

# التعرف التلقائي على لغة الإشارة العربية استنادًا إلى الإيماءات: نهج التعلم العميق الموزع

أحمد الزعبي<sup>1</sup>، توفيق الحضرمي<sup>2</sup>، أمجد البشايرة<sup>1</sup>، لجين بني يونس<sup>1</sup>

<sup>1</sup> قسم علوم الحاسوب، جامعة العلوم والتكنولوجيا الأردنية، إربد، الأردن

[amalbashayreh20@cit.just.edu.jo](mailto:amalbashayreh20@cit.just.edu.jo)

[agalzubi@just.edu.jo](mailto:agalzubi@just.edu.jo)

[lhbaniyounis19@cit.just.edu.jo](mailto:lhbaniyounis19@cit.just.edu.jo)

<sup>2</sup> كلية العلوم والتكنولوجيا، جامعة نوتنجهام ترينت، نوتنجهام، المملكة المتحدة

[tawfik.al-hadhrami@ntu.ac.uk](mailto:tawfik.al-hadhrami@ntu.ac.uk)

## الملخص:

يعد استخدام خوارزميات التعلم الآلي للتعرف على أنماط إيماءات اليد المخصصة للأشخاص ذوي الإعاقة اتجاهًا متناميًا 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 العميق الموزع باستخدام 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) **آلية المتوسط الموزع**. 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 الموزع، التعرف على الصور، إمكانية النفاذ، صعوبات التواصل.

## 1. المقدمة

بما أن لغة الإشارة تُعد 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., 2021). ونظراً للمجموعة المتنوعة من الفرص المتاحة فقد تمتد حلول التكنولوجيا المساعدة هذه إلى ما هو أبعد من مجرد الترجمة. فقد تسمح بيث 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 et al., 2019) (Rastgoo et al., 2021) 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 (Cheok et al., 2023) (Elsheikh et al., 2019).

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 العميق الموزع 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 العميق الموزع 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 طلاً 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 العميق الموزع وقد دفعنا هذا إلى معالجة هذه الفجوة من خلال الاستفادة من نموذج التعلم العميق الموزع 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 محلياً 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 العميق الموزع 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 الأحرف الأبجدية العربية بلغة الإشارة (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 خصيصاً 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).



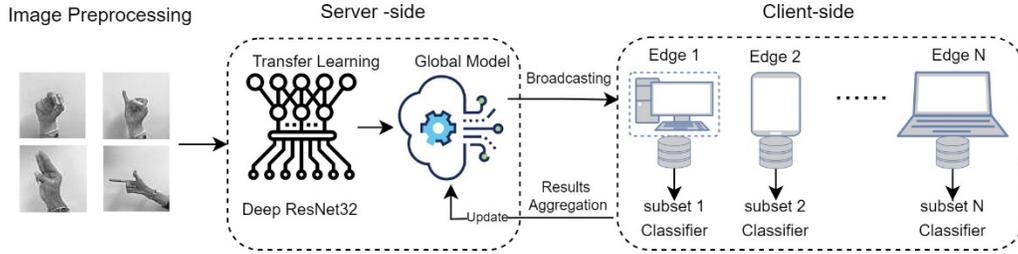
الشكل 1. عينة من صور الإشارات العربية من مجموعة بيانات ARASL.

تم تعديل بيانات أبجدية لغة الإشارة العربية بحيث تتيح تغيير حجم الصورة، وتحويل البيانات إلى تينسر (Tensor Conversion)، وتطبيعها (Normalization) إلى المدى [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 العميق الموزع.

### 3. المنهجية

يوضح الشكل 2 الإطار العام لهندسة التعلم العميق الموزع المقترحة والتي تتضمن خادماً مركزياً يتفاعل مع عملاء متعددين يعملون كعقد حوسبة موزعة. ويستضيف الخادم نموذجاً 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 درجات معاملات (معاملات) النموذج المحلي ويقوم

بتحديث النموذج الشامل ثم يرسله مرة أخرى إلى العملاء.



الشكل 2. مسار عملية التعلم العميق الموزع للتعرف على الإشارات العربية.

لقد استخدمت هذه الدراسة آلية المتوسط الموزع (McMahan 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.

تم استخدام (ResNet32) (He et al., 2016)، وهي بنية شبكة عصبية عميقة معروفة في إطار التعلم العميق الموزع 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.

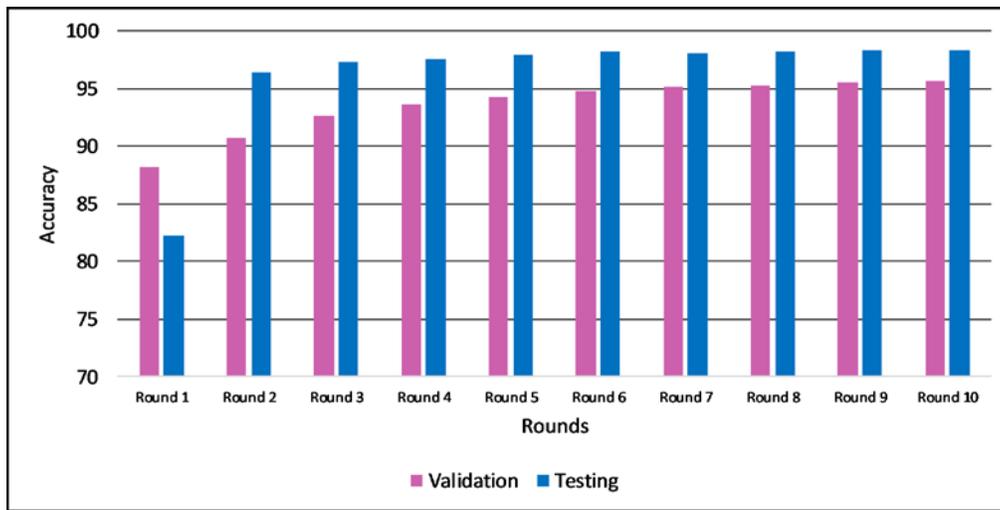
#### 4. النتائج التجريبية

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). يتم تجميع التدرجات باستخدام متوسط التوزيع الموزع (FedAvg) في 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

المؤشرات الصحيحة والخاطئة (TP) و (TN) و (FP) و (FN) 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 نظرًا لأن مجموعة بيانات لغة الإشارة العربية غير متوازنة غير متوازنة. ويعامل المتوسط الكلي للقياسات، مع معاملة جميع الفئات بالتساوي 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.

البحرث المبدع	النموذج	دقة الاختبار	عمليات المبرمج محلياً على
Kamruzzaman et al.	CNN	90.0	100
Aldehbi et al. (2022)	MobileNet	94.5	15
Zakariah et al. (2022)	EfficientNet	95.0	20
This Work	FL-ResNet32	98.3	10

## 5. مناقشة

تؤكد هذه الدراسة على مدى أهمية وجود بيانات حوسبة موزعة في إطار التعلم العميق الموزع 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 اتصالًا بالشبكة بنطاق ترددي كافٍ لمجموعات البيانات الكبيرة وزمن انتقال منخفض لضمان التنبؤات في الوقت المناسب (Diaz et al., 2023). بالإضافة إلى ذلك تتطلب تبيعة 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 العميق الموزع.

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 الممكنة 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) (Bennbaia, 2022) (Othman and El Ghouli, 2022) نحو تطوير تقنيات المترجم الافتراضي للإشارة (الأفاتار) بما يتناسب مع الثقافة المحلية ويمكن هذا الأمر الأفراد الصم وضعاف السمع من الانخراط في الحياة المجتمعية مما يؤدي إلى ظهور مناهج اتصال أكثر ديناميكية وقابلية للتكيف.

تعد 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 من الحلول التفاعلية المتقدمة لمشكلة إلى محتوى لغة الإشارة. وستستفيد هذه التقنية القائمة على الشخصيات الافتراضية من التعلم العميق الموزع 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 الإعاقات التواصلية.

## المراجع

- 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.
- 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.
- 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).
- 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.
- Bennbaia, S. (2022). Toward an evaluation model for signing avatars. *Nafath*, 6(20).
- 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.
- 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.
- 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.
- Elsheikh, A. (2023). Enhancing the Efficacy of Assistive Technologies through Localization: A Comprehensive Analysis with a Focus on the Arab Region. *Nafath*, 9(24).

- 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.
- Kamruzzaman, M. et al. (2020). Arabic sign language recognition and generating Arabic speech using convolutional neural network. *Wireless Communications and Mobile Computing*, 2020.
- Krishnan, R., & Manickam, S. (2024). Enhancing Accessibility: Exploring the Impact of AI in Assistive Technologies for Disabled Persons. *Nafath*, 9(25).
- Latif, G., Mohammad, N., Alghazo, J., AlKhalaf, R., and AlKhalaf, R. (2019). Arasl: Arabic alphabets sign language dataset. *Data in brief*, 23:103777.
- 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*.
- 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.
- 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*.
- Othman, A., & El Ghouli, O. (2022). BuHamad: The first Qatari virtual interpreter for Qatari Sign Language. *Nafath*, 6(20).
- Rastgoo, R., Kiani, K., & Escalera, S. (2021). Sign language recognition: A deep survey. *Expert Systems with Applications*, 164, 113794.
- Saleh, Y., & Issa, G. (2020). Arabic sign language recognition through deep neural networks fine-tuning. *International Association of Online Engineering*, 71-83.
- Semreen, S. (2023). Sign languages and Deaf Communities. *Nafath*, 9(24).
- Talov, M. C. (2022). SpeakLiz by Talov: Toward a Sign Language Recognition mobile application. *Nafath*, 7(20).
- 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.

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.

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.