للإشراف يقف عائقاً أمام توسيع نطاق التعرف على لغة الإشارة. وقد حاولت بعض الحلول في الآونة الأخيرة التغلب على هذه المشكلة باستخدام التعلم قليل الأمثلة للتعرف على الإشارات غير المرئية مع عدد قليل من العينات ذات التسميات التوضيحية (14؛ 15؛ 16؛ 3). ويشكل التعلم قليل الأمثلة تقنية لتعلم التمييز بين الفئات من خلال عدد محدود من العينات أو الأمثلة ذات التسميات

نقدم في هذه الورقة طريقة التعلم قليل الأمثلة للتعرف على لغة الإشارة وهي طريقة مصممة خصيصًا ليمكن تعميمها على الفئات غير المرئية سابقاً. وتقبل هذه الطريقة معلومات وضعية إيماءات الإشارة وتغذيها في مشفر المحول لاستخراج مجموعة من السمات التي تشفر المعلومات المكانية والزمانية. ثم ننقل هذه السمات من مساحة السمات إلى مساحة التضمين من خلال الاستفادة من تقنية انتشار التضمين (embedding propagation) و تقنية انتشار التسميات (label propagation) معاً. وقد تم تقييم هذه الطريقة المقترحة باستخدام مجموعة بيانات (WLASL-100)، وتوضح النتائج التي تم الحصول عليها فعالية الجمع بين هاتين التقنيتين في مجال التعلم قليل الأمثلة للتعرف على لغة الإشارة

لقد تم ترتيب هذه الورقة على النحو التالي. يبدأ القسم 2 بمراجعة الأدبيات ذات الصلة. ثم نقدم في القسم 3 طريقة التعلم قليل الأمثلة للتعرف على لغة الإشارة ويتم تقديم العمل التجريبي في القسم 4. وأخيراً يتم تقديم استنتاجاتنا وعملنا المستقبلي في القسم 5

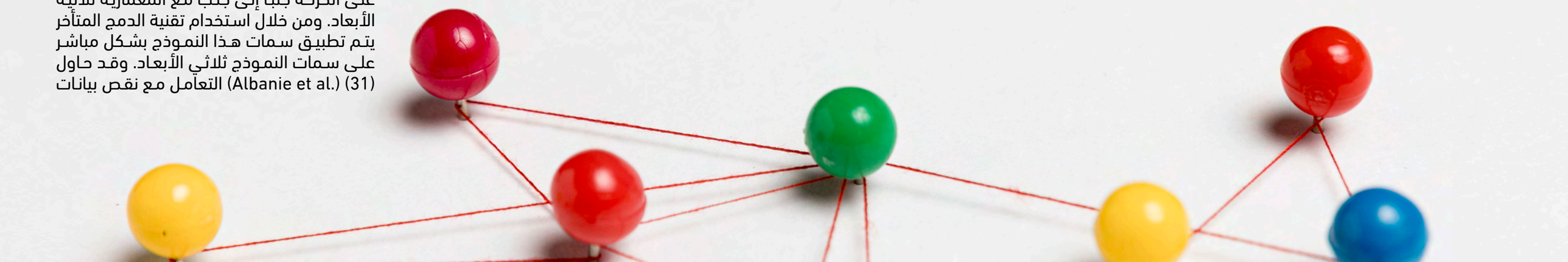
2 الأدبيات ذات الصلة

التعرف على لغة الإشارة (SLR): لقد تم تطوير العديد من التقنيات في العقدين الماضيين للتعرف على إيماءات لغة الإشارة (1؛ 2). وتركز غالبية هذه التقنيات بشكل أساسي على تتبع والتعرف على أيدي مستخدمي الإشارة (17؛ 18؛ 19؛ 20). حيث تمثل حركة اليدين الجزء اليدوي من لغة الإشارة بينما تمثل حركات الجسم وتعبيرات الوجه الجزء غير اليدوي من لغة الإشارة. ولم يحاول التعرف على الإشارات اليدوية وغير اليدوية في وقت واحد (21؛ 22؛ 23) سوى عدد قليل من الدراسات

كانت هناك عدة محاولات لتطوير أساليب التعرف على لغة الإشارة القائم على التعلم العميق في السنوات الأخيرة. حيث اقترح Camgoz et al.) (24 نموذجًا قائمًا على استخدام المحول للترجمة والتعرف المستمر على لغة الإشارة. وهنا يتم تعلم المعلومات الزمنية لإشارات الجملة بطريقة موحدة باستخدام فاقد التصنيف الزمني الاتصالي أو الارتباطي (CTC). وقد اقترحت دراسة سابقة (25) محولًا تقدميًا لترجمة جمل الكلام المنفصلة إلى تسلسلات مستمرة من التعبيرات ثلاثية الأبعاد. وقد استخدم (Tao et al.) (26) في هذا العمل زيادة تعزيزية لأبجدية الإشارة الأمريكية لمعالجة التشوهات (occlusions) غير المكتملة وتقليل تأثير تغييرات منظور الرؤية. ويتم بعد ذلك إرسال الصور المعززة الناتجة إلى الشبكات العصبية الالتفافية (CNN)بسيطة. وقد تم استخدام الشبكة الشبكات العصبية الالتفافية (CNN) في دراسة أخرى (27) لجمع العديد من التشكيلات المكانية والطيفية لصور إيماءات اليد لتوفير طريقة للكشف البصري عن هجاء الأصابع (fingerspelling) في الإيماءات. وتخلق هذه الطريقة المقترحة صورًا مكانية زمنية لحركات اليد في صيغ طيف (Gabor) ثم تستخدم الشبكة العصبية التلافيفية (CNN) المحسنة لتصنيف الإيماءات في مساحة مشتركة إلى فئات مناسبة

تم اقتراح (SAM-SLR) للتعرف على لغة الإشارة وهو إطار عمل متعدد الوسائط يتعرف على "بنية جسم الإنسان كوسيلة لاستغلال المعلومات متعددة الوسائط (28). وقد استخدم (Huang et al) شبكة عصبية تلافيفية ثلاثية الأبعاد لتعلم الجوانب المكانية والزمانية لإيماءات الإشارة (29). كما تم استخلاص مجموعة من السمات من أيدي المؤشر لتسليط الضوء على التغييرات المهمة في حركات اليد. وتم استخدام مجموعة بيانات تتكون من 25 إشارة لتقييم النهج المقترح والإبلاغ عن الحصول على دقة بنسبة 94.2٪. وفي عمل آخر تم تطوير نظام مختلف للتعرف على أبجدية لغة الإشارة وتم الإبلاغ عن الحصول على دقة بنسبة 98.9٪ (29)

واقترح المؤلفون في (30) نموذجًا لتقنية التعرف المعزول على لغة الإشارة باستخدام صور منتجة من إطارات ملونة تغطي تاريخ الحركات. وقد تم استخدام هذه التقنية لتلخيص المعلومات المكانية الزمنية لكل إشارة. كما تم تنفيذ نموذج يقبل صور (بنظام الألوان الأحمر والأخضر والأزرق) (RGB) وتاريخ الحركة كوحدات انتباه مكاني تعتمد على الحركة جنبًا إلى جنب مع المعمارية ثلاثية الأبعاد. ومن خلال استخدام تقنية الدمج المتأخر يتم تطبيق سمات هذا النموذج بشكل مباشر على سمات النموذج ثلاثي الأبعاد. وقد حاول (Albanie et al.) (31) التعامل مع نقص بيانات

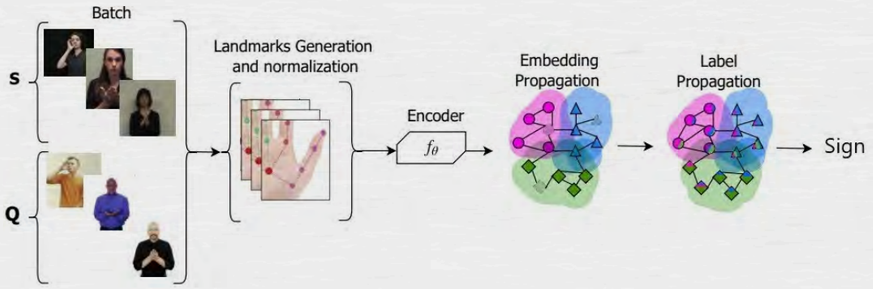


لغة الإشارة ذات التسميات التوضيحية من خلال التعرف على الكلمات الرئيسية في بث تلفزيوني لمعالج. ففي 1000 ساعة من الفيديو يتم تلقائيًا ترجمة 1000 إشارة من خلال الترجمات النصية المتوافقة زمنياً بشكل ضعيف إضافة إلى تحديد الكلمات الرئيسية الهامة. وقد قدم المؤلفون في (32) إطارًا متكاملًا للتعلم متعدد الأمثلة في أفلام لغة الإشارة.

التعلم قليل الأمثلة للتعرف على لغة الإشارة: وعلى النقيض من الطرق التقليدية التي تحتاج للإشراف للتعرف على لغة الإشارة تتعرف الطرق القائمة على التعلم قليل الأمثلة على فئات الإشارات غير المستكشفة إما من عدد قليل جدًا من عينات التدريب (أمثلة) أو بدون عينات تدريب بصرية. ويعد Cornerstone Network (CN) نموذج تعلم قليل الأمثلة اقترحه (14) ويمكنه التقليل من تأثير عينات الدعم في الظروف غير المناسبة. ويتم في هذه الشبكة استخلاص المتوسطات غير الدقيقة من عينات الدعم من عينات الإدخال واستخدامها كسمات إدخال. ويتم بعد ذلك استخدام الشبكات العصبية مع خوارزميات التجميع لتعلم المخطط من مساحة الإدخال إلى مساحة التضمين. وكما هو الحال مع الشبكات العصبية السيامية فقد تم تدريب شبكة استخراج السمات بنفس الطريقة بحيث يتم توزيع السمات من البيانات غير المتجانسة على أوسع نطاق ممكن. وبالمثل فقد قام (15) (Shovkoplias et al) بتقصي العديد من طرق التعلم قليل الأمثلة مثل التعلم المستقل عن النموذج والتعلم الفوقي وشبكات المطابقة والشبكات النموذجية لتصنيف تسجيلات مخطط كهربية العضلات لإيماءات الصم والبكم. وقد استخدم المؤلفون في (16) متنبئًا ذا نقطة رئيسية وهو مدرب مسبقاً للاحتفاظ فقط بالمعلومات المتعلقة بالجسم واليد والوجه وتجاهل المناطق الأخرى. ويسمح هذا الأمر بمقارنة أفضل بين تضمينات المتجهات (vector) حيث يتم تعلم التمثيلات الغنية من تسلسلات النقاط الرئيسية للجسم. ويتم تصنيف متجه الإدخال الجديد من خلال مقارنة مدى بعده عن بعض الأمثلة لكل فئة باستخدام خوارزمية (أقرب جيران k-nearest neighbors) (k) وتشابه جيب التمام والشبكات النموذجية.

قام (3) (Bilge et al.) بتطبيق التعلم من الصفر لتصنيف إيماءات لغة الإشارة دون أي عينات ذات تسميات توضيحية. حيث يتم إنشاء تمثيلات الفئة الدلالية من أوصاف الإشارات النصية المتوفرة بسهولة والمستمدة من قواميس لغة الإشارة. وتُستخدم هذه التمثيلات لرسم خرائط للإشارات أثناء الاستدلال على فئاتها المقابلة. وبالمثل يتم استخدام إطار التعلم من الصفر لتطوير نماذج مكانية زمنية لمناطق الجسم واليد باستخدام تمثيلات الفئة الدلالية (33). كما تم استخدام وسائط (RGB) والعمق في هذه الدراسة. وتتضمن هذه الطريقة نموذجين لمحول الرؤية يحددان أجزاء الجسم ويقسمانها إلى 9 أجزاء. ليتم بعد ذلك استخلاص مجموعة من السمات المرئية بواسطة المحول الثاني

3 المنهجية

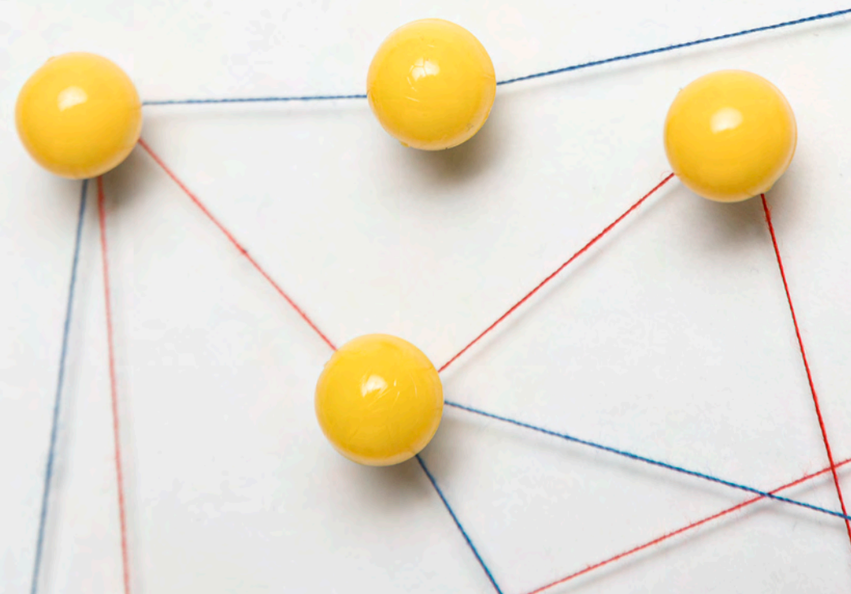


الشكل1: الإطار المقترح: تم أخذ تمثيلات تقنية انتشار التضمين وتقنية انتشار التسميات من (34)

نقدم في هذا القسم نظرة عامة على مسار العملية المقترح كما هو موضح في الشكل 1. حيث تمزج هذه العملية مُشفِّر المحول (35) مع تقنية انتشار التضمين (34). ويقوم مُشفِّر المحول في البداية باستخلاص السمات من كل حركة إشارة. ليتم بعد ذلك تعيين هذه السمات للتضمينات عبر مكون انتشار التضمين. ثم نقوم بتقييم طريقتين لصقل التضمين وهما انتشار التسميات والشبكة النموذجية. ويتم أخيرًا إدخال التضمينات المكررة في مُصنِّف لتصنيف كل إشارة إلى التسمية المقابلة لها.

3.1 نموذج المحول

يتم استخدام نموذج قائم على المحول تم اقتراحه بواسطة (35) في عمليتنا كمستخلص للسمات لتعلم تمثيلات وضعية الجسم. ويتم استخلاص السمات باستخدام مشفر المحول بينما يتم استبدال المشفر بمكون انتشار التضمين. ويخضع كل إطار فيديو لمرحلة تقدير الوضعية قبل المعالجة ويتم تحديد معالم الرأس والجسم واليد. وبهدف تجنب مشكلة صعوبة النموذج في تعميم البيانات الجديدة وتعزيز تعميمه على مختلف البيانات يتم تعزيز البيانات الهيكلية أثناء التدريب عبر التقنيات المقترحة في (35). حيث يتم تدوير كل إحداثيات المفصل في كل إطار بشكل عشوائي حتى زاوية 13 درجة. ثم يتم تحويل إحداثيات المفصل هذه إلى مستوى جديد مما يعطي الفيديو مظهرًا مائلًا. ويتم بعد ذلك تدوير المعلم (landmark) نسبياً مقابل المعلم الحالي أثناء مروره عبر النقاط الرئيسية لكلا اليدين. وبعد ذلك تتم إزالة السمات المكانية غير ذات الصلة إلى حد كبير عن طريق تطبيع نسب جسم المؤشر ومسافة الكاميرا وموقع الإطار مما ينتج متجه (vector) من أوضاع الجسم الطبيعية كمدخل للنموذج. ويتكون متجه (vector) وضعية كل إطار من 54 موقعًا مفصلياً يتم ترميزها بعد ذلك بمعلومات مواقعها. ويتم استخدام الترميز المكتسب بأبعاد 108 ويضاف حسب عناصره إلى متجه (vector) الوضعية. ثم يتم تغذية تسلسل الإدخال في طبقات ترميز المحول ويمر عبر وحدة الانتباه الذاتي وشبكة تغذية أمامية من طبقتين. وتتكون وحدة الانتباه الذاتي من تسعة رؤوس وست طبقات ترميز



3.2 انتشار التضمين

يعد انتشار التضمين تقنية لرسم خريطة للسمات في مجموعة من السمات المتداخلة تسمى التضمينات. وقد استخدمنا في هذا العمل تقنية انتشار التضمين المقترحة في (34). وتنقل هذه التقنية سمات الإدخال المستخلصة باستخدام مُشفِّر المحول إلى البيانات العرضية. ثم تنتج مجموعة من التضمينات z˜i في خطوتين. أولاً يتم حساب المسافة لكل زوج من السمات (i, j) على أنها (d2 = zi − zj2) ومصفوفة التجاور على أنها exp = (-d2 /o2) حيث o2 هو عامل للقياس وAii = 0 لجميع i. وبعد ذلك يتم حساب لابلاس لمصفوفة التجاور على النحو التالي:

$$L = D^{-1/2} * AD^{-1/2}, D_{ii} = \sum_j A_{ij} \quad (1)$$

ومن ثم يتم الحصول على التضمينات المحسَّنة على النحو التالي

$$P = (I - \alpha L)^{-1} \quad (2)$$

حيث I هي مصفوفة تحديد وR ϵ α هو عامل للقياس ويتم حساب التضمينات النهائية على النحو التالي

$$z_i = \sum_j P_{ij} Z_j \quad (3)$$

يؤدي انتشار التضمين إلى إزالة الضوضاء غير المرغوب فيها من متجهات السمات نظرًا لأن (Z_i) أصبحت الآن مجموعًا مرجحًا لجيرانها. وبهدف إجراء صقل متعدد على التضمين الناتج فقد قمنا بتقييم تقنيات انتشار التسميات والشبكة النموذجية (36). ويتم إجراء عمليات تحسين النموذج وتصنيفه على مخرجات تقنية الصقل

4. العمل التجريبي

مجموعة البيانات. لقد استخدمنا مجموعة بيانات لغة الإشارة الأمريكية على مستوى الكلمات (WLASL) لتدريب وتقييم نهجنا المقترح (37). وهي مجموعة بيانات للغة الإشارة الأمريكية تتألف من 100 إيماءة إشارة مميزة يؤدي كل منها العديد من المؤشرين مع قيام أكثر من ثلاثة مؤشرين بتنفيذ كل إشارة. وتتضمن مجموعة البيانات معلومات عن الوضعية المتخذة لأداء جميع الإشارات. وقد قسمنا البيانات في عملنا إلى ثلاث مجموعات: مجموعة أساسية تحتوي على 90 إيماءة ومجموعة تحقق تحتوي على 5 إيماءات ومجموعة فئة جديدة تحتوي على 5 إيماءات. وتم استخدام المجموعة الأساسية ومجموعة التحقق أثناء مرحلة ما قبل التدريب بينما تم استخدام المجموعة الجديدة أثناء مرحلة الاستدلال. وقمنا أثناء الاستدلال بتقسيم المجموعة الجديدة إلى مجموعات دعم واستعلام

إعداد التجارب. يتم تحسين النماذج باستخدام مُحسِّن (SGD) أثناء مرحلة التدريب بمعدل تعلم 0.0001 وقد تم اختياره تجريبيًا. وفي كل مرة يصل فيها النموذج إلى مرحلة ثبات وهو ما يحدث عندما لا ينخفض فاقد أو خسارة التحقق لمدة 10 عمليات مرور كاملة للبيانات نقوم بتقليل معدل التعلم بعامل 10.

الجدول 1: دقة التعرف للنظام المقترح مع عدد مختلف من العينات في مجموعة العينات الداعمة. أعلى دقة مكتوبة بخط غامق وثاني أعلى درجة مكتوبة بخط مسطر

النتائج والمناقشة. قمنا بتقييم النموذج المقترح باستخدام تكوينات مختلفة من خلال تغيير عدد العينات في مجموعة العينات الداعمةوتوضح النتائج في الجدول 1 تأثير انتشار التضمين على أداء النموذج في التعرف على لغة الإشارة مع توفر عينات محدودة. كما قمنا بتقييم مكونات النظام مع انتشار التضمين وبدونه لتسليط الضوء على فعالية هذه الطريقة. وكما هو موضح في الجدول فقد تم تحقيق دقة بنسبة 76.6٪ باستخدام طريقة انتشار التسميات جنبًا إلى جنب مع طريقة انتشار التضمين مقارنة بنفس الإعدادات بدون انتشار التضمين. وتم الحصول على ثاني أعلى معدل دقة 76.0٪ باستخدام الشبكات النموذجية مع انتشار التضمين مما يمثل تحسنًا بنحو 11٪ على نفس الإعدادات بدون استخدام طريقة انتشار التضمين

ومن الواضح أيضًا أن كل من تقنيات الصقل وانتشار التسميات والشبكات النموذجية قد أدت دورها بشكل فعال مع نموذج المحول باستخدام عدد صغير من العينات في مجموعة العينات الداعمةوعلى الرغم من أن زيادة عدد العينات قد أدي بشكل عام إلى تحسين أداء جميع التقنيات إلا أن بعض النماذج واجهت مشكلة عدم الكفاءة في التعامل مع البيانات الجديدة وهو ما قد يفسر انخفاض الأداء عند استخدام 10 عينات في مجموعة العينات الداعمة

Support set size	Without Embedding Propagation		With Embedding Propagation	
	Label Propagation	Prototypical Networks	Label Propagation	Prototypical Networks
1	72.2	67.2	70.8	68.6
5	72.4	73.4	**76.6**	72.2
10	69.8	65.4	68.8	<u>76.0</u>

5. الخاتمة

اقترحنا في هذه الورقة طريقة للتعلم قليل الأمثلة للتعرف على لغة الإشارة وهي طريقة مصممة للتعميم بشكل فعال على الفئات غير المرئية سابقاً. وتقوم طريقتنا بربط السمات في مساحة الإدخال بمساحة التضمين باستخدام انتشار التضمين جنبًا إلى جنب مع تقنيات انتشار التسميات. ويتم في البداية استخراج سمات إيماءة الإشارة من إطارات الإدخال باستخدام مشفر المحول. ثم يتم تعيين هذه السمات على مساحة التضمين من خلال طريقة انتشار التضمين متبوعة بانتشار التسمية لصقل هذه التضمينات. لقد قمنا بتقييم الطريقة المقترحة باستخدام مجموعة بيانات (WLASL-100) حيث توضح النتائج التجريبية تفوق الجمع بين انتشار التضمين وانتشار التسميات مقارنة بالشبكة النموذجية. وبالنسبة للعمل المستقبلي فنحن نخطط لتقييم طريقتنا وفق مجموعات بيانات لغة الإشارة المختلفة لتقييم قدرتها على التعميم على الفئات غير المرئية سابقاً بشكل أكبر.

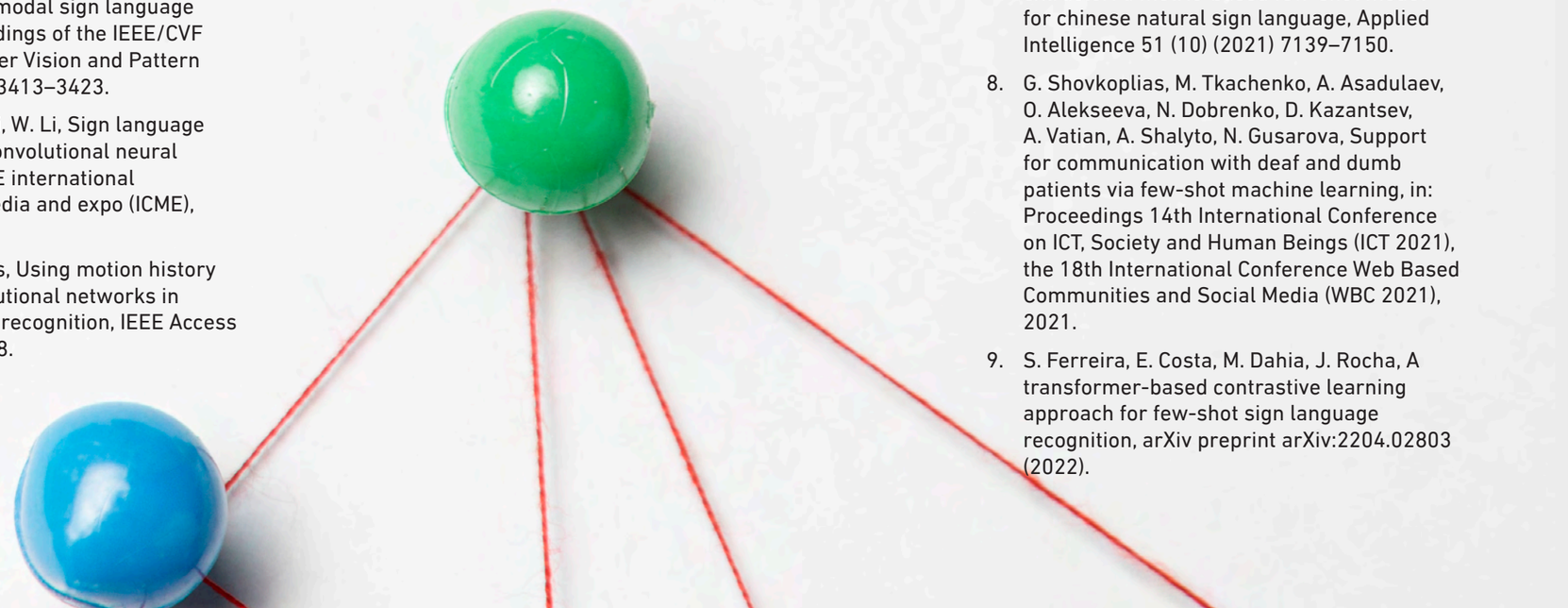
شكر وتقدير

يود المؤلفون أن يعربوا عن تقديرهم للدعم الذي تلقوه من الهيئة السعودية للبيانات والذكاء الاصطناعي (SDAIA) وجامعة الملك فهد للبترول والمعادن (KFUPM) في إطار منحة مركز أبحاث الذكاء الاصطناعي المشترك بين (SDAIA) وجامعة الملك فهد للبترول والمعادن رقم JRC-AI-RFP-14.

المراجع

1. E.-S. M. El-Alfy, H. Luqman, A comprehensive survey and taxonomy of sign lan- guage research, Engineering Applications of Artificial Intelligence 114 (2022) 105198.
2. S. Alyami, H. Luqman, M. Hammoudeh, Reviewing 25 years of continuous sign language recognition research: Advances, challenges, and prospects, Information Processing & Management 61 (5) (2024) 103774.
3. Y. C. Bilge, R. G. Cinbis, N. Ikizler-Cinbis, Towards zero-shot sign language recog- nition, IEEE Transactions on Pattern Analysis and Machine Intelligence (2022) 1–1doi:10.1109/TPAMI.2022.3143074.
4. Y. Wu, T. S. Huang, Vision-based gesture recognition: A review, in: International gesture workshop, Springer, 1999, pp. 103–115.
5. A. a. I. Sidig, H. Luqman, S. A. Mahmoud, Arabic sign language recognition using optical flow-based features and hmm, in: Recent Trends in Information and Com- munication Technology: Proceedings of the 2nd International Conference of Reli- able Information and Communication Technology (IRICT 2017), Springer, 2018, pp. 297–305.
6. C. Neidle, A. Thangali, S. Sclaroff, Challenges in development of the american sign language lexicon video dataset (asllvd) corpus, in: 5th workshop on the represen- tation and processing of sign languages: interactions between corpus and Lexicon, LREC, Citeseer, 2012.
7. C. Lucas, R. Bayley, Variation in sign languages: Recent research on asl and beyond, Language and Linguistics Compass 5 (9) (2011) 677–690.

1. C. Valli, C. Lucas, Linguistics of American sign language: An introduction, Gal- laudet University Press, 2000.
2. [9] R. Rastgoo, K. Kiani, S. Escalera, Sign language recognition: A deep survey, Expert Systems with Applications 164 (2021) 113794.
3. N. Cihan Camgoz, S. Hadfield, O. Koller, R. Bowden, Subunets: End-to-end hand shape and continuous sign language recognition, in: Proceedings of the IEEE inter- national conference on computer vision, 2017, pp. 3056–3065.
4. N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, R. Bowden, Neural sign language translation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7784–7793.
5. S. Stoll, N. C. Camgöz, S. Hadfield, R. Bowden, Sign language production using neural machine translation and generative adversarial networks, in: Proceedings of the 29th British Machine Vision Conference (BMVC 2018), British Machine Vision Association, 2018.
6. M. P. Lewis, F. Gary, Simons, and charles d. fennig (eds.). 2013. ethnologue: Lan- guages of the world (2015).
7. F. Wang, C. Li, Z. Zeng, K. Xu, S. Cheng, Y. Liu, S. Sun, Cornerstone network with feature extractor: a metric-based few-shot model for chinese natural sign language, Applied Intelligence 51 (10) (2021) 7139–7150.
8. G. Shovkoplias, M. Tkachenko, A. Asadulaev, O. Alekseeva, N. Dobrenko, D. Kazantsev, A. Vatian, A. Shalyto, N. Gusarova, Support for communication with deaf and dumb patients via few-shot machine learning, in: Proceedings 14th International Conference on ICT, Society and Human Beings (ICT 2021), the 18th International Conference Web Based Communities and Social Media (WBC 2021), 2021.
9. S. Ferreira, E. Costa, M. Dahia, J. Rocha, A transformer-based contrastive learning approach for few-shot sign language recognition, arXiv preprint arXiv:2204.02803 (2022).
10. S. Ravi, M. Suman, P. Kishore, K. Kumar, A. Kumar, et al., Multi modal spatio temporal co-trained cnns with single modal testing on rgb–d based sign language gesture recognition, Journal of Computer Languages 52 (2019) 88–102.
11. K. M. Lim, A. W. C. Tan, C. P. Lee, S. C. Tan, Isolated sign language recogni- tion using convolutional neural network hand modelling and hand energy image, Multimedia Tools and Applications 78 (14) (2019) 19917–19944.
12. A. Wadhawan, P. Kumar, Sign language recognition systems: A decade systematic literature review, Archives of Computational Methods in Engineering 28 (3) (2021) 785–813.
13. S. Aly, W. Aly, Deeparslr: A novel signer-independent deep learning framework for isolated arabic sign language gestures recognition, IEEE Access 8 (2020) 83199–83212.
14. H. Luqman, E.-S. M. El-Alfy, Towards hybrid multimodal manual and non-manual arabic sign language recognition: Marsl database and pilot study, Electronics 10 (14) (2021) 1739.
15. P. Kumar, P. P. Roy, D. P. Dogra, Independent bayesian classifier combination based sign language recognition using facial expression, Information Sciences 428 (2018) 30–48.



16. A. Sabyrov, M. Mukushev, V. Kimmelman, Towards real-time sign language inter- preting robot: Evaluation of non-manual components on recognition accuracy., in: CVPR Workshops, 2019.
17. N. C. Camgoz, O. Koller, S. Hadfield, R. Bowden, Sign language transformers: Joint end-to-end sign language recognition and translation, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 10023– 10033.
18. [25] B. Saunders, N. C. Camgoz, R. Bowden, Progressive transformers for end-to-end sign language production, in: European Conference on Computer Vision, Springer, 2020, pp. 687–705.
19. W. Tao, M. C. Leu, Z. Yin, American sign language alphabet recognition using convolutional neural networks with multiview augmentation and inference fusion, Engineering Applications of Artificial Intelligence 76 (2018) 202–213.
20. H. Luqman, E.-S. M. El-Alfy, G. M. BinMakhashen, Joint space representation and recognition of sign language fingerspelling using gabor filter and convolutional neural network, Multimedia Tools and Applications 80 (7) (2021) 10213–10234.
21. S. Jiang, B. Sun, L. Wang, Y. Bai, K. Li, Y. Fu, Skeleton aware multi-modal sign language recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3413–3423.
22. J. Huang, W. Zhou, H. Li, W. Li, Sign language recognition using 3d convolutional neural networks, in: 2015 IEEE international conference on multimedia and expo (ICME), IEEE, 2015, pp. 1–6.
23. O. M. Sincan, H. Y. Keles, Using motion history images with 3d convolutional networks in isolated sign language recognition, IEEE Access 10 (2022) 18608–18618.
24. S. Albanie, G. Varol, L. Momeni, T. Afouras, J. S. Chung, N. Fox, A. Zisserman, Bsl-1k: Scaling up co-articulated sign language recognition using mouthing cues, in: European conference on computer vision, Springer, 2020, pp. 35–53.
25. L. Momeni, G. Varol, S. Albanie, T. Afouras, A. Zisserman, Watch, read and lookup: learning to spot signs from multiple supervisors, in: Proceedings of the Asian Con- ference on Computer Vision, 2020.
26. R. Rastgoo, K. Kiani, S. Escalera, Zs-slr: Zero-shot sign language recognition from rgb-d videos (2021). doi:10.48550/ARXIV.2108.10059. URL https://arxiv.org/abs/2108.10059
27. P. Rodríguez, I. Laradji, A. Drouin, A. Lacoste, Embedding propagation: Smoother manifold for few-shot classification, in: European Conference on Computer Vision, Springer, 2020, pp. 121–138.
28. M. Bohácek, M. Hrúz, Sign pose-based transformer for word-level sign language recognition, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 182–191.
29. J. Snell, K. Swersky, R. Zemel, Prototypical networks for few-shot learning, Ad- vances in neural information processing systems 30 (2017).
30. D. Li, C. Rodriguez, X. Yu, H. Li, 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, 2020, pp. 1459– 1469.

ترجمة النص العربي إلى مصطلحات عربية للغة الإشارة

دعاء الغامدي
قسم هندسة الحاسب، كلية علوم الحاسب والمعلومات، جامعة الملك سعود doaaalghmdi@gmail.com الرياض 11543، المملكة العربية السعودية.

منصور السليمان
قسم هندسة الحاسب، كلية علوم الحاسب والمعلومات، جامعة الملك سعود msuliman@ksu.edu.sa الرياض 11543، المملكة العربية السعودية.

يوسف العوهلي
قسم علوم الحاسب، كلية علوم الحاسب والمعلومات، جامعة الملك سعود yousef@ksu.edu.sa الرياض 11495، المملكة العربية السعودية

محمد أ. بن شريف
قسم هندسة الحاسب، كلية علوم الحاسب والمعلومات، جامعة الملك سعود mmekhtiche@ksu.edu.sa الرياض 11543، المملكة العربية السعودية.

محمد الجابري
قسم علوم الحاسب ونظم المعلومات، كلية العلوم التطبيقية، جامعة المعرفة mohmahgabri@gmail.com الرياض، 13713، المملكة العربية السعودية.

محمد أ. مختيش
قسم هندسة الحاسب، كلية علوم الحاسب والمعلومات، جامعة الملك سعود mmekhtiche@ksu.edu.sa الرياض 11543، المملكة العربية السعودية.

الملخص - تعد لغة الإشارة العربية (ARSL) لغةً يستخدمها مجتمع الصم في جميع أنحاء الدول العربية ولكن عدم إلمام عموم السكان بهذه اللغة غالبًا ما يؤدي إلى العزلة الاجتماعية للصم. وتشكل الاختلافات البنيوية بين لغة الإشارة العربية والعربية المنطوقة تحديات كبيرة للترجمة الآلية. وقد قمنا في هذه الدراسة بتعزيز ترجمة المصطلحات من العربية إلى لغة الإشارة العربية من خلال استخدام تقنيات زيادة البيانات وتوسيع مجموعة البيانات من 600 إلى أكثر من 23328 عينة باستخدام (نماذج المحول من تسلسل إلى تسلسل) (Seq2seq) . وقد حقق نهجنا تحسنًا كبيرًا في الأداء مما أدى إلى تحسين درجة معيار التقييم ثنائي اللغة (BLEU) من 11.1٪ في النموذج الأساسي إلى 52.72٪ في مجموعة الاختبار الأصلية. وقد حقق أفضل نموذج نسبة 85.17٪ كدرجة معيار التقييم ثنائي اللغة (BLEU) في اختبار البيانات الجديدة المعززة مما يؤكد فعالية زيادة البيانات في تحسين جودة ترجمة لغة الإشارة العربية

الكلمات الرئيسية - : لغة الإشارة العربية (ARSL)؛ النص المعجمي؛ زيادة البيانات؛ الترجمة الآلية؛ نموذج تسلسل إلى تسلسل (Seq2seq)؛ درجة معيار التقييم ثنائي اللغة (BLEU).

1. المقدمة

يشكل مجتمع الصم وضعاف السمع العالمي أكثر من 5٪ من سكان العالم ويعتمد هذا المجتمع بشكل كبير على لغات الإشارة للتواصل [1]. إن لغات الإشارة هي لغات غنية بصرية مكانية تستخدم مزيجًا من إيماءات اليد وتعبيرات الوجه وحركات الجسم لنقل المعنى [2]. وتعمل لغة الإشارة العربية على وجه الخصوص كوسيلة أساسية للتواصل لمجتمع الصم في الدول العربية [3]. وعلى الرغم من أهميتها تظل لغة الإشارة العربية غير مألوفة إلى حد كبير للعموم مما يساهم في العزلة الاجتماعية للأفراد الصم. وتتمتع لغة الإشارة العربية بتركيبها النحوي وقواعدها ومعجمها الخاص على عكس اللغة العربية المنطوقة مما يجعل الترجمة بين هاتين اللغتين تحديًا معقدًا

لقد أرسى تطوير نظام ترجمة آلية قائم على القواعد الدلالية لتحويل النص العربي إلى مصطلحات لغة الإشارة العربية كما هو موضح في [4] أساسًا مهمًا. ومع ذلك بقيت هذه المقاربات مقيدة بتوافر بيانات التدريب وبالقيود المتأصلة في المنهجيات القائمة على القواعد. فقد استند العمل إلى مجموعة بيانات متوازية صغيرة نسبيًا مكونة من 600 جملة عربية مترجمة إلى معجم لغة الإشارة العربية. ورغم كون هذه المجموعة من البيانات مفيدة إلا أنها غير كافية للتعبير عن التنوع الكامل للغة الطبيعية. وقد حقق النظام القائم على القواعد نسبة 35% على معيار التقييم ثنائي اللغة (BLEU) مما يسلط الضوء على التحديات في مجال الحفاظ على المعنى المقصود والبنية النحوية في الترجمة. إن هذه القيود تحد من قابلية التوسع والتكيف لنماذج الترجمة مما يؤدي إلى ترجم منخفضة الدقة

إن فعالية أنظمة الترجمة الآلية وخاصة تلك المصممة لأزواج لغوية مختلفة مثل اللغة العربية ولغة الإشارة العربية تعتمد بشكل كبير على توافر مجموعات بيانات كبيرة وعالية الجودة. وتسمح مجموعة البيانات الأكثر شمولاً بالحصول على تدريب وتعميم أفضل مما يؤدي إلى ترجم أكثر دقة [5]. كما أظهرت التطورات في معالجة اللغة الطبيعية (NLP) وتقنيات التعلم الآلي مثل نماذج "تسلسل إلى تسلسل" (Seq2seq) إمكانات كبيرة في تحسين دقة الترجمة من خلال تعلم أنماط اللغة المعقدة وعلاقاتها من البيانات المتوفرة بشكل مباشر [6]

يهدف بحثنا إلى معالجة هذه القيود عبر سد الفجوة من خلال الاستفادة من تقنيات زيادة البيانات مثل استبدال الفراغات واستبدال المرادفات وإعادة صياغة الجملة لتوسيع مجموعة البيانات الأصلية من 600 إلى أكثر من 23328 جملة. كما نقوم بتقييم البيانات الناتجة بالاعتماد عل نماذج الترجمة الآلية المتقدمة "تسلسل إلى تسلسل" للغة العربية وتطبيق تقنيات مختلفة في ما يخص حجم البيانات لفحص تأثير حجم مجموعة البيانات على أداء النموذج. إن هذا النهج يجعل البيانات أكثر نفعاً للتدريب ويشمل مجموعة أوسع من التنوع اللغوي

وتتمثل مساهمة هذا العمل في شقين: (1) نستكشف تقنيات مختلفة لزيادة البيانات لتعزيز حجم مجموعة البيانات وجودة ترجمة لغة الإشارة العربية. (2) نقوم بالتحقيق في نماذج الترجمة الآلية المختلفة "تسلسل إلى تسلسل" ومقارنتها من خلال اختبار أدائها على كل من بيانات الاختبار الأصلية وبيانات الاختبار المعززة

2. الأدبيات ذات الصلة

إن ترجمة النص العربي إلى لغة الإشارة العربية أمر ضروري لدمج الأفراد الصم في مجتمعاتهم. ومع ذلك فإن تطوير أنظمة الترجمة الفعّالة يواجه تحديات بسبب ندرة مجموعات النصوص الموازية والتوثيق غير الكامل لقواعد اللغة العربية وبنيتها. ولا تزال الأبحاث في مجال ترجمة لغة الإشارة العربية في مراحلها الأولى مقارنة بلغات الإشارة الأخرى [7] مثل لغة الإشارة الأمريكية [8] (ASL) ولغة الإشارة البريطانية [9] (BSL). إن العديد من الأنظمة الحالية تعتمد على مقاربات قائمة على القواعد وتتطلب معرفة لغوية واسعة النطاق لربط النص المنطوق أو المكتوب بتعبيرات لغة الإشارة المقابلة

وفي سياق لغة الإشارة العربية فقد تم استكشاف العديد من المقاربات. حيث ركز [10] على ترجمة الجمل العربية المتعلقة بالصلاة إلى لغة الإشارة العربية باستخدام كتابة الإشارة وهكذا فقد كانت هذه المقاربة محدودة بمجموعاتها الصغيرة وضعف التغطية لهياكل الجمل المختلفة. أم [11] فقد استخدم نهج الترجمة الآلية القائمة على الأمثلة (EBMT) ولكن اعتمادهم على تشكيل جوجل وتشابه الأمثلة أدى إلى ارتفاع معدلات الخطأ. وطور [12] نظامًا قائمًا على القواعد حقق دقة عالية على مستوى الكلمة لكنه لم يعالج بشكل كافٍ الاختلافات النحوية على مستوى الجملة. واستكشف [13] تحويلات بنية الجملة النحوية لكنه اقتصر على هياكل نحوية محددة

وقد تضمنت التطورات الحديثة تقنيات تعتمد على البيانات مثل الترجمة الآلية الإحصائية (SMT) والترجمة الآلية القائمة على الأمثلة (EBMT) والتي تبدو واعدةً ولكنها تتطلب توافر مجموعات بيانات كبيرة وعالية الجودة. وقد طور [4] نظامًا قائمًا على القواعد لترجمة النص العربي إلى لغة الإشارة العربية باستخدام مجموعة من 600 جملة في مجال الصحة. وفي حين أن نظامهم قد حقق درجة دقة تزيد عن 80٪ فإن حجم مجموعة البيانات المحدود قد حد من قابلية تطبيقه على كامل اللغة وتعميمه على نطاق أوسع

لقد بُذلت جهود كبيرة لتحسين توافر بيانات لغة الإشارة وشرحها. حيث توفر أداة ترجمة لغة الإشارة "جملة" التي وصفها [14] حلاً قائمًا على الويب إنشاء تعليقات توضيحية للغة الإشارة القطرية (QSL) بنص عربي مكتوب مما يدعم إنشاء تعليقات توضيحية مثل قاموس جملة لغة الإشارة القطرية. وقد وسع [15] هذا العمل من خلال تطوير قاموس جملة (JUMLA-QSL-22) الذي يحتوي على 6300 سجل ممثل بالنصوص الإشارية والترجمة وهوية المؤشر والموقع. وتعد هذه الأدوات ومجموعات البيانات ضرورية لتطوير معالجة لغة الإشارة (SLP) كما تسلط الضوء على الجهود الجارية لإنشاء موارد لغوية أكثر شمولاً للغات الإشارة ذات الصلة باللغة العربية

لقد تطورت تقنيات الترجمة الآلية (MT) بشكل كبير مع ظهور نماذج الترجمة الآلية العصبية (NMT) مثل بنيات "تسلسل إلى تسلسل" كطرق حديثة في هذا المجال. وقد أظهرت هذه النماذج بما في ذلك الشبكات العصبية المتكررة (RNNs) والهندسة البرمجية القائمة على المحولات قدرة قوية على التعامل مع أزواج اللغات المعقدة من خلال التعلم من كميات هائلة من البيانات لتغطية الأنماط اللغوية والسياقات المختلفة [16]. وفي ما يخص ترجمة لغة الإشارة فقد تم تطبيق هذه النماذج على لغات إشارة أخرى [17] و [18] محققة درجات متفاوتة من النجاح. ويكمن التحدي في تطبيق هذه النماذج بشكل فعال على لغة الإشارة العربية حيث تشكل ندرة البيانات والاختلافات اللغوية عقبات كبيرة. ومع ذلك فإنه يمكن لاستخدام مجموعات البيانات المعززة جنبًا إلى جنب مع نماذج ترجمة لغة الإشارة العربية أن يعزز من دقة الترجمة ويساعد في سد الفجوة بين اللغة العربية المنطوقة ولغة الإشارة العربية

لقد أظهرنا في عملنا السابق [19] أداء نموذج (AraT5-V2) للترجمة الآلية للغة العربية والتي تم تقييمها باستخدام طرق مختلفة لزيادة البيانات بما في ذلك (BR) و(SP) و(SR). وتظهر النتائج التجريبية أن طريقة (BR) أظهرت أداءً متفوقًا. ويرجع ذلك على الأرجح إلى حجم مجموعة البيانات الأكبر الذي يبلغ 22404 عينة. وعلى النقيض من ذلك فقد أظهرت طريقتا (SP) و(SR) اللتان استخدمتا مجموعات بيانات أصغر حجمًا خسائر أعلى في التحقق ودرجات معيار التقييم ثنائي اللغة (BLEU) أقل بكثير. وقد تم إجراء مزيد من التحليل من خلال الجمع بين طرق الزيادة الثلاثة والمقارنة مع نماذج أخرى بما في ذلك نموذجي (AraT5 Base) و(mT5) الأصليين حيث تفوق نموذج (AraT5 V2) بحصوله على درجة 90.93 على معيار التقييم ثنائي اللغة (BLEU)

وقد قمنا في هذا العمل بتوسيع التحقيق من خلال استخدام تقنيات زيادة البيانات لتوسيع مجموعة البيانات الأصلية إلى أكثر من 23000 عينة واختبار النماذج على مجموعة الاختبار الأصلية حيث كنا قد اختبرنا النماذج بواسطة مجموعة الاختبار المعززة في دراستنا السابقة. كما أننا في هذه الدراسة نطبق تقنيات نسبة البيانات لدراسة تأثير حجم مجموعة البيانات. ونقوم في جميع التجارب بمقارنة أداء نماذج (AraT5) الأساسية و (Arat5 v2) و (mT5).

3. المنهجية

سنحدد في هذا القسم المنهجية المستخدمة لتحسين الترجمة من النص العربي إلى النص الإشاري. ويركز نهجنا على توسيع حجم مجموعة البيانات بشكل كبير من خلال تقنيات مختلفة لزيادة البيانات وتنفيذ نماذج الترجمة الآلية المتقدمة. نحن نهدف إلى معالجة قيود الأنظمة السابقة القائمة على القواعد وتحسين دقة الترجمة وموثوقيتها من خلال إثراء مجموعة البيانات والاستفادة من نماذج "تسلسل إلى تسلسل"

يوضح الشكل 1 مثالاً لعملية الترجمة من اللغة العربية المنطوقة إلى لغة الإشارة بما يشمل التمثيل الإشاري الوسيط والذي يعمل كخطوة حاسمة في سد الفجوة بين بناء الجملة في اللغة العربية المنطوقة وقواعد لغة الإشارة العربية. وستكون الترجمة الدقيقة بلغة الإشارة أسهل بالنسبة لنماذج الترجمة الآلية من خلال ترجمة اللغة المنطوقة إلى النص الإشاري

الشكل 1. توضيح لعملية الترجمة من اللغة العربية المنطوقة إلى النص الإشاري ثم إلى لغة الإشارة.

3.1 تقنيات زيادة البيانات

لقد قمنا باستخدام أساليب زيادة البيانات لتوسيع مجموعة البيانات لتصل إلى 23328 جملة بهدف تجاوز القيود التي فرضتها محدودية مجموعة بيانات لغة الإشارة العربية الأصلية والتي كانت تتألف من 600 جملة فقط وهي تركز في الغالب على مجال الرعاية الصحية. وتشمل تقنيات الزيادة الأساسية المستخدمة (BR) و(SR) و(SP). وقد تم اختيار هذه التقنيات بناءً على قدرتها على تعزيز تنوع وقوة بيانات التدريب وهو أمر ضروري للتعامل مع السمات اللغوية المعقدة للغة العربية ولغة الإشارة العربية

يعد تطوير واستخدام خوارزمية الفهرسة أحد المكونات الأساسية لمنهجيتنا فهي تضمن المعالجة المنهجية والمطابقة الدقيقة بين النص العربي وما يقابله من النص الإشاري. وتقوم خوارزمية الفهرسة بتعيين مؤشرات فهرسة لكل كلمة في كل من الجمل العربية الأصلية وما يقابلها من النص الإشّاري مع الحفاظ على المحاذاة والاتساق طوال عملية زيادة البيانات. ويعد هذا أمراً ضروريًا للحفاظ على المعنى الدلالي للجمل عند تطبيق تقنيات زيادة البيانات لأنه يضمن الانعكاس الدقيق لأي تعديلات يتم إجراؤها على النص الأصلي على نسخة النص الإشاري. ويسهل استخدام هذه الخوارزمية التكامل السلس لعينات البيانات الجديدة مما يسمح بزيادة البيانات بشكل فعال وقابل للتطوير ووضع الأساس لإنشاء مجموعة بيانات متنوعة وعالية الجودة

استبدال الفراغات (BR) هو منهجية لزيادة البيانات تستخدم في معالجة اللغة الطبيعية لمحاكاة الكلمات المفقودة أو غير المعروفة وتعزيز أداء نماذج التعلم الآلي. وتتضمن هذه التقنية إخفاء كلمات محددة في الجملة والتنبؤ بهذه الكلمات بناءً على السياق المحيط باستخدام أداة ملء الفراغ ونموذج (20)[AraELECTRA]. وتتيح تقنية استبدال الفراغات إثراء مجموعة البيانات بما يصل إلى 21804 عينة جديدة من خلال إنشاء نسخ معدلة من الجمل الأصلية وتوليد كلمات مرشحة جديدة بدل الكلمات المخفية

يقدم استبدال المرادفات (SR) تنوعًا معجميًا عن طريق استبدال الكلمات بمرادفاتها من قاموس محدد مسبقًا ما يؤدي بالتالي إلى زيادة تنوع المفردات مع الحفاظ على المعنى العام للجمل. وتستخدم هذه الطريقة قاموس مخصص بناءً على قاموس لغة الإشارة السعودية ومجموعة بيانات لغة الإشارة العربية مما يضمن كون المرادفات ذات صلة بلغة الإشارة العربية من حيث السياق. ومن خلال تعريض النموذج لاختيارات معجمية مختلفة تنقل معاني مماثلة تعزز هذه الطريقة قدرة النموذج على تغطية التعبيرات اللغوية المختلفة وتحسن قدرته على التكيف مع استخدامات الكلمات المختلفة مما أدى إلى إنشاء 684 جملة جديدة

تُستخدم طريقة **إعادة صياغة الجملة (SP)** لتزويد بيانات التدريب ببعض التنوع ومساعدة النموذج على تعلم طرق مختلفة لتقديم نفس المعلومات من خلال إنشاء نسخ معاد صياغتها من الجمل من خلال الترجمة العكسية (أي ترجمة الجمل إلى الإنجليزية ثم إلى العربية مرة أخرى). وتولد هذه العملية عبارات بديلة تحافظ على المعنى الأصلي ولكنها تختلف من حيث البنية. ويعد هذا التنوع أمرًا بالغ الأهمية للتعامل مع الاختلافات البنيوية الكبيرة بين اللغة العربية ولغة الإشارة العربية. ويصبح النموذج من خلال التدريب على البيانات المعاد صياغتها أكثر مرونة في التعرف على مجموعة واسعة من بنى الجمل وترجمتها بدقة وبالتالي تتحسن قدرته على التقاط المعنى والحفاظ على الاتساق النحوي في الترجمة. وقد نتج عن استخدام طريقة إعادة صياغة الجملة 840 جملة جديدة.

يوضح الشكل 2 أمثلة لمجموعات من البيانات المعززة التي تغطي مجموعة واسعة من المفردات وبنى الجمل مما يوفر للنموذج الخبرة اللازمة للتعامل مع الفروق الدقيقة اللغوية.

الشكل 2. عينات من مجموعات البيانات من اللغة العربية إلى النص الإشاري العربي.

يعزز استخدام تقنيات زيادة البيانات كما هو مفصل في عملنا السابق بشكل كبير من حجم مجموعة البيانات وتنوعها [19]. وتوفر مجموعة البيانات الغنية هذه أساسًا متينًا لتدريب نماذج الترجمة الآلية الفعالة ودعم تطوير نظام قوي قادر على ترجمة النص العربي بدقة إلى لغة الإشارة العربية وتحسين إمكانية النفاذ والتواصل لمجتمع الصم

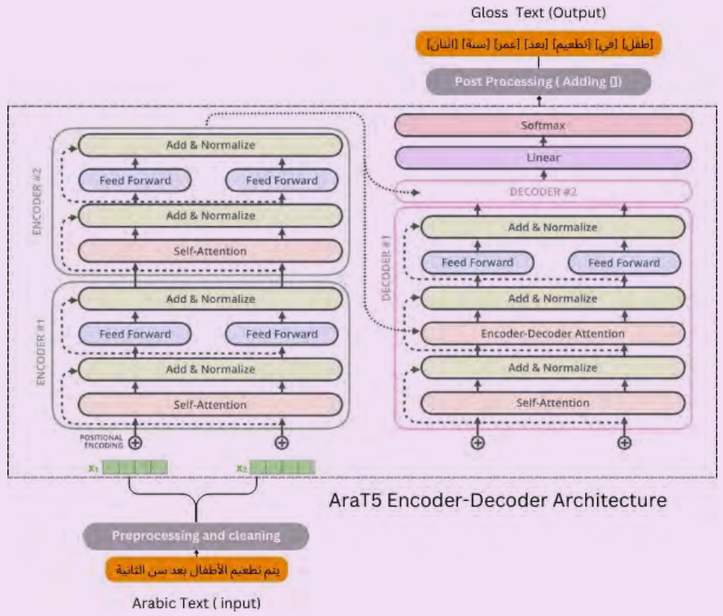
3.2 نموذج الترجمة الآلية "تسلسل إلى تسلسل"

أصبحت نماذج المحولات "من تسلسل إلى تسلسل" تشكل حجر الزاوية في معالجة اللغة الطبيعية وخاصة للمهام التي تتطلب تحويل تسلسلات الإدخال إلى تسلسلات إخراج مثل الترجمة الآلية. وقد تم تصميم هذه النماذج للتعامل مع تسلسلات الإدخال والإخراج ذات الأطوال المتغيرة مما يجعلها مناسبة لترجمة النص من لغة إلى أخرى مع الحفاظ على معنى العينات. إن نموذج T5 (محول لتحويل النص إلى نص) هو أحد هذه النماذج المتطورة والذي يوحد مهام معالجة اللغة الطبيعية المختلفة تحت إطار واحد عن طريق تحويلها إلى مهام "نص إلى نص" [6]. وتعد هذه البنية مناسبة بشكل خاص لمهام الترجمة نظرًا لقدرتها على التعامل مع الأنماط اللغوية المتنوعة والسياق بشكل فعال

نستخدم أيضًا في دراستنا (AraT5-V2) وهو أحد أشكال تدريب نموذج T5 مخصص للغة العربية [21]. ويستفيد (AraT5-V2) من البنية القوية لـ T5 والتي تم تطويرها للتعامل مع تعقيدات بناء الجملة والدلالات العربية. ويتكون النموذج من مُشفِّر ذي طبقات متعددة لكل منها آلية انتباه ذاتي و"شبكة تغذية للأمام" يتبعها مُفكِّك تشفير يتضمن أيضًا انتباهًا إلى مخرجات المُشفِّر. ويسمح هذا الهيكل للنموذج بإنشاء ترجمات تعكس بدقة معنى اللغة المصدر وتتوافق مع المعايير النحوية للغة الهدف. ويوضح الشكل 3 بنية الترجمة الآلية للمصطلحات العربية باستخدام نموذج (AraT5)

لمنع حدوث مشكلة عدم الكفاءة عند التعامل مع البيانات الجديدة (overfitting). ويُعتبر النموذج جاهزًا للاختبار بمجرد أن يؤدي أداءً مرضيًا وفق مجموعة التحقق

إن تدريب نموذج (AraT5-V2) يجعله قادراً على أخذ جملة عربية كمدخل وتوليد ترجمة مناسبة لها إلى نص الإشارة. وتضمن بنية التشفير وفك التشفير محافظة المخرجات على المعنى الدلالي واتباعها للقواعد النحوية للغة الإشارة مما يعكس المعرفة المكتسبة أثناء التدريب. ويتم تقييم أداء النموذج باستخدام مقاييس مثل درجة (BLEU) والتي تقيس دقة الجمل المترجمة مقابل الإشارات البشرية



AraT5 Encoder-Decoder Architecture

الشكل 3. بنية نموذج (AraT5) في الترجمة الآلية إلى النص الإشاري.

كما هو موضح في الشكل 3 فإن عملية تدريب (AraT5-V2) لترجمة النص العربي إلى النص الإشاري تتضمن سلسلة من الخطوات المنظمة بعناية لتحسين أداء النموذج. بدايةً يقوم النموذج المدرب مسبقًا والذي يفهم بالفعل بنى اللغة العربية العامة بتدريب إضافي على مجموعة بيانات موازية من الجمل العربية وترجمتها الإشارية المقابلة

ويقوم المشفر أثناء التدريب بمعالجة المدخلات وتحويلها إلى سلسلة من التمثيلات السياقية التي تغطي معنى الجملة. ثم يتم تمرير هذه التمثيلات إلى المشفر الذي يولد ناتج الشرح بطريقة تلقائية - يتنبأ برمز واحد في كل مرة مع الاستفادة من الرموز التي تم إنشاؤها مسبقًا في عملية التنبؤ بالرمز التالي. ويتم تعديل معلمات النموذج باستخدام الانتشار العكسي (backpropagation) حيث يتم تقليل الاختلافات بين جمل الإشارة المتوقعة وجمل الإشارة الفعلية باستخدام خوارزميات التحسين مثل (AdamW). كما يتم تدريب المعلمات الفائقة (عوامل التدريب) بما في ذلك معدل التعلم وحجم الدفعة لتحقيق الأداء الأمثل ويتم استخدام تقنيات مثل التوقف المبكر

4. النتائج التجريبية

لقد استخدمنا نموذج (AraT5 V2) في تجاربنا وهو نموذج ترجمة آلية عصبية متطور مصمم خصيصًا للنص العربي. وقمنا بتقييم نموذجين آخرين وهما [22] (mT5) وهو نموذج محول متعدد اللغات قادر على التعامل مع لغات مختلفة و [21] (AraT5 Base) وهو الإصدار الأساسي من (AraT5) والذي تم تصميمه خصيصًا للغة العربية ولكنه يفتقد للتحسينات الموجودة في الإصدار الثاني V2. كما أن مقياس التقييم الأساسي المستخدم في هذه التجارب هو درجة معيار التقييم ثنائي اللغة (BLEU) والتي تُستخدم عادةً في مهام الترجمة الآلية لتقييم جودة الترجمة. وبالإضافة إلى درجة (BLEU) نقوم أيضًا بقياس خسائر التحقق والاختبار ومقارنة تنبؤات النموذج بمجموعات الاختبار المرجعية. لقد تم إجراء التدريب باستخدام معدل التعلم التكيفي مع مُحسِّن (AdamW) جنبًا إلى جنب مع معدل فاقد 0.1 لمنع حدوث مشكلة عدم الكفاءة عند التعامل مع البيانات الجديدة. وقد استخدمنا حجم دفعة يتراوح من 8 إلى 128 مع تعديله بناءً على حجم مجموعة البيانات ومجدول معدل التعلم الخطي. وقد تم إجراء 22 عملية مرور محلية على كامل بيانات التدريب (epochs) مع إجراء التقييم وحفظ النموذج

كل 500 خطوة لمراقبة التقدم ومنع حدوث مشكلة عدم الكفاءة عند التعامل مع البيانات الجديدة

4.1 أداء تقنيات زيادة البيانات

لقد قمنا بحساب درجات (BLEU) لكل طريقة على حدة وذلك بهدف تقييم أداء نموذج (AraT5-V2) المدرب باستخدام طرق مختلفة لزيادة البيانات. حيث تم تقييم كل طريقة لزيادة البيانات باستخدام التحقق المقسم (validation split) أثناء التدريب بينما تم حساب درجات (BLEU) باستخدام مجموعة اختبار لغة الإشارة العربية الأصلية المكونة من 90 عينة لتحديد التأثير الكلي على عملية الترجمة. ويستعرض الجدول 1 نتائج نموذج (AraT5-V2) الخاصة بكل طريقة لزيادة البيانات جنبًا إلى جنب مع النتائج من مجموعة البيانات الأصلية قبل الزيادة

الجدول 1. مقاييس أداء (AraT5-V2) لمختلف طرق زيادة البيانات.

Metric	BR	SP	SR	Original ArSL
Val. Loss	0.260	2.173	1.795	3.116
Val. BLEU	92.778	25.75	29.30	15.273
Test Loss	2.413	2.396	2.383	2.159
Test BLEU	52.71	13.33	12.90	11.069
Dataset Size	22404	1440	1284	600

لقد استمرت طريقة (BR) كما هو موضح في الجدول 1 بتقديم أفضل أداء مع خسارة في التحقق من الصحة قدرها 0.260 ودرجة (BLEU) للتحقق من الصحة قدرها 92.778. وكانت درجة (BLEU) للاختبار ل (BR 52.71) وهي أعلى بكثير من درجات الطرق الأخرى. من المحتمل أن يكون حجم مجموعة البيانات الكبير البالغ 22404 عينة لطريقة (BR) قد ساهم في أدائها المتفوق مما يسمح للنموذج بتعلم أنماط ترجمة أكثر قوة والتعامل بشكل فعال مع البيانات الجديدة عليه

وفي المقابل أظهرت طريقتي (SP) و(SR) أداءً أضعف مع خسائر في التحقق من الصحة بلغت 2.173 و1.795 على التوالي ودرجات (BLEU) للتقييم 25.75 ل(SP) و29.30 ل(SR).أما طريقة (SR) التي استخدمت أصغر حجم لمجموعة البيانات والمكونة من 1284 عينة فقد كانت صاحبة أدنى درجة اختبار (BLEU) وهي 12.90. مما يشير إلى أن التغطية المحدودة للبيانات والمفردات قد قللت بشكل كبير من فعاليتها. ومع ذلك نرى أن طريقتي (SP) و(SR) تعملان بشكل أفضل من مجموعة بيانات لغة الإشارة العربية الأصلية غير المعززة والتي حصلت على درجة اختبار (BLEU) قدرها 11.069 مما يوضح قيمة تعزيز البيانات في تحسين جودة الترجمة

لقد قمنا بدمج جميع طرق زيادة البيانات الثلاثة (BR و SP وSR) مما أدى إلى تكوين مجموعة بيانات من 23328 عينة تُستخدم لتدريب نموذج (AraT5 V2) ومقارنتها بنماذج (AraT5 Base) و(mT5) باستخدام مجموعة البيانات المعززة المختلطة. ومن المهم هنا ملاحظة أننا استخدمنا مجموعة الاختبار الأصلية لتقييم النماذج يوضح الجدول 2 نتائج مقارنة هذه النماذج. وكما هو موضح في الجدول 2 فقد حقق (AraT5 V2) أعلى درجات (BLEU) وأقل خسائر في مجال التحقق والاختبار مع درجة (BLEU) للتحقق من الصحة 86.16 ودرجة (BLEU) للاختبار 69.41. كما حقق النموذج الأساسي أداءً جيدًا من حيث درجة (BLEU) للتحقق من الصحة حيث وصل إلى 35.190 ولكنه حصل على درجة (BLEU) للاختبار أقل بكثير وهي 33.62. كما أظهر نموذج (mT5) أداءً معتدلاً مع درجة (BLEU) للتحقق من الصحة 72.380 ودرجة (BLEU) للاختبار أقل بكثير 15.157.

Model	Val. Loss	Val. BLEU	Test Loss	Test BLEU
AraT5 V2	**0.492**	**86.16**	0.174	**69.41**
Base	1.610	35.190	0.979	33.62
mT5	0.586	72.380	**0.265**	15.157

الجدول 2. مقارنة بين نماذج الترجمة الآلية المختلفة.

بهدف التحقق من تأثير زيادة البيانات على أداء نموذج الترجمة الآلية الخاص بنا قمنا بإجراء تجربة على نسب زيادة البيانات. وتبحث هذه الدراسة في كيفية تأثير تغيير نسب البيانات المعززة على وجه التحديد 20٪ و 40٪ و 80٪ و 100٪ على دقة ترجمة نموذج (AraT5 V2). وتهدف هذه التجربة إلى تحديد التوازن الأمثل بين تنوع البيانات وحجمها لتحسين أداء النموذج وذلك من خلال ضبط كمية البيانات المعززة بشكل منهجي مع الحفاظ على إعدادات التدريب. وسيتم تقييم فعالية هذه النسب باستخدام درجات (BLEU) مما يوفر معلومات حول كيفية مساهمة مستويات مختلفة من زيادة البيانات في قوة ودقة نماذج الترجمة. ويوضح الجدول 3 حجم نسب البيانات لمختلف طرق زيادة البيانات

Method	Split	20%	40%	80%	100%
BR	Train. Size	3,908	7,398	14,374	17,863
	Val. Size	526	962	1,834	2,270
	Test Size	526	962	1,834	2,270
	Total Size	4,360	8,722	17,443	21,804
SP	Train. Size	1,985	1,942	1,907	2,173
	Val. Size	18.45	22.45	23.55	25.75
	Test Size	106	123	157	174
	Total Size	168	336	672	840
SR	Train. Size	531	640	859	968
	Val. Size	103	117	144	158
	Test Size	103	117	144	158
	Total Size	137	274	547	684
All	Train. Size	4,995	8,728	16,191	19,923
	Val. Size	735	1,202	2,135	2,602
	Test Size	735	1,202	2,135	2,602
	Total Size	4,665	9,332	18,662	23,328

الجدول 3. نسبة البيانات وأحجام مجموعات البيانات في مختلف طرق زيادة البيانات.

يوضح الجدول 3 أحجام مجموعات البيانات لمختلف طرق زيادة البيانات المختلفة بما في ذلك (BR وSP و SR) بشكل منفصل وبشكل جماعي بنسب 20% و40% و80% و100%. ويمثل الحجم الإجمالي عدد العينات المقابلة لكل نسبة من مجموعة البيانات الكاملة. وقد تم تحديد أحجام التحقق والاختبار من خلال تقسيم مجموعة البيانات الإجمالية وتضمين 90 عينة إضافية من مجموعة بيانات لغة الإشارة العربية الأصلية في مجموعة الاختبار

الجدول 4. درجات اختبار (BLEU) لمختلف طرق ونسب زيادة البيانات.

Method	20%	40%	80%	100%
BR	42.49	72.54	90.97	**91.57**
SP	12.63	16.52	20.27	21.77
SR	12.01	15.89	15.71	18.49
All	34.73	65.29	82.46	85.17

لقد تفوقت طريقة (BR) على الطرق الأخرى في جميع النسب كما هو موضح في الجدول 4 حيث حققت أعلى درجة اختبار (BLEU) وهي 91.57 بنسبة 100%. ويشير هذا التفوق إلى أن طريقة (BR) توفر بيانات التدريب الأكثر قوة للنموذج ويرجع ذلك على الأرجح إلى قدرتها على التعامل مع أنماط لغوية متنوعة بشكل فعال. وحتى في النسب المنخفضة فقد أظهرت طريقة (BR) تحسينات كبيرة مع درجة (BLEU) تبلغ 42.49 بنسبة زيادة 20% مما يسلط الضوء على تأثيرها القوي حتى مع وجود بيانات أقل

أما طريقتا (SP وSR) فقد حققتا درجات (BLEU) اختبارية أقل نسبيًا في جميع النسب حيث حققت طريقة (SP) أعلى درجة (BLEU) بنسبة 21.77% عند زيادة 100% وبلغت طريقة (SR) ذروتها عند 18.4%. وتشير هذه النتائج إلى أنه في حين تساهم (SP وSR) في تحسين أداء النموذج إلا أن تأثيرهما أقل وضوحًا مقارنة بـ(BR). وقد يعود انخفاض فعالية (SP) إلى القيود المتأصلة

إن هذا النهج يضمن إمكانية تقييم تأثير كل طريقة في ظل ظروف موحدة مما يوفر معلومات حول فعالية أحجام البيانات المختلفة واستخدامها معاً على دقة الترجمة. ويوضح الجدول 4 نتائج اختبار (BLEU) لكل طريقة زيادة بيانات والنسب المستخدمة

في الترجمة العكسية والتي تنتج أحيانًا عبارات معاد صياغتها تشبه إلى حد كبير الأصل أو تسبب ضوضاء لا تعزز عملية التدريب. ويمكن أن يُعزى الأداء الضعيف نسبيًا لـ(SR) إلى التغطية المحدودة للمفردات وأهمية السياق لقاموس المرادفات المستخدم والذي ربما أدى إلى استبدالات لم تغير بيانات التدريب بشكل كبير أو في بعض الحالات شوهت معنى الجملة

لقد أظهر الاستخدام الجماعي لطرق زيادة البيانات (جميع الطرق) أداءً متوازنًا حيث حقق درجة اختبار (BLEU) بنسبة 85.17% عند زيادة بنسبة 100% مما يشير إلى أن مزيجًا من تقنيات الزيادة يمكن أن يقدم أداء جيداً ولكنه قد لا يتجاوز فعالية طريقة (BR) لوحدها. وبشكل عام تؤكد هذه النتائج على أهمية اختيار طرق زيادة البيانات المناسبة وتحسين تنفيذها لتعزيز دقة الترجمة الآلية فضلاً عن الحاجة إلى مقاربات أكثر تطورًا لتحسين طريقتي (SP وSR)

5. الخاتمة

لقد قمنا في هذه الدراسة بتعزيز ترجمة النص العربي إلى نص إشاري عربي باستخدام تقنيات زيادة البيانات المتقدمة ونموذج (AraT5 V2). وقد أظهرت نتائجنا أن طريقة استبدال الفراغات (BR) كانت صاحبة أعلى دقة في الترجمة في حين أن طريقة التكبير المختلطة حسنت الأداء أيضًا ولكنها لم تتفوق على (BR). ومع ذلك فقد استخدمنا في هذه الدراسة مجموعة بيانات صغيرة تم تطويرها في مجال الصحة والتي لا تغطي مجموعة شائعة من الكلمات العربية. وسيركز العمل المستقبلي على إكمال المرحلة الثانية لترجمة النص الإشاري إلى نص ثم إلى حركات لغة الإشارة العربية مما يؤدي إلى تطوير أنظمة ترجمة لغة الإشارة لتصبح أكثر كفاءة ودقة في خدمة مجتمع الصم الناطقين باللغة العربية بشكل أفضل.

المراجع

1. Kushalnagar, R. (2019). Deafness and hearing loss. Web accessibility: A foundation for research, pages 35–47.
2. Luqman, H., Mahmoud, S. A., et al. (2017). Transform-based arabic sign language recognition. Procedia Computer Science, 117:2–9.
3. Al-Fityani, K. and Padden, C. (2010). Sign language geography in the arab world. Sign languages: A Cambridge survey, 20.
4. Luqman, H. and Mahmoud, S. A. (2019). Automatic translation of arabic text-to-arabic sign language. Universal Access in the Information Society, 18(4):939–951.
5. Koehn, P. and Knowles, R. (2017). Six challenges for neural machine translation. arXiv preprint arXiv:1706.03872.
6. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of machine learning research, 21(140):1–67.
7. Sidig, A. a. I., Luqman, H., and Mahmoud, S. A. (2018). Arabic sign language recog- nition using optical flow-based features and hmm. In Recent Trends in Information and Communication Technology: Proceedings of the 2nd International Conference of Reliable Information and Communication Technology (IRICT 2017), pages 297–305. Springer.
8. [8] Zhao, L., Kipper, K., Schuler, W., Vogler, C., Badler, N., and Palmer, M. (2000). A machine translation system from english to american sign language. In Envisioning Machine Translation in the Information Future: 4th Conference of the Association for Machine Translation in the Americas, AMTA 2000 Cuernavaca, Mexico, October 10–14, 2000 Proceedings 4, pages 54–67. Springer.

9. Marshall, I. and Sáfár, É. (2003). A prototype text to british sign language (bsl) trans- lation system. In The companion volume to the proceedings of 41st annual meeting of the association for computational linguistics, pages 113–116.
10. Almasoud, A. M. and Al-Khalifa, H. S. (2012). Semsignwriting: A proposed semantic system for arabic text-to-signwriting translation.
11. Almohimeed, A., Wald, M., and Damper, R. I. (2011). Arabic text to arabic sign language translation system for the deaf and hearing-impaired community. In Proceedings of the second workshop on speech and language processing for assistive technologies, pages 101–109.
12. El, A., El, M., and El Atawy, S. (2014). Intelligent arabic text to arabic sign language translation for easy deaf communication. International Journal of Computer Applica- tions, 92(8).
13. Al-Rikabi, S. and Hafner, V. (2011). A humanoid robot as a translator from text to sign language. In 5th Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC 2011), pages 375–379.
14. 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.
15. Othman, A., El Ghoul, O., Aziz, M., Chemnad, K., Sedrati, S., and Dhouib, A. (2023). Jumla-qsl-22: Creation and annotation of a qatari sign language corpus for sign lan- guage processing. In Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments, pages 686–692.
16. Angelova, G., Avramidis, E., and Möller, S. (2022). Using neural machine translation methods for sign language translation. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pages 273–284.
17. Jang, J. Y., Park, H.-M., Shin, S., Shin, S., Yoon, B., and Gweon, G. (2022). Automatic gloss-level data augmentation for sign language translation. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 6808–6813.
18. Kayahan, D. and Güngör, T. (2019). A hybrid translation system from turkish spo- ken language to turkish sign language. In 2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA), pages 1–6. IEEE.
19. Alghamdi, D., Alsulaiman, M., Alohali, Y., Bencherif, M. A., and Algabri, M. (2024). Arabic gloss machine translation through data augmentation. In Proceedings of the Third SmartTech Conference (Manuscript submitted for publication). King Saud Uni- versity.
20. Antoun, W., Baly, F., and Hajj, H. (2020). Araelectra: Pre-training text discriminators for arabic language understanding. arXiv preprint arXiv:2012.15516.
21. Nagoudi, E. M. B., Elmadany, A., and Abdul-Mageed, M. (2021). Arat5: Text-to-text transformers for arabic language generation. arXiv preprint arXiv:2109.12068.
22. Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., and Raffel, C. (2020). mt5: A massively multilingual pre-trained text-to-text transformer. arXiv preprint arXiv:2010.11934.

تطوير أنظمة الذكاء الاصطناعي لترجمة لغة الإشارة في الفيديو

الأستاذ المشارك أوزر سيليك
ozer@ogu.edu.tr
قسم الرياضيات وعلوم الكمبيوتر، كلية العلوم، جامعة إسكي شهير عثمان غازي، إسكي شهير، تركيا

أحمد أفجي أوغلو
ahmet.avcioglu@engelsizceviri.com
جامعة إسكي شهير عثمان غازي مبنى ETGB تكنوبارك رقم: 44/106 إسكي شهير/تركيا

الملخص - غالبًا ما يواجه الأشخاص الصم أو ضعاف السمع تحديات في فهم الترجمة المكتوبة بالكامل في مقاطع الفيديو بسبب الاختلافات بين لغة الإشارة واللغة المنطوقة، حيث يتمتع كل منهما بقواعد نحوية وبنية فريدة. وفي حين توفر العديد من منصات الوسائط الاجتماعية ترجمة آلية إلا أنها غالبًا ما تكون غير كافية لتحقيق إمكانية النفاذ. لا بد من تضمين ترجمة لغة الإشارة لجعل محتوى الفيديو متاحًا بالكامل. وتتضمن إضافة لغة الإشارة إلى مقاطع الفيديو تقليديًا عملية تستغرق وقتًا طويلاً لتسجيل وإدخال مقاطع فيديو ترجمة منفصلة يجب إعادة تسجيلها عند إجراء أي تغييرات على الفيديو الأصلي. ويبسط البرنامج المساعد الإضافي (plugin) هذه العملية من خلال ترجمة التحديثات ديناميكيًا مباشرة من ملفات الترجمة مما يقلل بشكل كبير من الوقت والجهد والتكلفة. فهو يسمح بدعم لغة الإشارة دون الحاجة لتعديل المحتوى الأصلي. ويتكامل البرنامج المساعد الإضافي مع مشغلات الفيديو مما يوفر نافذة لغة إشارة قابلة للتخصيص يتحكم فيها المستخدمو يمكن تنشيطها أو نقلها أو تغيير حجمها أو إيقاف تشغيلها حسب الرغبة.

الكلمات الرئيسية
ترجمة لغة الإشارة، الذكاء الاصطناعي، البرمجة اللغوية العصبية، النفاذ الرقمي.

المقدمة

غالبًا ما تخطي الحواجز السمعيةالتي يواجهها الأفراد الصم أو ضعاف السمع التحديات السمعية، وهي تستبعدهم من أنشطة أساسية في الحياة اليومية وخاصة عندما يتعلق الأمر بفهم اللغة المكتوبة. وتختلف اللغات المنطوقة ولغة الإشارة اختلافاً كلياً ليس فقط من حيث الشكل ولكن أيضًا في هياكلها النحوية وقواعدها ودلالاتها. ويعني هذا الاختلاف الأساسي أن الاعتماد على الترجمة المكتوبة وحدها غير كافٍ لضمان النفاذ الكامل لمجتمعات الصم وضعاف السمع. وحتى مع الانتشار الواسع للترجمة الآلية عبر منصات الوسائط الاجتماعية وخدمات البث فإن العديد من الأفراد يواجهون صعوبة في تفسير هذه التسميات التوضيحية لأنها لا تعكس تمامًا قواعد اللغة أو الفروق الدقيقة في لغة الإشارة. وبالتالي فإن جهود إمكانية النفاذ الرقمي التي تعتمد فقط على الترجمة المكتوبة تتجاهل جانبًا أساسيًا من جوانب التواصل - المعلومات المنقولة من خلال لغة الإشارة.

وقد أبرزت العديد من الدراسات أهمية لغة الإشارة في تعزيز إمكانية النفاذ للصم وضعاف السمع مؤكدة أن الترجمة على الرغم من كونها مفيدة ليست بديلاً عن الترجمة بلغة الإشارة [1]. وعلى سبيل المثال يسلط تقرير صادر عن الاتحاد العالمي للصم الضوء على أن أكثر من 70 مليون شخص حول العالم يستخدمون لغة الإشارة كوسيلة أساسية للتواصل مما يؤكد على أهمية توفير وسائل الإعلام القابلة للنفاذ والتي تتضمن ترجمة لغة الإشارة [3]. ونتيجة لذلك فإنه من الأهمية بمكان تجاوز تدابير إمكانية النفاذ غير الكافية ودمج الترجة بلغة الإشارة في محتوى الفيديو لتلبية احتياجات الأشخاص الصم وضعاف السمع.

يستكشف هذا البحث الأساليب التقليدية لدمج لغة الإشارة في محتوى الفيديو والقيود المفروضة على هذه الأساليب وكيف يمكن للتكنولوجيا الحديثة وخاصة الذكاء الاصطناعي إحداث ثورة في هذه العملية. ونقترح مجموعة من البرامج المساعدة الإضافية المبتكرة للويب والفيديو وملفات PDF التي تسمح بإنشاء ترجمة بلغة الإشارة في الوقت الفعلي متزامنة مع الترجمة. ويوفر هذا النظام حلاً مستدامًا وبأسعار معقولة ويمكنه التكيف مع التحديثات على الفيديو الأصلي مما يضمن للمستخدمين الصم وضعاف السمع النفاذ المستمر إلى المحتوى القابل للنفاذ بالكامل.

المنهجية

تتطلب عملية تطوير منصة محتوى فيديو قابل للنفاذ بالكامل للصم وضعاف السمع التغلب على العديد من التحديات سواء الفنية أو العملية. ويتضمن توفير الترجمة بلغة الإشارة في مقاطع الفيديو تقليديًا عملية متعددة الخطوات تبدأ بترجمة النص إلى لغة الإشارة بواسطة مترجم مدرب. وتتطلب هذه العملية إعداد استوديو احترافي حيث يتم تسجيل فيديو المترجم وتحريره ثم إدخاله على محتوى الفيديو الأصلي. وفي كل مرة يتم فيها تحديث الفيديو أو الترجمة الأصلية يجب أيضًا إعادة تسجيل فيديو لغة الإشارة وهو أمر يستغرق وقتًا طويلاً ويعتبر مكلفًا. إن هذه الطريقة تحد من إمكانية وقابلية التوسع في مجال إضافة دعم لغة الإشارة إلى مجموعة واسعة من محتوى الفيديو

وغير هذا النهج المقترح هذه العملية من خلال الاستفادة من الذكاء الاصطناعي وتقنيات مزامنة الترجمة المتقدمة. حيث يستخرج النظام الذي طورناه المعلومات من ملف الترجمة في الفيديو ويستخدمها لتوليد ترجمة بلغة الإشارة في الوقت الفعلي لك إطار فيديو. ويلغي هذا الأمر الحاجة إلى إعادة التسجيل المكلفة أو التحرير بعد الإنتاج كلما تم إجراء تغييرات على محتوى الفيديو الأصلي أو الترجمة (الشكل 1)

الشكل 1. لقطة شاشة من فيديو مؤسسة إسطنبول للفنون والثقافة حيث يتم استخدام البرنامج الإضافي (SignForDeaf) لترجمة لغة الإشارة بالفيديو من موقعهم على الويب. (https://www.iksv.org/tr/haberler/iksv-alt-kat-yepyeni-bir-cevrimici-seriye-basliyor-kim-bu-cizgi-filmciler).

ويكمن جوهر هذا الحل في قدرته على ترجمة التحديثات من الترجمة النصية إلى لغة الإشارة بشكل مستدام مما يوفر ترجمة بلغة الإشارة تكون مستمرة ودقيقة ومتزامنة مع الفيديو. كما أن دمج هذا النظام مع منصات الفيديو الحالية مثل اليوتيوب يعد أمراً بسيطاً ولا يتطلب أي تعديل على ملف الفيديو الأصلي. حيث يعمل البرنامج الإضافي بدلاً من ذلك كطبقة فوقية مما يسمح للمستخدمين بتنشيط أو نقل أو تغيير حجم أو تعطيل نافذة لغة الإشارة بناءً على تفضيلاتهم. وتضمن هذه الميزة سهلة الاستخدام إمكانية تخصيص النظام وفقًا للاحتياجات المحددة لكل مشاهد مما يعزز إمكانية النفاذ دون المساس بالسلامة البصرية للفيديو

كما تم تصميم البرنامج الإضافي مع أخذ القدرة على التكيف في المستقبل في الحسبان. فمع إضافة لغات إشارة جديدة يمكن تحديث النظام بسهولة لاستيعاب أقاليم ولغات مختلفة مما يضمن تطبيقه في سياقات لغوية متنوعة. ويسمح التصميم المعياري للبرنامج أيضًا بمزيد من التطورات في مجال الذكاء الاصطناعي مثل التحسينات في التعرف على اللغة في الوقت الفعلي والترجمة المتنوعة لبنى الجمل المعقدة ليتم دمجها دون الحاجة إلى إجراء تغييرات كبيرة في الإطار الحالي

المزامنة التلقائية للترجمة

إن أحد أكثر الجوانب المبتكرة في البرنامج الإضافي الخاص بنا هو قدرته على الاستفادة من الذكاء الاصطناعي لمزامنة الترجمة مع الترجمة المناسبة بلغة الإشارة. حيث أنه يجب في الأنظمة التقليدية ضبط توقيت مقاطع فيديو لغة الإشارة بعناية ومزامنتها يدويًا مع محتوى الفيديو وهي عملية شاقة خاصة عند التعامل مع مقاطع الفيديو التي يتم تحديثها بشكل متكرر. ففي كل مرة تتم فيها إضافة ترجمة جديدة أو تعديلها يلزم إعادة تسجيل الترجمة بلغة الإشارة بالكامل متبوعة بإعادة دمجها في الفيديو الأمر الذي يستهلك الوقت والموارد

ويتكيف حل (SignForDeaf) المدعوم بالذكاء الاصطناعي بشكل ديناميكي مع أي تغييرات يتم إجراؤها في ملفات الترجمة ويقوم تلقائيًا بتحديث الترجمة بلغة الإشارة المقابلة دون الحاجة إلى تدخل يدوي. وتتيح هذه الميزة مزامنة تستجيب للتحديثات المتكررة، ولا يضمن مستوى الأتمتة توفير الدقة فحسب بل يوفر أيضًا القدرة على التعامل مع كميات كبيرة من المحتوى مما يجعله قابلاً للتطوير للمؤسسات التي تنتج مخرجات فيديو متكررة ومتنوعة وهي ميزة تفتقر إليها الحلول السابقة

نافذة لغة الإشارة القابلة للتخصيص

تتمثل إحدى الميزات الرئيسية التي تميز البرنامج الإضافي الخاص بنا في نافذة لغة الإشارة القابلة للتخصيص بالكامل. حيث إنه غالبًا ما يتم انتقاد حلول إمكانية النفاذ لكونها غير مرنة ولكن هذا النظام يعطي الأولوية لتجربة المستخدم في تصميم نظام ترجمة بلغة الإشارة بالفيديو. ويتمتع المشاهدون بالتحكم الكامل في عرض نافذة لغة الإشارة مما يضمن إمكانية تكييفها مع تفضيلاتهم الفردية. ويتضمن هذا التخصيص خيارات لتحريك النافذة أو تغيير حجمها أو حتى تعطيلها حسب الحاجة مما يمنح المستخدمين المرونة لضبط العرض بناءً على بيئة المشاهدة والراحة الشخصية. فقد يفضل المستخدم الذي يشاهد مقطع فيديو على شاشة صغيرة مثل الهاتف المحمول على سبيل المثال تصغير حجم نافذة لغة الإشارة أو نقلها إلى زاوية الشاشة حيث لا تحجب العناصر المرئية المهمة. وعلى العكس من ذلك فقد يختار المستخدم الذي يشاهد على شاشة أكبر تكبير النافذة لمنحه مزيداً من مجال الرؤية

ويضمن هذا النهج الذي يركز على المستخدم ألا تتداخل الترجمة بلغة الإشارة مع محتوى الفيديو الرئيسي مع الحفاظ على إمكانية النفاذ إليه بسهولة حسب الحاجة. كما يمكن للمستخدمين تمكين أو تعطيل ميزة لغة الإشارة في أي وقت مما يضمن مشاهدة الفيديو دون تشتيت لأولئك الذين يختارون تعطيلها. ويمكن أن تعمل هذه المرونة على تحسين إمكانية النفاذ إلى محتوى الفيديو وشموله بشكل كبير وخاصة بالنسبة لمجتمعات الصم وضعاف السمع

تقليل الوقت والتكلفة

عادةً ما تكون الطرق التقليدية لإضافة الترجمة بلغة الإشارة مكلفة وبطيئة. حيث تتضمن العملية عادةً توظيف مترجم لغة إشارة محترف وتسجيل ترجمته في الاستوديو ثم دمج فيديو لغة الإشارة في المحتوى الرئيسي من خلال التحرير بعد الإنتاج. وتتطلب هذه الطريقة موارد بشرية ومالية كبيرة مما قد يجعلها غير مجدية لمنشئي المحتوى الصغار أو المنظمات ذات الميزانيات المحدودة. وعلاوة على ذلك فإن أي تغييرات على الفيديو الأصلي أو نص الترجمة ستتطلب تكرار هذه العملية بالكامل مما يؤدي إلى خسائر في التكلفة والوقت

ويقضي البرنامج الإضافي الخاص بنا على هذه التحديات من خلال إزالة الحاجة إلى إعادة التصوير المستمر والتكامل اليدوي. فبمجرد تنصيب البرنامج الإضافي يتم إنشاء الترجمة بلغة الإشارة تلقائيًا بناءً على الترجمة النصية مما يعني أن أي تغييرات على الترجمة ستنعكس على الفور في الترجمة بلغة الإشارة. ويقلل هذا بشكل كبير من خسائر الوقت والتكلفة المطلوبة للحفاظ على محتوى الفيديو القابل للنفاذ مما يتيح للمزيد من المبدعين والمعلمين والمنظمات تقديم الترجمة بلغة الإشارة دون العبء المالي الإضافي. ولنأخذ مؤسسة تعليمية كبيرة أو منصة إعلامية على سبيل المثال حيث يتم إنتاج مئات مقاطع الفيديو كل شهر وهنا يمكن أن يكون توفير الوقت كبيراً مما يسمح بتخصيص الموارد لإنشاء محتوى جديد أو تحسين ميزات إمكانية النفاذ الأخرى بدلا من ذلك

بالإضافة إلى ذلك فإن انخفاض تكلفة التنفيذ يعني أن المنظمات الأصغر أو المبدعين المستقلين الذين ربما لم يتمكنوا في السابق من تحمل تكاليف دعم لغة الإشارة سيمكنهم الآن تقديم مقاطع فيديو قابلة للنفاذ بالكامل. ويوسع هذا الأمر من نطاق المحتوى القابل للنفاذ عبر منصات وصناعات مختلفة من مقاطع الفيديو التعليمية والدورات التدريبية عبر الإنترنت إلى المواد الترفيهية والتدريبية للشركات. وفي نهاية المطاف لن تنحصر هذه الفائدة في مجتمع الصم وضعاف السمع من خلال توفير محتوى أكثر شمولاً، بل ستشجع أيضًا على تبني ممارسات إمكانية النفاذ على نطاق أوسع عبر المشهد الإعلامي

الخاتمة

يمثل تطبيق أنظمة الترجمة الآلية بلغة الإشارة تقدماً مهماً في جعل المحتوى الرقمي أكثر نفاذاً بالنسبة لمجتمعات الصم وضعاف السمع. ومن خلال معالجته لقيود النفاذ التقليدية القائمة على الترجمة فإن هذا الحل يقدم خيارًا أكثر شمولاً وديناميكية لتوفير ترجمة متزامنة بلغة الإشارة. ولا يخفف هذا النهج الأعباء الزمنية والمالية المرتبطة عادةً بإنشاء مقاطع فيديو بلغة الإشارة فحسب بل يضمن أيضًا إمكانية تحديث ميزات إمكانية النفاذ بسهولة مع تطور المحتوى. وهذا أمر مهم في السياقات التعليمية والترفيهية والمهنية حيث يتم تحديث المحتوى بشكل متكرر.[2]

يفتح هذا النظام إمكانيات جديدة لمستقبل إمكانية النفاذ في وسائل الإعلام. فمع استمرار الذكاء الاصطناعي في التقدم في مجالات مثل معالجة اللغة الطبيعية والتعرف على الإيماءات يمكننا أن نتوقع ترجمات أكثر تطورًا ودقة من حيث السياق مما قد يؤدي إلى سد الفجوات بين لغات الإشارة المختلفة واللغات المنطوقة في جميع أنحاء العالم.[5] ويعتبر هذا الأمر هاماً ليس فقط لمجتمع الصم وضعاف السمع ولكن للمجتمع ككل لأنه يعزز الشمول ويكسر حواجز الاتصال عبر الخطوط اللغوية والثقافية.[4] فمن خلال دمج الترجمة بلغة الإشارة المدعومة بالذكاء الاصطناعي في منصات الفيديو السائدة ستتاح الفرصة لمنشئي المحتوى والمنظمات لتحسين تجربة المشاهدة بشكل كبير لجمهورهم. وتسمح الطبيعة القابلة للتخصيص وسهلة الاستخدام للبرنامج الإضافي للترجمة بلغة الإشارة بالفيديو للأفراد بتخصيص تجربة إمكانية النفاذ الخاصة بهم مما يجعله أداة متعددة الاستخدامات يمكنها استيعاب مجموعة واسعة من الاحتياجات الشخصية. كما يمكننا مع تزايد عدد المنظمات التي تتبنى هذه التقنيات أن نتوقع تحولاً ثقافياً أوسع نحو تطبيع إمكانية النفاذ إلى المحتوى الرقمي مما يعود بالنفع ليس فقط على مجتمع الصم وضعاف السمع بل وأيضاً على فهم المجتمع للشمول.[6]

في الختام نرى أنه للاستخدام الواسع النطاق لأدوات الترجمة بلغة الإشارة المدعومة بالذكاء الاصطناعي القدرة على إعادة تشكيل مشهد الفيديو على شبكة الإنترنت مما يجعل المحتوى الرقمي متاحًا للجميع. إنه حل لا يعالج تحديات إمكانية النفاذ الحالية فحسب بل يمهد الطريق أيضًا للتقدم المستقبلي مما يضمن عدم تخلف أي شخص عن الركب مع تقدم التكنولوجيا.

الشكر

نود أن نشكر زملائنا وشركاء البحث على مساهماتهم القيمة في هذا المشروع

المراجع

1. Wilson, M., & Moffat, P. (2018). The impact of subtitles and sign language on video accessibility. Journal of Deaf Studies and Deaf Education, 23(2), 204-215. doi:10.1093/deafed/eny012
2. Arfé, B., Rossi, C., & Sicoli, S. (2014). The role of sign language in reading comprehension for deaf individuals. Frontiers in Psychology, 5, 1174. doi:10.3389/fpsyg.2014.01174
3. World Federation of the Deaf (2021). Global accessibility report on sign language use in media. WFD Publications, pp. 12-36.
4. Liddell, S. K. (2003). Grammar, Gesture, and Meaning in American Sign Language. Cambridge University Press, pp. 45-89. ISBN: 9780521016505
5. Berke, J. (2020). AI and Accessibility: Bridging the communication gap for the deaf community. AI Journal of Linguistics, 15(3), 235-250. doi:10.1111/ail.153235
6. Napier, J., Leigh, G., & Goswell, D. (2016). Sign Language Interpreting: Theory and Practice in Australia and New Zealand. Federation Press, pp. 102-145. ISBN: 9781760021162