Original

Dimensionality Reduction for Offline Alphabet Arabic Sign Language Recognition using Deep Learning

Sari Awwad (PhD)[1]*, Subhieh M. El-Salhi (PhD)[2], Bashar Igried (PhD)[1]

[1]Department of Computer Science and Applications, Faculty of Prince Al-Hussein Bin Abdullah II for Information Technology, The Hashemite University, Zarqa, Jordan
[2]Department of Computer Information Systems, Faculty of Prince Al-Hussein Bin Abdullah II for Information Technology, The Hashemite University, Zarqa, Jordan

*Corresponding author: Sari Awwad
Department of Computer Science and Applications, Faculty of Prince Al-Hussein Bin Abdullah II for Information Technology, The Hashemite University, Zarqa, Jordan
E-mail: sari@hu.edu.jo

Received: 18 March 2025
Accepted: 19 August 2025

ABSTRACT

Background: Arabic Sign Language (ArSL) recognition remains limited in terms of technological development, compared to American Sign Language (ASL). This disparity restricts communication accessibility for individuals with hearing impairments in Arabic-speaking regions, in offline environments with limited computational resources.

Objective: This study aimed to develop a robust offline recognition system for ArSL by integrating Principal Component Analysis (PCA) for dimensionality reduction, Scale-Invariant Feature Transform (SIFT) for feature extraction, and Convolutional Neural Networks (CNNs) for gesture classification.

Material and Methods: This experimental, quantitative research used a curated dataset of ArSL gestures, obtained from Kaggle. Preprocessing involved normalization, contrast enhancement, and noise reduction. SIFT was used to extract invariant features, while PCA reduced computational complexity. CNN architectures were trained to recognize gestures, assessed using accuracy, precision, recall, F1-score, loss, confusion matrix, and Receiver Operating Characteristic (ROC) curve.

Results: The system achieved an accuracy of 86.64%, surpassing conventional models, such as SIFT combined with Support Vector Machines (SIFT+SVM) (84.45%). The integration of PCA and SIFT enhanced recognition efficiency and reduced model complexity. Deep learning methods showed superior adaptability and precision across gesture types.

Conclusion: This study presents a robust offline ArSL recognition system that enhances communication, education, and social participation for individuals with hearing impairments in Arabic-speaking regions.

Keywords

Sign Language; Deep Learning; Image Processing; Computer-Assisted; Gesture; Principal Component Analysis

Introduction

The Arabic Sign Language (ArSL) is heavily employed for communication among deaf persons in Arabic-speaking areas. ArSL recognition faces significant challenges, mainly due to the considerable variation and complexity found in sign languages [1]. This study aimed to develop a robust offline recognition system, considering the linguistic and cultural variations in Arabic-speaking communities.

The previous research mostly concentrated on American Sign

Language (ASL), often for real-time applications [2], while studies addressing ArSL especially in offline settings, remain scarce [1,3]. Offline ArSL recognition is an essential facet of accessibility, flexibility, and robustness, especially in an environment of intermittent internet connectivity. The current work stresses the need to create an ArSL recognition tool that can work offline, enabling individuals with hearing impairments to communicate comfortably in varied situations without any dependency on a networked resource. Unlike some other gesture-based communication languages, such as ASL or British Sign Language (BSL), ArSL has received little attention in the development of computer-based recognition systems. Principal Component Analysis (PCA) is taken into account for the dimensionality-reduction step, while Scale-Invariant Feature Transform (SIFT) is selected to extract stable image features that are robust against variations in illumination, scale, and rotation [4]. Hence, the system decreases the computation while increasing the reliability of its recognition. Furthermore, this study departs from the conventional Support Vector Machines (SVM) discriminative mechanisms for deep mechanisms, especially Convolutional Neural Networks (CNNs), because of their stronger ability to learn high-level complex patterns from gesture data [5].

In the present study, the principal aim is to investigate whether combining PCA, SIFT, and deep learning models can help increase the accuracy and efficiency of offline ArSL recognition systems. More specific objectives include (1) improving the accuracy of deep learning algorithms for ArSL gesture recognition, (2) preprocessing ArSL images and extracting significant features, and (3) assessing the accuracy of the developed system through standard performance measures.

The significance of this study extends beyond technical innovation. By addressing the lack of effective offline ArSL recognition tools, the research promotes communication accessibility and social inclusion for individuals with hearing impairments. Accurate ArSL gesture recognition empowers users to engage in educational settings, professional environments, and daily social interactions. Additionally, it enhances educational equity by enabling students to receive content in a comprehensible format through translated ArSL gestures [6].

This research contributes both technically and socially by presenting a hybrid PCA–SIFT–CNN model designed for offline ArSL recognition. It responds to the needs of Arabic-speaking deaf communities and lays a foundation for inclusive, context-aware assistive technologies.

ArSL recognition, however, is gaining more attention to help improve the communication between hearing-impaired persons [7]. It remains lagging due to the linguistic specificities, prioritization of high-resource languages, and lack of investment in research [8]. Hence, this research fills all those gaps by interfacing deep learning and dimensionality reduction to establish an offline ArSL recognition system for Arabic-speaking communities [9].

Resource limitations hinder robust recognition systems. ArSL research faces constraints: lack of large annotated datasets [10], limited hardware access impeding deep learning [11], shortage of experts combining ArSL linguistics and Machine Learning (ML) [12], and scarce funding [12].

CNNs have transformed computer vision tasks like classification. Their importance in sign recognition stems from effectively representing spatial and temporal features [13].

While ASL research has advanced, ArSL studies remain limited. Few applied dimensionality reduction methods like PCA, which optimizes feature representation [14]. This study addresses that gap by combining PCA with SIFT.

CNNs extract spatial features, while Recurrent Neural Networks (RNNs) address temporal dependencies. The DeepArSLR framework achieved promising performance [15].

Offline ArSL systems serve users in resource-constrained settings [16]; pretrained model adaptation improves performance. Multi-modal systems integrating visual and kinesthetic data enhance gesture recognition]; text-to-ArSL translation facilitates bidirectional communication [16].

Review studies highlight deep learning and multimodal approaches, yet few focus on ArSL [17]. Hybrid models are emerging, but none have integrated SIFT with PCA for ArSL character recognition. This research proposes a novel, resource-efficient framework.

Material and Methods

Experimental Quantitative Research

This study adopts an experimental and quantitative design rooted in applied computer vision research. A curated dataset of ArSL gestures was sourced as the foundation for system development. It includes diverse samples from multiple signers across various settings, with annotated labels enabling supervised learning tasks [18].

Preprocessing of the Images

The preprocessing techniques aimed to optimize the input data for gesture recognition by enhancing image clarity, ensuring visual consistency, and reducing noise across all samples. These preprocessing techniques included normalization, noise reduction using median or Gaussian filters, and contrast enhancement to ensure consistent feature quality across images [19, 20]. Figure 1 shows the preprocessing steps, including raw image input, denoising, contrast enhancement, and normalization.

Feature Extraction

This study employs SIFT for robust representation of ArSL gestures and uses CNN-based architectures to extract hierarchical features from preprocessed images [21]. Models, such as MobileNet, Visual Geometry Group (VGG), and ResNet were explored, each leveraging multiple convolutional layers to abstract spatial and temporal gesture patterns.

Dimensionality Reduction

To improve efficiency and mitigate overfitting, PCA was applied for dimensionality reduction while preserving essential gesture features.

Classification

CNNs were used for multi-class classification due to their proven performance in image tasks. Training involved optimizing model

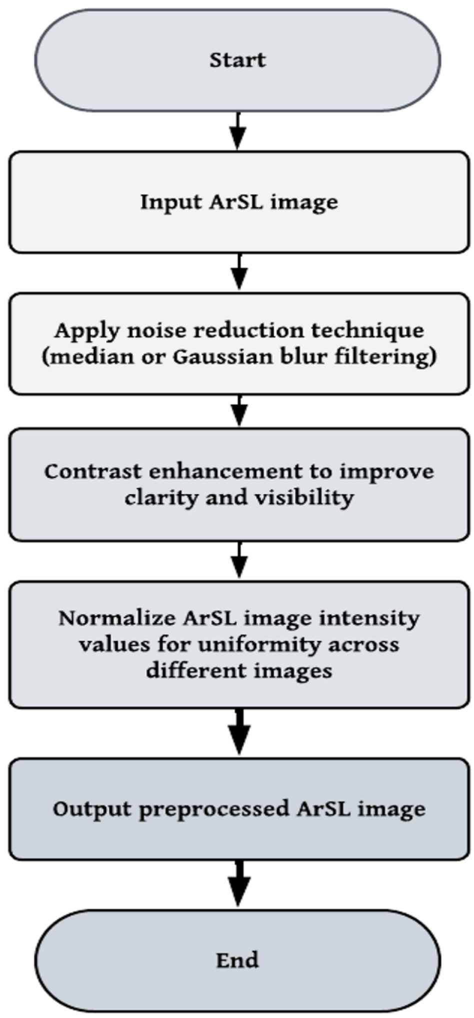

Figure 1: Arabic Sign Language (ArSL) Image Preprocessing

parameters to accurately distinguish among ArSL gesture classes.

Results

1. Experiments

1.1 Dataset

The ArSL dataset, obtained from Kaggle, contains diverse gestures, including alphabets, numbers, and expressions captured by various signers under different conditions [22]. Each gesture is manually labeled for supervised learning. Preprocessing was applied to standardize lighting, background, and image dimensions. A sample is shown in Figure 2.

1.2 Comparison with Existing Approaches

The proposed system surpasses existing models, including Bidirectional Long Short-Term Memoryand (BI-LSTM), Bag of Features with Bag of Poses (BoF+Bop), SIFT combined with Support Vector Machines (SIFT+SVM), and 3D-CNN, by integrating PCA and SIFT, which improve recognition accuracy and reduce input complexity. PCA compresses features while retaining key information, and SIFT extracts distinctive gesture traits, enhancing robustness [23]. Although SVMs provide a standard benchmark, deep learning approaches yield superior results.

1.3 Deep Learning Methodologies Consideration

CNNs outperform traditional classifiers like SVM by automatically learning hierarchical features from raw ArSL data [23]. Their adaptability to gesture variations boosts both classification accuracy and scalability.

1.4 Performance Metrics Analysis

To comprehensively assess the recognition system, standard evaluation metrics were used to gauge accuracy, discriminatory power, and reliability [23]. Accuracy indicates the proportion of correctly classified ArSL gestures among all samples.

Precision measures the rate of true positives among predicted positives, highlighting the system's ability to reduce false alarms. Sensitivity, or recall, measures how many of the actual positive cases are correctly identified by the model, showing how well the system detects the right gestures [23]. F1-Score, the harmonic mean of precision and recall, is especially useful in addressing class imbalance.

1.5 Additional Evaluation Metrics

Loss quantifies the gap between predicted and actual values during training. Cross-entropy loss was employed to penalize

Figure 2: Arabic Sign Language Dataset

misclassifications and guide model optimization. A downward trend across epochs signals improved learning and reduced error [24]. A confusion matrix provides a tabular comparison of predicted versus actual classes, where diagonal entries represent correct predictions and off-diagonal entries indicate misclassifications. It helps identify systematic errors for model refinement [24]. Receiver Operating Characteristic (ROC) curve analysis assesses classifier performance across different thresholds by plotting true positive rates against false positive rates. A higher Area Under the Curve (AUC) signifies better class separability and supports threshold adjustment and model tuning.

2. Experimental Results

In ArSL gestures, the system's effectiveness was evaluated using accuracy, loss, precision, recall, F1-score, confusion matrix analysis, and ROC curve metrics. Key performance results are presented in the following subsections.

2.1 Accuracy Rates

To validate the model's ability to classify previously unseen ArSL gestures, high accuracy rates were achieved on the test dataset. As illustrated in Figure 3, the accuracy graph compares training and test performance, demonstrating the system's generalizability across datasets. Table 1 compares the proposed method with existing approaches.

The comparison results show that the proposed method has an efficient accuracy over 5 epochs.

2.2 Loss Analysis

The validation and training convergence over successive epochs are illustrated by the loss graph (Figure 4). For optimizing performance and controlling prediction errors, the ability of the model is indicated by the decreasing trend in loss.

2.3 Confusion Matrix

As shown in Figure 5, the confusion matrix offers detailed insights into the model's classification performance across ArSL gesture classes, identifying true positives, false positives, false negatives, and true negatives. It is instrumental in detecting frequent misclassifications and guiding system improvements.

2.4 ROC Curve Analysis

The ROC curve plots the true positive rate

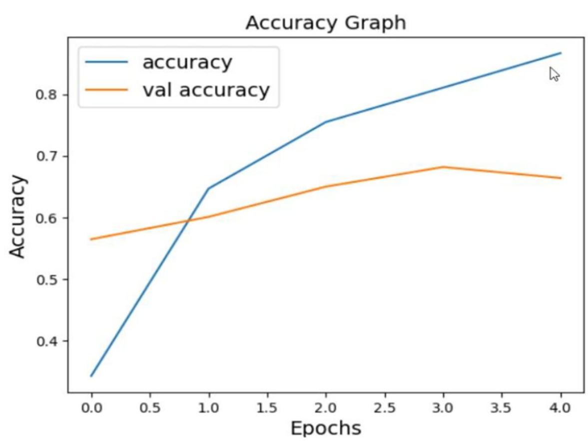

Figure 3: Accuracy Graph

Table 1: Accuracy comparison between the proposed method and existing approaches

No.	Method	Reference	Accuracy (%)
1	Bidirectional LSTM (BI-LSTM)	[15]	85.56
2	Bag of Features + Bag of Poses (BoF+Bop)	[25]	66.95
3	3D Convolutional Neural Network (3D-CNN)	[26]	34.90
4	SIFT + Support Vector Machine (SIFT+SVM)	This study	84.45
5	Proposed PCA–SIFT–CNN Hybrid Method	This study	86.64

LSTM: Long Short-Term Memoryand, SIFT: Scale-Invariant Feature Transform, PCA: Principal Component Analysis against the false positive rate, illustrating the trade-off between sensitivity and specificity. A larger AUC indicates stronger discriminatory power in recognizing ArSL gestures.

These results confirm the system's robustness and applicability in enhancing communication for individuals with hearing impairments in Arabic-speaking contexts. Metric-based optimization further supports its real-world deployment.

Discussion

This study proposes a hybrid framework for ArSL recognition that integrates SIFT, PCA, and CNNs for efficient offline usage. The system is designed for low-resource environments and achieved 86.64% accuracy, outperforming other models, such as SIFT+SVM (84.45%), BoF+Bop (66.95%) [25], and 3D-CNN (34.9%) [26].

A high F1-score and AUC (>0.92) validate the model's strong classification performance. PCA reduced feature complexity, SIFT ensured invariance, and CNNs extracted deep patterns for better generalization [26].

Unlike end-to-end models like BI-LSTM and 3D-CNN [26] that require extensive data and powerful hardware, the proposed hybrid model delivers strong results under moderate resource constraints—aligning with trends favoring interpretable, efficient systems. Latif et al. [27], MobileNetV2 enables real-time, offline recognition, ideal for regions with limited internet access. Future work may build by integrating skeletal or sEMG data, as their study demonstrated that combining visual inputs with muscle activity signals improves gesture recognition accuracy and robustness.

By developing an offline ArSL system, this study aligns with the recommendations] to create culturally inclusive, deployable solutions that improve access for Arabic-speaking deaf communities. Standard metrics (accuracy, precision, recall, F1-score, loss, and ROC) and statistical tests (t-test, ANOVA) confirmed significant improvements (*P*-value<0.05) and strong model reliability [28].

Limitations include reliance on isolated gestures, lack of signer-independent testing, and exclusion of multimodal inputs. While PCA enhances efficiency, it may discard subtle, low-variance features. Overall, the PCA–SIFT–CNN framework offers a robust, offline-compatible ArSL solution, advancing socially responsive assistive technologies.

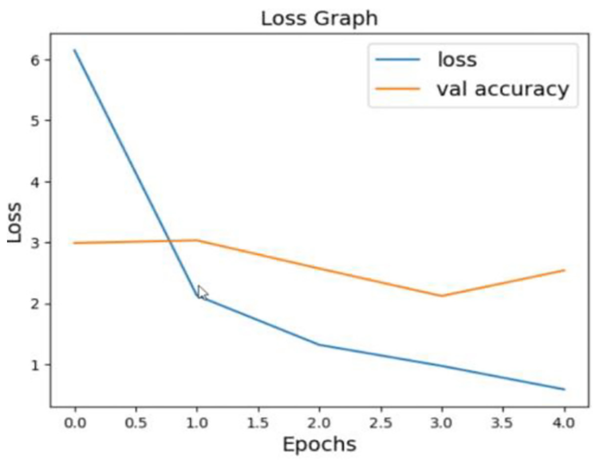

Figure 4: Loss Analysis Graph

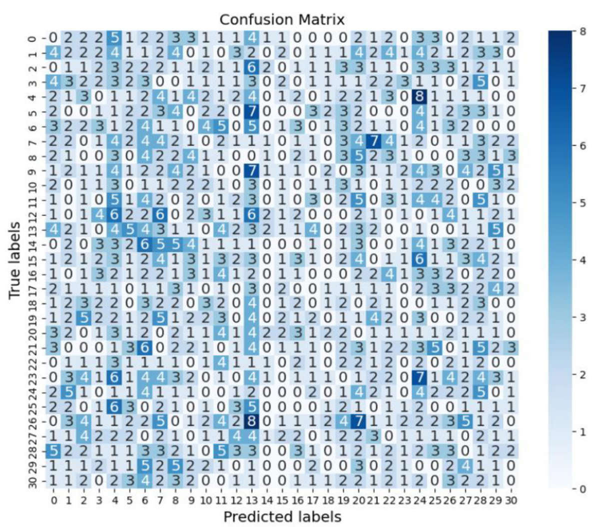

Figure 5: Confusion Matrix

Conclusion

This paper presents a novel approach for ArSL recognition by integrating PCA-based dimensionality reduction, SIFT feature extraction, and deep learning. Experimental results demonstrate that the proposed hybrid framework outperforms standalone techniques in terms of accuracy and efficiency. These findings confirm the practicality and effectiveness of combining handcrafted features with deep models for ArSL recognition. Future work will focus on expanding the gesture range and dataset diversity to further enhance system robustness and generalizability.

To advance the proposed system, future work should focus on expanding the dataset to include a broader range of ArSL gestures from diverse regions and dialects, as well as exploring multilingual sign language support. Enhancing the model for real-time performance and optimizing it for mobile or low-powered devices will improve its practical deployment. Incorporating temporal learning models, such as LSTM networks, could enable recognition of sequential and sentence-level gestures. Finally, usability should be prioritized through intuitive interface design and active involvement of the deaf community in testing, ensuring the system remains accessible, relevant, and effective in real-world contexts.

Acknowledgment

The authors would like to thank the Department of Computer Science and Applications at The Hashemite University for providing technical support and access to computing resources essential to this research.

Authors' Contribution

S. Awwad conceptualized the research framework and supervised the experimental design. SM. El-Salhi contributed to methodology development and dataset preprocessing. B. Igried led the model implementation and evaluation. All authors contributed to writing, reviewing, and approving the final manuscript.

Ethical Approval

This study did not involve human participants, animals, or sensitive personal data and therefore did not require ethical approval.

Funding

The authors received no financial support from any funding agency or institution for the research, authorship, and/or publication of this article.

Conflict of Interest

None

References

1. Abdullah A, Abdul-Kadir NA, Harun FC. Research and development of IMU sensors-based approach for sign language gesture recognition. *J Telecommun Electron Comput Eng.* 2017;**9**(3-9):33-9.
2. Wadhawan A, Kumar P. Sign language recognition systems: A decade systematic literature review. *Arch Comput Methods Eng.* 2021;**28**:785-813. doi: 10.1007/s11831-019-09384-2.
3. Alghazo JM, Latif G, Elhassan A, Alzubaidi L, Al-Hmouz A, Al-Hmouz R. An online numeral recognition system using improved structural features–a unified method for handwritten Arabic and Persian numerals. *J Telecommun Electron Comput Eng.* 2017;**9**(2-10):33-40.
4. Tharwat A, Gaber T, Hassanien AE, Shahin MK, Refaat B. Sift-based arabic sign language recognition system. In: Afro-European Conference for Industrial Advancement: Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014; Springer; 2015. p. 359-70.
5. Elleuch M, Maalej R, Kherallah M. A new design based-SVM of the CNN classifier architecture with dropout for offline Arabic handwritten recognition. *Procedia Comput Sci.* 2016;**80**:1712-23. doi: 10.1016/j.procs.2016.05.512.
6. Fayyaz S, Ayaz Y. CNN and traditional classifiers performance for sign language recognition. In: Proceedings of the 3rd international conference on machine learning and soft computing; New York, NY, USA: Association for Computing Machinery; 2019. p. 192-6.
7. Ghanim TM, Khalil MI, Abbas HM. Comparative study on deep convolution neural networks DCNN-based offline Arabic handwriting recognition. *IEEE Access.* 2020;**8**:95465-82. doi: 10.1109/ACCESS.2020.2994290.
8. Kumar P, Gauba H, Roy PP, Dogra DP. A multi-

modal framework for sensor based sign language recognition. *Neurocomputing.* 2017;**259**:21-38. doi: 10.1016/j.neucom.2016.08.132.

9. Hanecioğlu N. Gesture recognition for Turkish Sign Language with CNN and RNN [Master's thesis]. Ankara: Ankara Yildirim Beyazit University; 2019. Available from: file:///C:/Users/ORGINAL/Downloads/955cd1eb-935d-498c-b216-6a46d-6e764c5.pdf.
10. Ahmed R, Dashtipour K, Gogate M, Raza A, Zhang R, Huang K, et al. Offline arabic handwriting recognition using deep machine learning: A review of recent advances. In International conference on brain inspired cognitive systems; Guangzhou, China: Springer; 2019. p. 457-68.
11. Zakariah M, Alotaibi YA, Koundal D, Guo Y, Mamun Elahi M. Sign language recognition for Arabic alphabets using transfer learning technique. *Comput Intell Neurosci*. 2022;**2022**(1):4567989. doi: 10.1155/2022/4567989.
12. Rowel MMK, Gunasekara ADAI, Uwanthika GAI, Wijesinghe DB. E-Learning Platform for Hearing Impaired Children with Handwritten Character Recognition using CNN. *IJRC*. 2022;**1**(1):37-43. doi: 10.64701/ijrc/345/5297.
13. Pisharady PK, Saerbeck M. Recent methods and databases in vision-based hand gesture recognition: A review. *Comput Vis Image Underst.* 2015;**141**:152-65. doi: 10.1016/j.cviu.2015.08.004.
14. Nikam AS, Ambekar AG. Sign language recognition using image based hand gesture recognition techniques. In 2016 online international conference on green engineering and technologies (IC-GET); Coimbatore, India: IEEE; 2016. p. 1-5.
15. Aly S, Aly W. DeepArSLR: A novel signer-independent deep learning framework for isolated arabic sign language gestures recognition. *IEEE Access*. 2020;**8**:83199-212. doi: 10.1109/ACCESS.2020.2990699.
16. Ibrahim NB, Selim MM, Zayed HH. An automatic Arabic sign language recognition system (ArSLRS). *J King Saud Univ Comput Inf Sci.* 2018;**30**(4):470-7. doi: 10.1016/j.jksuci.2017.09.007.
17. Cheok MJ, Omar Z, Jaward MH. A review of hand gesture and sign language recognition techniques. *Int J Mach Learn Cybern*. 2019;**10**(1):131-53. doi: 10.1007/s13042-017-0705-5.
18. Anderson R, Wiryana F, Ariesta MC, Kusuma GP. Sign language recognition application systems for deaf-mute people: A review based on input-process-output. *Procedia Comput Sci.* 2017;**116**:441-8. doi: 10.1016/j.procs.2017.10.028.
19. Goldin-Meadow S, Brentari D. Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behav Brain Sci.* 2017;**40**:e46. doi: 10.1017/S0140525X15001247. PubMed PMID: 26434499. PubMed PMCID: PMC4821822.
20. Wang F, Li C, Liu CW, Zeng Z, Xu K, Wu JX. An approach based on 1D fully convolutional network for continuous sign language recognition and labeling. *Neural Comput Appl*. 2022;**34**(20):17921-35. doi: 10.1007/s00521-022-07415-x.
21. Oyedotun OK, Khashman A. Deep learning in vision-based static hand gesture recognition. *Neural Comput Appl*. 2017;**28**(12):3941-51. doi: 10.1007/s00521-016-2294-8.
22. Alayed A. Machine Learning and Deep Learning Approaches for Arabic Sign Language Recognition: A Decade Systematic Literature Review. *Sensors (Basel)*. 2024;**24**(23):7798. doi: 10.3390/s24237798. PubMed PMID: 39686335. PubMed PMCID: PMC11644898.
23. Hu Y, Wong Y, Wei W, Du Y, Kankanhalli M, Geng W. A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition. *PLoS One.* 2018;**13**(10):e0206049. doi: 10.1371/journal.pone.0206049. PubMed PMID: 30376567. PubMed PMCID: PMC6207326.
24. Wu D, Pigou L, Kindermans PJ, Le ND, Shao L, Dambre J, Odobez JM. Deep Dynamic Neural Networks for Multimodal Gesture Segmentation and Recognition. *IEEE Trans Pattern Anal Mach Intell.* 2016;**38**(8):1583-97. doi: 10.1109/TPAMI.2016.2537340. PubMed PMID: 26955020.
25. Joksimoski B, Zdravevski E, Lameski P, Pires IM, Melero FJ, Martinez TP, et al. Technological solutions for sign language recognition: a scoping review of research trends, challenges, and opportunities. *IEEE Access.* 2022;**10**:40979-98. doi: 10.1109/ACCESS.2022.3161440.
26. Al-Hammadi M, Muhammad G, Abdul W, Alsulaiman M, Hossain MS. Hand gesture recognition using 3D-CNN model. *IEEE Consumer Electronics Magazine.* 2019;**9**(1):95-101. doi: 10.1109/MCE.2019.2941464.
27. Latif G, Alghazo J, Mohammad N, Alghazo R. Communicating with the deaf and hard of hearing through automatic Arabic sign language translator. *In J Phys Conf Ser*. 2021;**1962**(1):012055. doi: 10.1088/1742-6596/1962/1/012055.
28. Elshaer AM, Ambioh Y, Soliman Z, Ahmed O, Elnakib M, Safwat M, et al. Enhancing arabic alphabet sign language recognition with vgg16 deep learning investigation. In 14th International Conference on Electrical Engineering (ICEENG); Cairo, Egypt: IEEE; 2024. p. 184-6.