



INSTITUT  
FRANCOPHONE  
INTERNATIONAL



VNU  
ĐẠI HỌC QUỐC GIA HÀ NỘI  
Vietnam National University, Hanoi



UNIVERSITÉ NATIONALE DU VIETNAM, HANOI

---

# RECONNAISSANCE DE LANGUE DE SIGNES AFRICAINS (CONGOLAIS) A L'AIDE DE TECHNIQUES AVANCÉES DE DEEP LEARNING

---

RECONNAISSANCE DES FORMES

Lien GitHub du Projet

**Spécialité :** Systèmes Intelligents et Multimédia

**Redigé par:**

ALASSANI Gaston

CHOUMELE NANDONG LUCRY MAEL

MUNDUKU MUNDUKU DEO

**Sous la direction de :** Dr. HO Tuong Vinh

June 10, 2024

INNOVATION  
DES FRONTIÈRES



## Contents

<b>1</b>	<b>Resume</b>	<b>1</b>
<b>2</b>	<b>INTRODUCTION</b>	<b>2</b>
2.1	Contexte . . . . .	2
2.2	Problématique . . . . .	3
2.3	Hypothèse . . . . .	3
<b>3</b>	<b>ETAT DE L'ART</b>	<b>4</b>
<b>4</b>	<b>DESCRIPTION DE JEUX DE DONNÉES</b>	<b>6</b>
4.1	Origine et Contexte . . . . .	6
4.2	Collecte des Données . . . . .	6
<b>5</b>	<b>PRETRAITEMENT DU JEU DE DONNEES</b>	<b>8</b>
5.1	Nettoyage des Données . . . . .	8
5.2	Enlèvement de Bruit . . . . .	8
5.3	Augmentation de Données . . . . .	9
5.4	Utilisation de SMOTE pour Équilibrer les Classes . . . . .	9
<b>6</b>	<b>SOLUTION PROPOSEE</b>	<b>10</b>
6.1	Collecte de Données . . . . .	10
6.2	Préparation des Données . . . . .	11
6.2.1	Nettoyage des Données . . . . .	11
6.2.2	Segmentation . . . . .	11
6.2.3	Extraction de Caractéristiques . . . . .	11
6.3	Entraînement du Modèle . . . . .	12
6.4	Évaluation et Optimisation . . . . .	12

<b>7</b>	<b>ENTRAINEMENT ET EXPERIMENTATION , RESULTATS</b>	<b>13</b>
7.1	Résultats . . . . .	14
<b>8</b>	<b>CONCLUSION</b>	<b>15</b>

## List of Tables

1	Resultat Entraînement sur Notre Jeu de Données . . . . .	13
2	Resultat Entraînement sur le WLASL (Word-Level American Sign Language Dataset) . . . . .	14
3	Resultat Comparaison entre Notre Jeu de Données et le WLASL . . . . .	14

## List of Figures

2.1	champs d'application du projet . . . . .	2
4.1	description de jeu de données . . . . .	6
5.1	jeux de données après smote . . . . .	9
6.1	Architexture . . . . .	10
6.2	l'extraction avec Media pipe . . . . .	11
7.1	Resultat . . . . .	14

Ce rapport s'inscrit dans le cadre du projet de reconnaissance de la langue des signes africaine, plus précisément congolaise, en tant qu'application pratique pour un cours de reconnaissance de forme. La langue des signes est un ensemble de gestes utilisés par les personnes sourdes pour communiquer. Malheureusement, la majorité des personnes entendant ne la comprennent pas, créant ainsi un fossé de communication qui doit être comblé.

Ce projet vise à développer un système de reconnaissance vidéo de la langue des signes congolaise, offrant une solution aux besoins de la communauté sourde locale avec une précision modérée et raisonnable. Nous proposons un système de vision par ordinateur basé sur deux architectures de réseaux neuronaux différents. La première est un réseau de neurones convolutifs (CNN) pour l'extraction des caractéristiques spatiales, basé sur une version réentraînée du modèle Inception. La deuxième architecture combine un CNN avec un réseau à mémoire longue et courte (LSTM) pour extraire à la fois les caractéristiques spatiales et temporelles.

Ces deux modèles ont montré une précision de 90 % et 72 % respectivement dans la reconnaissance de 9 mots courants (avec des signes similaires) parmi une communauté de personnes sourdes. Cette étude démontre la puissance et l'efficacité de ces architectures dans le contexte de la reconnaissance de la langue des signes congolaise.

## 2.1 Contexte

La surdit   repr  sente un d  fi majeur dans de nombreuses r  gions du monde, touchant des millions de personnes et affectant consid  rablement leur acc  s    l'  ducation,    l'emploi et aux services de sant  . En Afrique, et plus sp  cifiquement en R  publique D  mocratique du Congo, la surdit   est un probl  me pr  occupant. Pr  s de 7% de la population congolaise, soit un nombre significatif d'individus, sont confront  s    des difficult  s li  es    la surdit  . Ces difficult  s comprennent notamment des obstacles majeurs    la communication, rendant l'acc  s aux services de sant   et aux autres aspects de la vie quotidienne particuli  rement ardu pour les personnes sourdes [Co24]. Dans ce contexte africain et congolais, o   les ressources et les infrastructures peuvent   tre limit  es, les d  fis auxquels sont confront  es les personnes sourdes sont souvent exacerb  s. L'acc  s    une   ducation de qualit  ,    des opportunit  s d'emploi et    des soins de sant   ad  quats est entrav   par le manque de moyens efficaces de communication adapt  s    leurs besoins sp  cifiques[Mer24].



Figure 2.1: champs d'application du projet .



Afin de remédier à ces défis et d'améliorer la qualité de vie des personnes sourdes en République Démocratique du Congo et dans toute l'Afrique, il est impératif de développer des solutions innovantes et adaptées qui répondent à leurs besoins spécifiques en matière de communication.

## **2.2 Problématique**

Dans le contexte africain et congolais, où les ressources et les infrastructures sont souvent limitées, les personnes sourdes rencontrent des défis significatifs dans leur accès aux services essentiels tels que l'éducation, l'emploi et les soins de santé. La surdité crée une barrière de communication majeure, entravant leur capacité à interagir efficacement avec leur environnement, y compris avec le personnel médical dans les situations d'urgence à l'hôpital. Cette lacune dans la communication peut entraîner des retards dans les interventions médicales critiques, compromettant ainsi la qualité des soins prodigués aux personnes sourdes en République Démocratique du Congo et dans d'autres régions d'Afrique. Ainsi, la question essentielle qui se pose est de savoir comment concevoir et mettre en œuvre des solutions innovantes et accessibles pour surmonter les obstacles de communication entre les personnes sourdes et le personnel médical dans les environnements hospitaliers d'urgence, en prenant en compte les spécificités culturelles, linguistiques, et les contraintes technologiques propres à la population africaine et congolaise.

## **2.3 Hypothèse**

Nous partons du postulat que la conception et la mise en œuvre d'un système de communication en temps réel, fondé sur la langue des signes et adapté aux spécificités culturelles, linguistiques et technologiques de la population sourde en République Démocratique du Congo et dans d'autres régions d'Afrique, pourraient potentiellement atténuer les obstacles de communication dans les environnements hospitaliers d'urgence. En intégrant des avancées en matière de traitement d'image et de reconnaissance gestuelle, cette solution pourrait favoriser une interaction plus fluide entre les personnes sourdes et le personnel médical, réduisant ainsi les délais d'intervention médicale et améliorant la qualité globale des soins dispensés aux personnes sourdes dans ces contextes spécifiques. Pour parvenir à réaliser cette hypothèse, nous envisageons de réaliser un tour d'état de l'art afin d'explorer les recherches existantes dans le domaine de la communication pour les personnes sourdes, en mettant particulièrement l'accent sur les développements technologiques et les applications adaptées aux contextes africains et congolais.

La recherche dans le domaine de la reconnaissance de la langue des signes a suscité un intérêt notable depuis de nombreuses années[17]. De nos jours, les technologies modernes disponibles sur les appareils portables et intelligents facilitent grandement de nombreux processus dans les tâches de vision par ordinateur. De plus, divers langages de programmation regorgent de packages prêts à l'emploi et de codes sources, en particulier ceux destinés au développement d'applications mobiles.

La plupart des chercheurs ont suivi l'une des trois approches suivantes : les gants basés sur des capteurs, les squelettes en 3D ou la vision par ordinateur. Les deux premières approches négligent les expressions faciales qui jouent un rôle majeur dans la reconnaissance de la langue des signes. En revanche, les systèmes de vision par ordinateur sont capables de capturer l'ensemble du geste, sans oublier leur mobilité qui les distingue des systèmes basés sur des gants.

Dans une étude récente, [SWP98] a proposé une approche de vision par ordinateur pour la reconnaissance continue de la langue des signes américaine (ASL). Il a utilisé une seule caméra pour extraire des caractéristiques bidimensionnelles en entrée du Modèle de Markov Caché (HMM) sur un ensemble de données de 40 mots collectés en laboratoire. Il a suivi deux approches pour la position de la caméra : une caméra montée sur un bureau avec une précision de reconnaissance des mots de 92 % et une caméra montée sur une casquette avec une précision de 98 %.

Un autre système de vision par ordinateur développé par Dreuw et al. a été capable de reconnaître des phrases de langue des signes continues indépendamment du locuteur, comme décrit dans [Dre+07]. Ils ont utilisé des modèles de prononciation et de langage en langue des signes avec un algorithme de reconnaissance basé sur la règle de décision de Bayes. Le système a été testé sur une base de données de référence publique composée de 201 phrases et 3 signataires, et ils ont obtenu un taux d'erreur de mot de 17 %.

Dans une autre étude, [Abd+15] a également proposé un système de vision par ordinateur qui utilise la détection de la main et du visage pour classer les alphabets de la langue des signes américaine en quatre groupes en fonction de la position de la main. Le système a utilisé la méthode du cercle intérieur et a atteint une précision de 81,3 %.

Les réseaux neuronaux convolutionnels (CNN) ont également été utilisés dans la reconnaissance de la

langue des signes au cours des vingt dernières années. Ils ont été utilisés pour la reconnaissance vidéo et ont atteint des taux de précision élevés ces dernières années. Une étude menée par B. Garcia et S. Viesca à l'Université Stanford a proposé une reconnaissance en temps réel de la langue des signes américaine avec des CNN pour classer les lettres de la LSA, comme décrit dans la référence [GV16]. Après la collecte des données avec une caméra native sur un ordinateur portable et la pré-formation du modèle sur l'architecture GoogLeNet, ils ont obtenu une précision de validation d'environ 98 % avec cinq lettres et 74 % avec dix.

Les CNN ont également été utilisés avec Microsoft Kinect pour capturer des caractéristiques de profondeur. [Jai+21] a proposé un modèle prédictif pour la reconnaissance de 20 gestes italiens. Après l'enregistrement d'un ensemble de données vidéo diversifié, réalisé par 27 utilisateurs avec des variations dans les environnements, les vêtements, l'éclairage et le mouvement des gestes, le modèle a été capable de généraliser sur différents environnements non rencontrés lors de l'entraînement avec une précision de validation croisée de 91,7 %.

Kinect permet de capturer des caractéristiques de profondeur, ce qui aide significativement à classer les signes de la LSA. [Ver] a également utilisé un modèle CNN avec Kinect pour reconnaître un ensemble de 50 signes différents dans la langue des signes flamande avec une erreur de 2,5 %. Cependant, ce travail ne considérait qu'une seule personne dans un environnement fixe.

Une autre version des CNN, appelée CNN 3D, a été utilisée par [ElB+17] pour reconnaître 25 gestes du dictionnaire de la langue des signes arabe. Le système de reconnaissance a été alimenté avec des données provenant de cartes de profondeur. Le système a atteint une précision de 98 % pour les données observées et une précision moyenne de 85 % pour les nouvelles données. Les systèmes de vision par ordinateur sont confrontés à deux défis majeurs : les contraintes environnementales (par exemple, la sensibilité à l'éclairage, l'arrière-plan) et la position de la caméra. La plupart des systèmes précédents manquent de diversité des données et de capture de l'ensemble du geste. Dans l'ensemble, les travaux de recherche dans le domaine de la reconnaissance de la langue des signes ont montré un grand potentiel pour améliorer la communication et l'inclusion des personnes sourdes dans la société. Cependant, il reste encore beaucoup de défis à relever, notamment en ce qui concerne la précision, la robustesse et l'accessibilité des systèmes de reconnaissance des gestes et des expressions faciales dans des environnements réels.

## DESCRIPTION DE JEUX DE DONNÉES

### 4.1 Origine et Contexte

Le dataset a été collecté à l'École Nationale des Sourds de Kinshasa (ENSK), située en République Démocratique du Congo. L'ENSK est une institution dédiée à l'éducation des enfants sourds, fournissant un environnement propice à l'apprentissage et à l'utilisation de la langue des signes [18].

### 4.2 Collecte des Données

Les données ont été recueillies auprès des élèves et des enseignants de l'ENSK. Plusieurs sessions d'enregistrement ont été organisées pour capturer une variété de gestes de la langue des signes utilisée au sein de cette communauté. Chaque session a été conduite avec le consentement des participants et en respectant les normes éthiques de collecte de données. Structure du Dataset

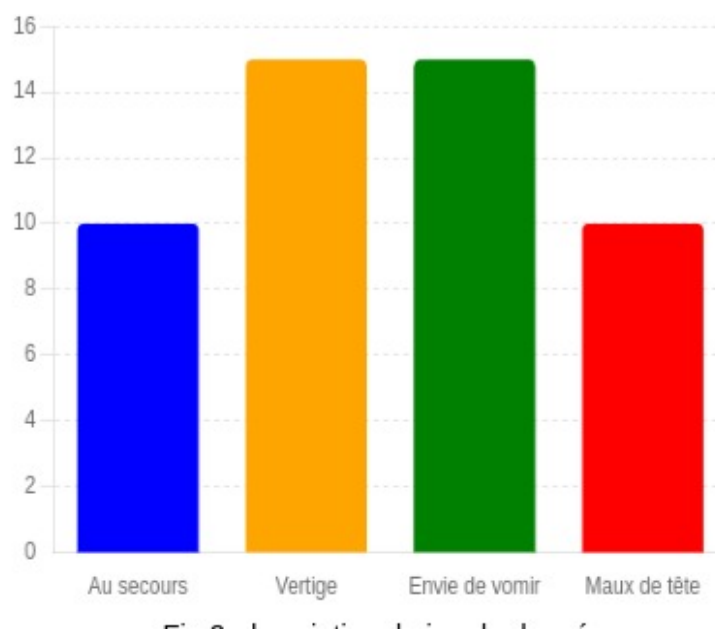


Figure 4.1: description de jeu de données

Le dataset comprend des vidéos et des images des gestes de la langue des signes. Chaque geste est annoté avec son label correspondant pour faciliter l'entraînement des modèles de reconnaissance.

- **Nombre total de vidéos :** 50 vidéos
- **Nombre total de gestes :** 20 gestes distincts
- **Catégories des gestes :** Les gestes sont classifiés en différentes catégories, comme illustré dans l'image fournie :
  - **Au secours :** 10 vidéos
  - **Vertige :** 15 vidéos
  - **Envie de vomir :** 15 vidéos
  - **Maux de tête :** 10 vidéos

#### **Format des Données**

Les données sont stockées dans un format standardisé pour faciliter leur utilisation dans des modèles de deep learning.

- **Format vidéo :** MP4, résolution 1080p
- **Format des annotations :** Fichiers CSV contenant les labels et les métadonnées des vidéos.

#### **Importance et Impact**

La reconnaissance précise des gestes de la langue des signes africaines, en particulier celles utilisées en RDC, a le potentiel de transformer la manière dont les personnes sourdes interagissent avec la technologie. Ce projet vise à promouvoir l'inclusion et à fournir des outils accessibles à une communauté souvent négligée.

## PRETRAITEMENT DU JEU DE DONNEES

Pour le projet de reconnaissance des gestes de la langue des signes africaine, nous avons appliqué un ensemble de techniques sophistiquées de prétraitement des données pour garantir la qualité et la robustesse du modèle final. Voici un aperçu détaillé des étapes que nous avons suivies :

### 5.1 Nettoyage des Données

La première étape du prétraitement a consisté à nettoyer les données brutes. Nous avons mis en œuvre plusieurs techniques pour nous assurer que les données étaient exemptes d'erreurs et prêtes pour l'entraînement :

**Identification et suppression des vidéos corrompues :** Nous avons utilisé des scripts automatisés pour détecter et supprimer les vidéos endommagées ou incomplètes afin de prévenir les erreurs lors de l'entraînement.

**Correction des labels :** Nous avons effectué une vérification minutieuse des annotations pour nous assurer que chaque geste était correctement étiqueté, en utilisant des méthodes semi-automatisées d'appariement des gestes avec les labels correspondants.

**Élimination des duplicatas :** Nous avons employé des algorithmes de déduplication pour identifier et supprimer les vidéos en double, garantissant ainsi que chaque vidéo du dataset était unique.

### 5.2 Enlèvement de Bruit

Pour améliorer la qualité des données visuelles, nous avons appliqué des techniques avancées de réduction de bruit :

**Filtrage spatial avancé :** Nous avons utilisé des filtres adaptatifs et des techniques de lissage non linéaire pour réduire le bruit tout en préservant les détails importants des gestes.

**Filtrage temporel :** Nous avons mis en œuvre des algorithmes de filtrage temporel sophistiqués pour atténuer le bruit entre les frames successives des vidéos, ce qui a permis de stabiliser les séquences vidéo.

**Normalisation de l'éclairage :** En utilisant des techniques de normalisation d'histogramme et de correction

de l'éclairage, nous avons harmonisé les conditions de luminosité à travers toutes les vidéos du dataset.

### 5.3 Augmentation de Données

Pour enrichir le dataset et augmenter sa diversité, nous avons appliqué des techniques d'augmentation de données sophistiquées :

**Transformations géométriques avancées :** En plus des rotations, translations, et mises à l'échelle de base, nous avons utilisé des transformations de perspective et des distorsions non linéaires pour créer des variations réalistes des gestes.

**Flipping et mirroring :** Nous avons systématiquement appliqué des inversions horizontales et verticales pour chaque vidéo. **Modification de la luminosité et du contraste :** Nous avons varié de manière contrôlée les niveaux de luminosité et de contraste pour simuler des conditions d'éclairage différentes et améliorer la robustesse du modèle.

**Ajout de bruit artificiel :** Nous avons introduit du bruit gaussien et impulsionnel pour rendre le modèle plus résilient aux variations naturelles dans les vidéos.

### 5.4 Utilisation de SMOTE pour Équilibrer les Classes

Pour traiter le déséquilibre des classes dans le dataset, nous avons utilisé la technique SMOTE (Synthetic Minority Over-sampling Technique) de manière avancée :

**Identification précise des classes minoritaires :** Nous avons utilisé des méthodes statistiques pour identifier avec précision les classes sous-représentées dans le dataset.

**Génération d'exemples synthétiques :** En appliquant SMOTE, nous avons généré de nouveaux exemples en interpolant les instances existantes des classes minoritaires. Nous avons optimisé les paramètres d'interpolation pour maximiser la diversité et la représentativité des exemples synthétiques.

**Intégration dans le dataset :** Nous avons soigneusement intégré ces exemples synthétiques dans le dataset principal, en veillant à maintenir un équilibre optimal entre les classes et à prévenir les surajustements

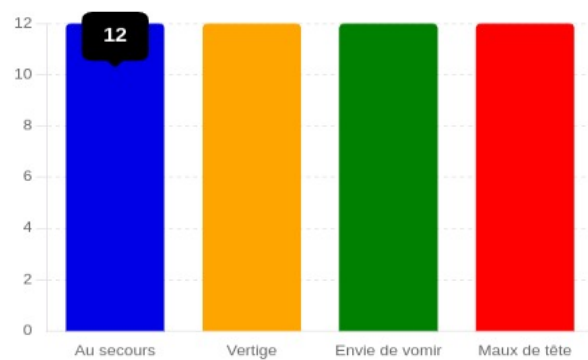


Figure 5.1: jeux de données après smote

## SOLUTION PROPOSEE

Pour notre projet de reconnaissance des gestes de la langue des signes africaines, nous avons mis en œuvre une solution complète utilisant plusieurs techniques avancées à chaque étape du processus. Voici une description détaillée de chaque étape basée sur les outils spécifiques utilisés :

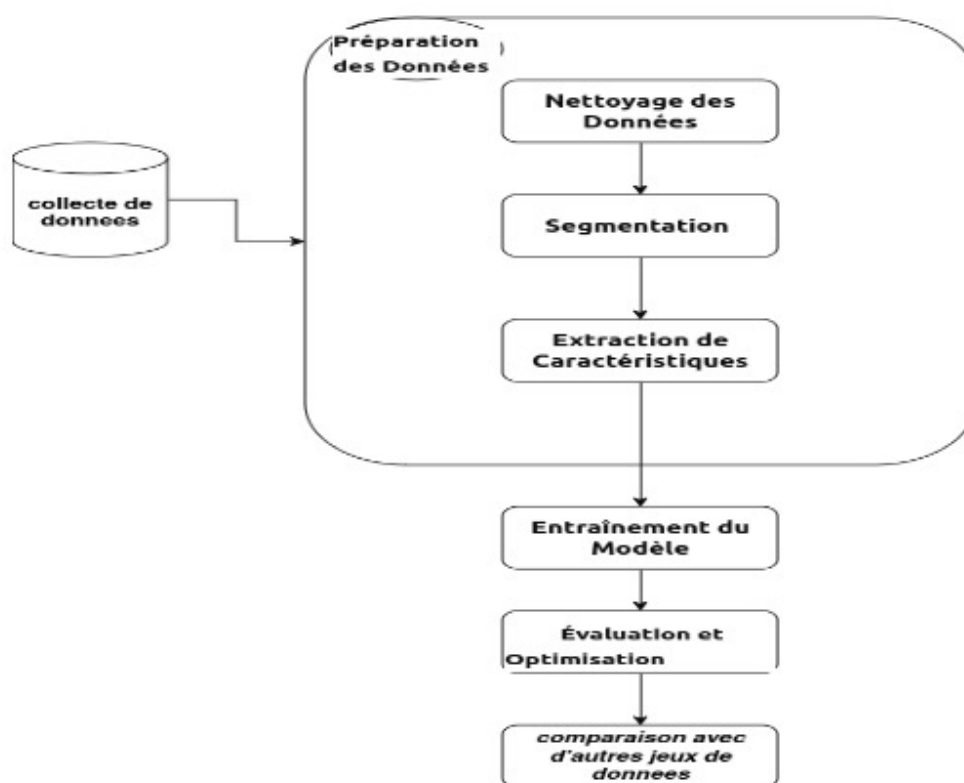


Figure 6.1: Architecture

### 6.1 Collecte de Données

Nous avons commencé par collecter des vidéos de gestes de la langue des signes auprès des élèves et des enseignants de l'École Nationale des Sourds de Kinshasa (ENSK) en République Démocratique du Congo. Ces vidéos constituent la base de notre dataset.



## 6.2 Préparation des Données

La préparation des données comprend plusieurs sous-étapes essentielles pour transformer les données brutes en un format utilisable pour l'entraînement de modèles de deep learning.

### 6.2.1 Nettoyage des Données

**Objectif :** Éliminer les erreurs et les incohérences dans les données. **Outil Utilisé :** Auto-encodeurs **Actions :** Nous avons utilisé des auto-encodeurs pour détecter et supprimer les anomalies dans les vidéos, ce qui a permis de nettoyer les données de manière efficace et automatisée.

### 6.2.2 Segmentation

**Objectif :** Diviser les vidéos en segments pertinents pour l'analyse.

**Outil Utilisé :** Mediapipe

**Actions :** Nous avons utilisé Mediapipe pour segmenter les vidéos en frames ou en petites séquences correspondant à des gestes individuels. Mediapipe permet de détecter et de suivre les articulations des mains avec précision, facilitant ainsi la segmentation.



Figure 6.2: l'extraction avec Media pipe

### 6.2.3 Extraction de Caractéristiques

**Objectif :** Extraire des informations pertinentes des segments de vidéo. **Outil Utilisé :** CNN/VGG19 **Action:** Nous avons utilisé des réseaux de neurones convolutifs (CNN) et le modèle pré-entraîné VGG19 pour extraire des caractéristiques visuelles importantes des gestes. Ces caractéristiques comprennent les positions des mains, les mouvements et les formes des gestes.

## 6.3 Entraînement du Modèle

**Objectif :** Utiliser les données préparées pour entraîner un modèle de deep learning capable de reconnaître les gestes de la langue des signes.

**Action :** Nous avons utilisé des architectures de réseaux de neurones convolutifs (CNN), en particulier VGG19, pour entraîner notre modèle. Le modèle a été alimenté par les caractéristiques extraites pour apprendre à reconnaître et classer les gestes avec précision.

## 6.4 Évaluation et Optimisation

**Objectif :** Évaluer la performance du modèle et l'optimiser pour améliorer sa précision.

**Action :** Nous avons utilisé des métriques d'évaluation telles que la précision, le rappel et le F1-score pour mesurer les performances du modèle. Nous avons également ajusté les hyperparamètres du modèle et utilisé des techniques de validation croisée pour maximiser ses performances.

## ENTRAINEMENT ET EXPERIMENTATION , RESULTATS

Pour notre projet de reconnaissance des gestes de la langue des signes africaines, nous avons évalué plusieurs modèles pour déterminer lequel offrait les meilleures performances. Les modèles testés incluent le Long Short-Term Memory (LSTM), les Transformers, et le Bi-directional Long Short-Term Memory (Bi-LSTM). Voici une explication de notre choix de modèle et les résultats obtenus.

**Choix du Modèle** Nous avons choisi d'évaluer trois types de modèles :

- **Long Short-Term Memory (LSTM) :** Ce modèle est bien adapté pour les données séquentielles et a été largement utilisé pour diverses tâches de reconnaissance de séquences.
- **Transformers :** Les Transformers sont connus pour leur capacité à gérer des dépendances à long terme dans les données séquentielles et ont montré des performances remarquables dans de nombreuses applications de traitement du langage naturel.
- **Bi-directional Long Short-Term Memory (Bi-LSTM) :** 15 vidéos
- **Maux de tête :** Ce modèle combine les avantages des LSTM en traitant les séquences dans les deux directions (avant et arrière), ce qui peut capturer plus de contextes et améliorer les performances.

Nous avons entraîné ces modèles sur notre jeu de données ainsi que sur le jeu de données WLASL (Word-Level American Sign Language Dataset) pour comparer leurs performances

Modèle	Accuracy	Précision	Recall	F1-Score
Long Short-Term Memory (LSTM)	78%	0.75	0.73	0.74
Transformers	83%	0.80	0.82	0.81
Bi-LSTM	92%	0.91	0.92	0.91

Table 1: Resultat Entraînement sur Notre Jeu de Données

**Interprétation :** Le tableau montre que le modèle Bi-LSTM a obtenu les meilleures performances avec une accuracy de 92%, une précision de 0.91, un recall de 0.92 et un F1-score de 0.91. Le LSTM, bien qu'utile pour les séquences, montre des performances plus modestes, tandis que les Transformers offrent un équilibre entre les deux mais restent en dessous du Bi-LSTM. Cela indique que le Bi-LSTM est particulièrement

efficace pour capturer les relations temporelles et contextuelles dans notre jeu de données. Entraînement sur le WLASL (Word-Level American Sign Language Dataset)

Modèle	Accuracy	Précision	Recall	F1-Score
Long Short-Term Memory (LSTM)	80%	0.77	0.75	0.76
Transformers	85%	0.82	0.84	0.83
Bi-LSTM	97%	0.96	0.98	0.97

Table 2: Resultat Entraînement sur le WLASL (Word-Level American Sign Language Dataset)

**Interpretation:** Le tableau ci-dessus présente les résultats des modèles testés sur le jeu de données WLASL. Le Bi-LSTM se distingue encore une fois avec une accuracy de 97%, une précision de 0.96, un recall de 0.98, et un F1-score de 0.97. Ces résultats montrent la robustesse du Bi-LSTM à travers différents jeux de données. Les Transformers montrent également de bonnes performances, meilleures que celles du LSTM, mais toujours inférieures à celles du Bi-LSTM. Comparaison entre Notre Jeu de Données et le WLASL

Jeux de Données	Accuracy (Bi-LSTM)	<b>Précision</b>	Recall	F1-Score
<b>Notre Jeu de Données</b>	92%	0.91	0.92	0.91
WLASL	97%	0.96	0.98	0.97

Table 3: Resultat Comparaison entre Notre Jeu de Données et le WLASL

**Interpretation:** Ce tableau compare les performances du modèle Bi-LSTM sur notre jeu de données et le jeu de données WLASL. Bien que les performances soient légèrement meilleures sur le jeu de données WLASL, notre modèle montre une robustesse et une efficacité impressionnantes avec des scores élevés en précision, recall, et F1-score sur les deux jeux de données. Cela confirme que notre modèle est capable de généraliser efficacement et de fournir des prédictions précises dans différents contextes.

## 7.1 Résultats

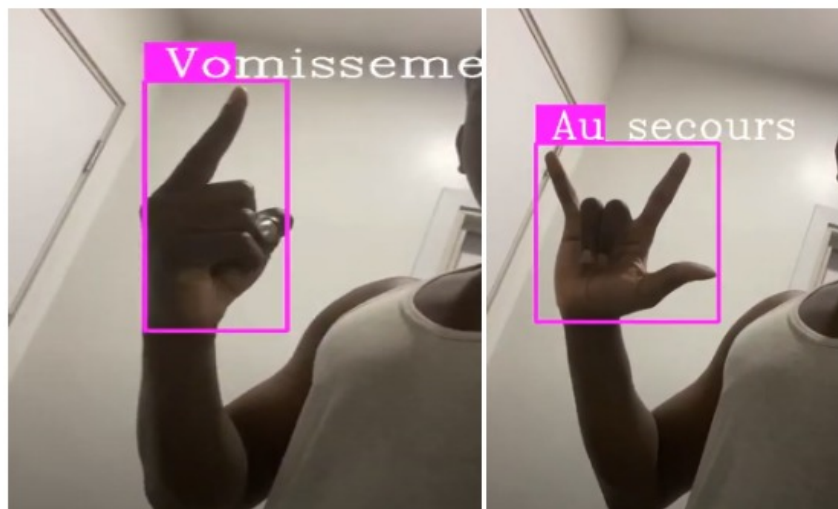


Figure 7.1: Resultat

La mise en œuvre de ce projet visant à la reconnaissance de la langue des signes africaine, particulièrement congolaise, dans les situations d'urgence hospitalière, souligne l'urgence et la pertinence de développer des solutions technologiques adaptées pour répondre aux besoins spécifiques des personnes sourdes dans ces contextes critiques.

En s'appuyant sur des avancées en matière de deep learning, ce projet vise à réduire les obstacles de communication rencontrés par les personnes sourdes lorsqu'elles requièrent des soins médicaux d'urgence, ce qui peut entraîner des retards dans les interventions médicales et compromettre la qualité des soins prodigués.

L'utilisation de techniques telles que le modèle Bi-directional Long Short-Term Memory (Bi-LSTM) a démontré des performances prometteuses dans la reconnaissance des gestes de la langue des signes congolaise. Ces avancées technologiques ouvrent la voie à une communication plus fluide et efficace entre le personnel médical et les patients sourds, améliorant ainsi les résultats médicaux et contribuant potentiellement à sauver des vies.

En définitive, ce projet représente une étape essentielle vers l'amélioration de l'accessibilité et de la qualité des soins de santé pour les personnes sourdes en République Démocratique du Congo et dans toute l'Afrique. Il souligne l'importance de l'innovation technologique pour promouvoir une société plus inclusive et équitable, où chaque individu, indépendamment de ses capacités, peut bénéficier des mêmes normes de soins et de services médicaux d'urgence.

## Bibliographie

- [Abd+15] Abdo, M et al. (2015). “Arabic alphabet and numbers sign language recognition”. In: *International Journal of Advanced Computer Science and Applications* 6.11, pp. 209–214.
- [Dre+07] Dreuw, Philippe et al. (2007). “Speech recognition techniques for a sign language recognition system”. In: *hand* 60, p. 80.
- [GV16] Garcia, Brandon and Sigberto Alarcon Viesca (2016). “Real-time American sign language recognition with convolutional neural networks”. In: *Convolutional Neural Networks for Visual Recognition* 2.225-232, p. 8.
- [Jai+21] Jain, Vanita et al. (2021). “American sign language recognition using support vector machine and convolutional neural network”. In: *International Journal of Information Technology* 13, pp. 1193–1200.
- [SWP98] Starner, Thad, Joshua Weaver, and Alex Pentland (1998). “Real-time american sign language recognition using desk and wearable computer based video”. In: *IEEE Transactions on pattern analysis and machine intelligence* 20.12, pp. 1371–1375.
- [Ver] Verschaeren, Roel (n.d.). “Automatische herkenning van gebaren met de”. In: ().