ISSN: 2312-8135 | Print ISSN: 1992-0652

info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

A Hybrid Approach to Arabic Sign Language Recognition by HSV Space and Deep Learning Classification

Noor Fadel Hussain[1*], Hiba Al-Khafaji[2]
[1]Information Security, College of Information Technology, University of Babylon, noor.fadel@uobabylon.edu.iq, Babil, Iraq.
[2]Software, College of Information Technology, University of Babylon, hibamj.alkhafaji@uobabylon.edu.iq , Babil, Iraq.
*Corresponding author email: noor.fadel@uobabylon.edu.iq ; mobile: 07501921586

نهج هجين للتعرف على لغة الإشارة العربية باستخدام مساحة الالوان وتصنيف التعلم العميق(HSV)

Accepted: **27/10/2025** **Published:** **31/12/2025**

ABSTRACT

Understanding and automatically recognizing sign language is crucial for equal opportunities for deaf and hard-of-hearing individuals in society. In the world of Arabic sign language communication, people share their ideas, thoughts, and feelings and engage in meaningful daily interactions. While artificial intelligence and computer vision technologies have improved, studies focused on Arabic sign language processing have come first and continue to be a research frontier compared to other world languages. This research proposed a methodology to leverage the HSV color space model for initial segmentation to eliminate background noise to refine the input data. This process segments the hand or semantic sign area. Then, the data is passed onto a Convolutional Neural Network (CNN) for precise sign classification. This method's novelty is the coordinated interplay of differential data simplification through color space manipulation and the profound representational power of the CNN for improved overall system performance for hood sign recognition.

The research also seeks to fill an existing gap in the field of Arabic signal processing by presenting an integrated model, opening up broad horizons for applications in the fields of education, healthcare, and smart government services. The proposed method showed promising results, as the results of the used criteria ranged between Accuracy 95.3%, Precision 93.5%, Recall 92.8%, and F1-score 93.1%.

Background:

Recognizing sign languages has always been a vital topic in societies, especially with the recent massive expansion of the deaf and mute community. It serves as the main form of communication for over 70 million deaf and hard-of-hearing individuals globally. Thus, the development of simple, accurate, and effective methods that utilize readily available resources deserves attention. Among the various forms of sign language, sign language and Arabic are especially significant, as Arabic is predominantly -spoken in the Middle East and North Africa. Yet, the challenges that arise from the communication gap between this population and the hearing members of society continue to affect various areas of the deaf and hard-of-hearing individuals educational, social, and professional for educational, social, and, professional integration. Hence, creating automated systems that can comprehend Arabic Sign Language is an essential development in helping socially integrate and bridge communication gaps [1].

Materials and Methods:

The proposed hand gesture recognition system relies on a multi-stage methodology aimed at high-resolution image processing and efficient feature extraction. The methodology begins with the data collection phase, where a set of images and videos representing letters and words in Arabic Sign Language is collected on the Kaggle Dataset.

Results:

The proposed Arabic Sign Language recognition system was developed through two principal stages: (1) segmentation using the HSV color space with experimentally defined thresholds, and (2) classification via a Convolutional Neural Network (CNN). The segmentation stage, relying on the HSV model with the defined lower range (H=0,S=30,V=60) and upper range (H=20,S=150,V=255), demonstrated effectiveness in isolating the hand region from complex backgrounds, contributed significantly to noise reduction and the removal of irrelevant small objects, and enhanced the clarity of the extracted hand region prior to classification.

Conclusion:

Recognizing sign languages has always been a vital topic in societies, especially with the recent massive expansion of the deaf and mute community. Therefore, it is important to focus on easy and highly accurate techniques that can be relied upon with relatively available resources. The proposed methodology is highly flexible and scalable, as the HSV color space boundaries can be modified or the CNN architecture optimized to suit research requirements and the characteristics of different datasets.

Key words:

Sign language, Arabic Sign language, HSV, convolutional neural network.

INTRODUCTION

Recognizing sign languages has always been a vital topic in societies, especially with the recent massive expansion of the deaf and mute community. It serves as the main form of communication for over 70 million deaf and hard-of-hearing individuals globally. Thus, the development of simple, accurate, and effective methods that utilize readily available resources deserves attention. Among the various forms of sign language, sign language and Arabic are especially significant, as Arabic is predominantly spoken in the Middle East and North Africa. Yet, the challenges that arise from the communication gap between this population and the hearing members of society continue to affect various areas of the deaf and hard-of-hearing individuals educational, social, and professional for educational, social, and, professional integration. Hence, creating automated systems that can comprehend Arabic Sign Language is an essential development in helping socially integrate and bridge communication gaps [1].

Over the past few decades, the advancements in AI and computer vision have significantly advanced the creation of systems that can recognize and interpret signs and gestures. However, the predominant focus is still on globally spoken languages, like English and American Standard Sign Language, while Arabic is still in the nascent stages of development. This illustrates the need for innovative models that can address the nuances of the Arabic language and its diverse dialects. [2].

Difficulties in recognizing Arabic Sign Language. Language recognition systems encounter numerous obstacles, as follows:

• Underdeveloped benchmark databases for Arabic signs focused recognition. The vast differences in signed languages across the Arab world.

• Problematics of isolating the arm or hand from the scene, more so in uncontrolled scenarios.

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية
ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

• Inability to design accurate classification models that build on generalization with small-sized data sets.

Given the outlined challenges, foremost problem-solving approaches must concentrate on the segmentation problem, which serves to separate the hand from the backdrop, thus lessening the effort classified deep-learning networks have to exert for classification. [3].

The HSV (Hue, Saturation, Value) color model ranks among the top color models used in computer vision image segmentation applications, especially in human hand recognition and scene understanding. This model is more useful than the traditional RGB color model since human color perception is more aligned with the HSV model, as hue (color), saturation (color purity), and value (brightness) are separated. This color model aids in image segmentation tasks as it helps identify various elements, even in complex scenarios with multiple background colors, and low light levels. Skin tone detection as a segmentation technique is less affected by low light conditions due to the robust design of the HSV model. Converting images to the HSV space enhances the quality of the extracted features, thereby improving the accuracy of recognition and classification models. The stability of image segmentation algorithms is improved as the model reduces the effect of shadows and enhances light segmentation. Thus, there are no HSV color space image segmentation tasks in the literature.

Employing certain limits in the HSV ranges allows for the more effective separation of the human hand or the signaling region from its background. At this point, the data is simplified, and visual clutter is minimized to form a clearer and more accurate representation of the information that the neural network needs to analyze. This benefits the system's overall performance and eliminates the recognition errors [5].

Combining image simplification via HSV with the representation learning capabilities of CNNs leads to the construction of an integrated system capable of handling the diversity and fine-grained details of Arabic signs, enhancing accuracy and reliability.

Although previous research attempts have existed, most have focused on ideal environments or small datasets, without systematically expanding on Arabic sign language [6]. Furthermore, integration between preprocessing steps such as segmentation and deep learning is absent in much of previous work. Hence, the value of this research approach [7]. Which combines two complementary approaches: HSV for segmentation and CNN for classification, becomes apparent.

Machine learning techniques have been used in many biological applications and in several fields such as attack detection [8], various image processing [9] and computer vision[10][11]. In our current research, some techniques were used as a comparison with the proposed method.

The research is structured as follows: The second section contains related works and a comparison table. The third section contains the proposed method. The discussion and results are concentrated in the fourth section. Finally, the research ends with a conclusion.

.

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

RELATED WORK

[12] This study focused on building a hybrid system based on two sensors: a Kinect V2 to capture body and arm motion, and a Leap Motion to capture fine finger movements with high accuracy. The research involved classifying signals using classical machine learning methods, specifically Support Vector Machines (SVMs). Accurately distinguishing between signals that differ only minimally during finger movements was facilitated by sensor feature fusion. While this exemplifies the value of using multiple modalities for sign language recognition, it introduces costs and complexities, as evidenced by having to two separate devices.

[13] For this research, the Google MediaPipe framework was used for hand feature detection and a combination of a CNN for spatial feature extraction and an LSTM for temporal feature extraction. No specific accuracy was stated. However, the research is problematic because it has not only a complex data processing requirement, but also, likely, sensitivity to signal variation.

[14] This research developed a real-time system for recognizing Arabic alphabetic characters. It only used pre-trained CNNs (e.g. VGGNet) and transfer learning for sign recognition, from still images captured via a webcam. Its implementation for isolated letter recognition is simple and performs quite well, but the system is limited in that it does not process words or sentences.

[15] This study addressed the toughest issue of all: understanding complete sentences in sign language instead of just individual signs. They utilized a Kinect sensor, which provided 3D skeletal tracking data of the joints in the body. Then, they implemented recurrent neural networks, specifically long short-term memory networks, which work with time series data and are able to learn and retain information over long periods, to translate gestures into sign language sentences in real time.

[16] This study conducted research in the area of computer vision and used an ordinary RGB camera to capture videos. They manually extracted features from the frames first and then used convolutional neural networks to learn features directly from images. The authors used a fuzzy logic-based decision-making system to increase system accuracy in ambiguous conditions and to increase system robustness in the presence of inter-user variability.

Table 1: Comparison of related work

Study	Techniques used	Accuracy	Limitations	Main Benefits and Contributions
[12]	Kinect V2, Leap Motion, SVM	95% - 98%	Expensive system (two sensors), non-portable, operates in a controlled environment.	Combining two sensors to compensate for each other's weaknesses and achieve very high accuracy for isolated signals.
[13]	Google Media Pipe with CNN and LSTM	--	requires complex data processing and is sensitive to signal variation	Combining CNN and LSTM to improve the accuracy of Arabic gesture recognition
[14]	Webcam, Transfer Learning (VGGNet), CNN	96% - 98%	For static images (single frames) only, does not handle motion or sentences.	A simple, low-cost, and effective system for real-time isolated alphabetic character recognition.
[15]	Kinect (3D Skeletal Data), RNN-LSTM	85% - 90%	Lower accuracy than isolated signal systems. Complexity in training sentences, depends on the accuracy of Kinect joint tracking.	Addressing the problem of recognizing continuous sign language (sentences) using advanced LSTM models, which is a greater challenge than isolated signals.
[16]	Computer Vision (RGB Camera), CNN, Fuzzy Logic	92% - 95%	Performance depends on good lighting and background, difficulty recognizing continuous signals.	Integrating artificial intelligence techniques (deep learning and fuzzy logic) to improve robustness and generalization.
Our	HSV With CNN	95.3	Performance depends on good lighting	Integrating artificial intelligence techniques (deep learning and HSV) to improve robustness detect hand and generalization classification with clutter backgrounds.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com
مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

MATERIALS AND METHODS

For the proposed hand gesture recognition system, high-resolution image processing and feature extraction must adhere to a multi-stage process. This begins with the collection phase, which gathers images and videos depicting the letters and words communicated through ASL, available in the Kaggle Dataset. Following this, during the preprocessing stage, the images are transformed from the traditional RGB (Red, Green, Blue) color model into the HSV (Hue, Saturation, Value) color model, as it is more aligned with the human perception of colors and handles changes in illumination more efficiently.

A segmentation process is then applied to isolate the hand from the background using specific skin color boundaries, allowing only the hand region to be extracted as the primary input for subsequent stages. In the next phase, the processed images are fed into a deep neural network for classification. Two networks were used for evaluation: the first is a traditional CNN, which relies on a hierarchical sequence of convolutional and pooling layers to extract spatial features. The second is the ResNet (Residual Network), which features skip connections that enable the model to overcome the problem of vanishing gradients in deep networks. However, it requires additional resources beyond the computer **Processor** .The proposed method was developed on a computer with specifications Intel(R) Core(TM) i5-6600U CPU @ 2.60GHz. **(RAM):** 8.00 GB. Windows 10. Memory: 980M. Cores: 4. The figure1 shows the main details of the proposed method. The proposed uses a computer camera with a video quality of 720p (16:9, 30 fps) .

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

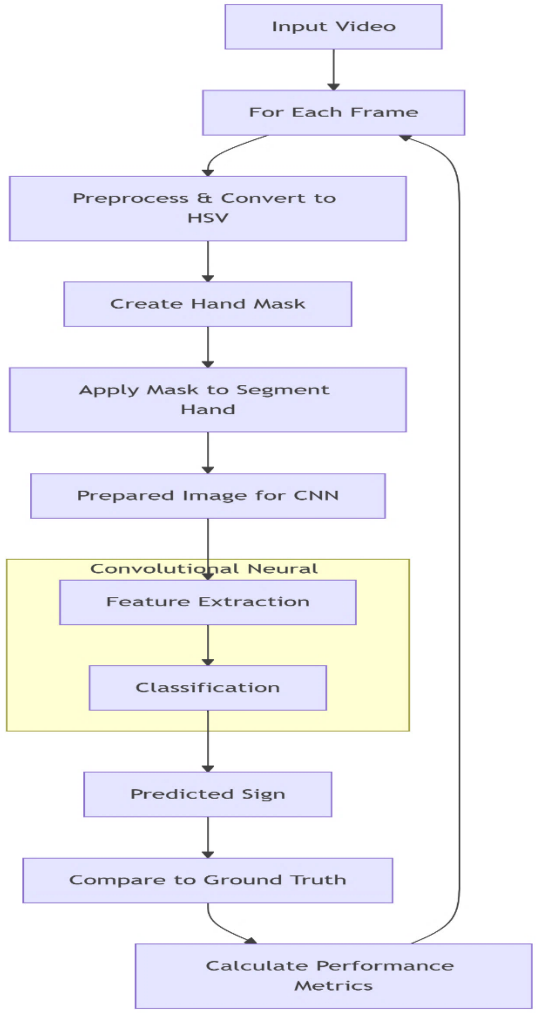


Figure 1: Arabic Sign Language Recognition System

Figure 1 above illustrates the methodology, which begins with mask creation: A specific range of hues, saturations, and values representing human skin color is defined.

lower = (H = 0, S =30, V= 60)

upper = (H =20, S =150, V =255)

A binary image (mask) is created, with white pixels within the skin color range and black pixels representing everything else.

Apply Mask: The final mask is used to isolate the hand from the background. The result is an image where the background is black and only the hand is visible, as shown in Figure 2.

The images are fed into a CNN, which extracts the spatial features and visual patterns associated with the various signs. The network relies on convolution and pooling layers to extract representative features, leading up to fully connected layers that perform classification. This stage is the focus of the research, as it enables the model to distinguish between different letters and signs based on their distinct features.

Article
JOURNAL OF UNIVERSITY OF BABYLON
For Pure and Applied Sciences (JUBPAS)
Vol. 33 ; No.4 | 2025

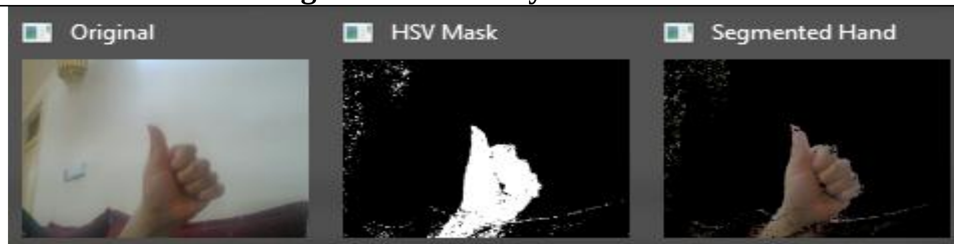
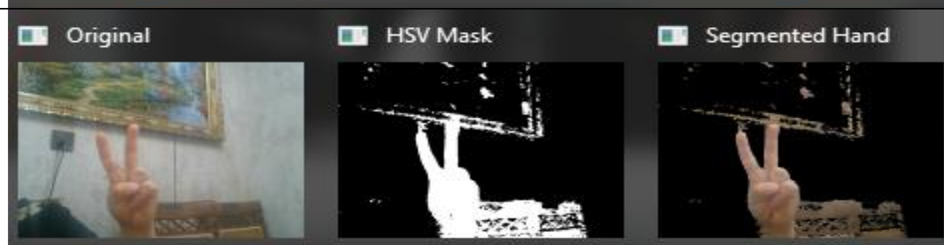
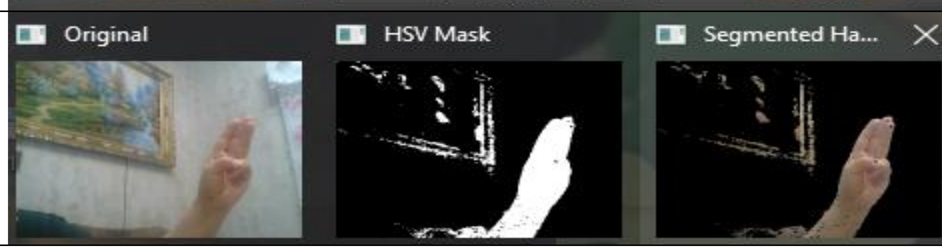
Sign	Segmented Hand by HSV Mask
Alif	
Baa	
Taa	
Thaa	

Figure 2: Segmented Hand

The approach encompasses an evaluation and testing phase in which the model is deployed to a separate dataset that wasn't considered during training to assess accuracy in recognition and the effectiveness of the model under diverse imaging conditions and system performance.

The data preparation, the segmented hand image, printed as a three-channel color image, is resized to the to the dimensions expected by the neural network model (224x224 pixels).

After completing the first step, the network architecture consisted of three convolutional layers with pooling, which was sufficient for extracting visual features edges, shapes, hand pose, and one Dense layer for deep representation. Input Layer: Receives the prepared image. A series of convolutional layers, ReLU activation functions, and pooling layers automatically detect and learn hierarchical features from the image (such as edges, shapes, and complex patterns that represent specific signs). For the first layer, we started with a 32-unit filter and multiplied it to the third layer.

Conv Layer 1: 32 filters, kernel size (3×3), activation ReLU.
MaxPooling 1: (2×2).
Conv Layer 2: 64 filters, kernel size (3×3), activation ReLU.

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية
ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

MaxPooling 2: (2×2).
Conv Layer 3: 128 filters, kernel size (3×3), activation ReLU.
MaxPooling 3: (2×2).
Flatten Layer.
Dense Layer 1: 128 units, activation ReLU, dropout 0.5 (to avoid overfitting).
Dense Layer 2 (Output Layer): 29 units number of Arabic letters, used Softmax activation. The number of epochs was 30–50. The optimizer, Adam, was used with a learning rate starting at 0.001 and settling at 0.0001. Output Layer: Contains nodes corresponding to the number of sign language classes to be recognized (29 nodes for Arabic letters). The Softmax function is used to convert the results to probabilities. The node with the highest probability represents the model's final prediction.

In this network, the filter size (kernel) was set to 5×5 in the first layer only to improve the extraction of wide edges. The kernel size was then settled on 3×3, which is an excellent standard option for processing small and medium-sized images.

The number of images from the Kaggle Arabic Sign Language dataset is less than 20,000, so it relies heavily on augmentation rotation, flipping, changing illumination, adding noise, see figure 3. The Train/Test split is 80% for training, 10% for validation, and 10% for testing.

The extracted features are flattened and passed through fully connected layers to perform the final classification task.

Output Prediction: The outcome of a CNN model is the predicted class label (e.g., "the letter Alif")

Comprehensive metrics such as precision, recall, and the F1 index are used to determine the system's performance, while potential limitations and scenarios that may affect recognition accuracy, such as complex backgrounds or inhomogeneous lighting, are analyzed. Metrics predictions out of the total predictions. Classification Accuracy: The model predicted Calculation: Standard performance metrics are calculated, such as: Accuracy: The percentage of correct a specific class, what is its correctness? Recall: The percentage of correct predictions out of the total that should have been correct. F1-Score: A harmonic mean between classification accuracy and recall, which provides a better picture when there is data skewing.

RESULTS AND DISCUSSION

The proposed Arabic Sign Language recognition system was developed through two principal stages: (1) segmentation using the HSV color space with experimentally defined thresholds, and (2) classification via a Convolutional Neural Network (CNN). The segmentation stage, using the HSV model with the defined lower range (H=0,S=30,V=60) and upper range (H=20,S=150,V=255) demonstrated effectiveness in isolating the hand region from complex backgrounds, aided significantly in the noise reduction, and removal of irrelevant small objects, contributing to the improvement in the clarity of the extracted hand region before the classification.

After segmentation, the CNN architecture received the processed, augmented, and resized hand images.

We performed augmentation on the image, and the results were as follows:

- Rotation 25° → Rotate the image by 25 degrees.
- Horizontal Flip → Flip the image horizontally.
- Rescale + Crop → Enlarge the image by 20% and then crop it.
- Gaussian Noise → Add Gaussian noise to simulate blurring.

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية
ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

- Brightness + → Increase brightness and contrast for improved visibility . Figure 3 illustrates the augmentation process we need to enhance data diversity in different environments.

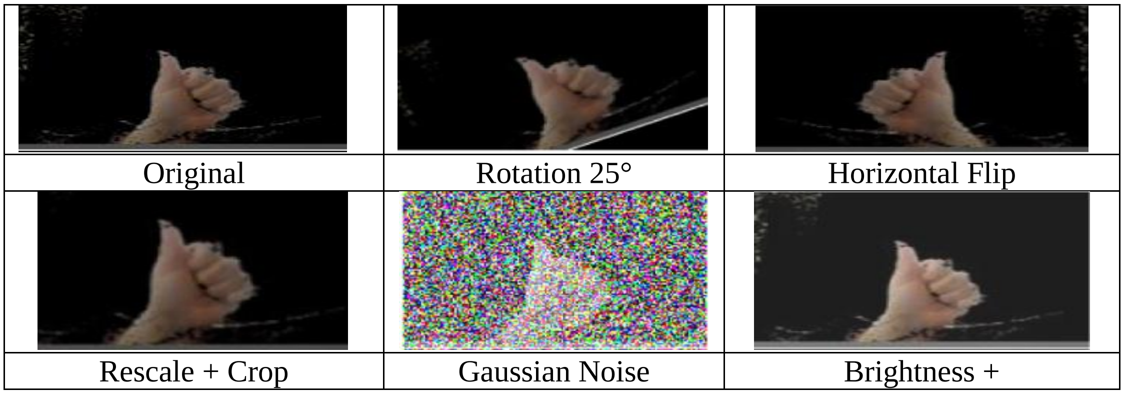


Figure 3-: Augmentation Process

Proficient knowledge and abundant data regarding modern machine learning added on numerous benefits. The up to date knowledge incorporated to new models and the successful proven results give the opportunity to keep moving forward. The improvement of knowledge in machine learning assists in the advanced work. Hence, the current state of knowledge and vast resources initiated the step in improving the models. The successful results give more room to improve and constantly work on more ideas that are modern. The improvement and more knowledge in the particular field helps in the more advanced work and in the state of achieving new models.

The outcomes showed how successful the segmentation technique based on the HSV color space worked in processing the input images to reduce the unwanted background distractions so that the processing concentrated primarily on the hand area. Outcomes also recognized the unique ability of convolutional neural networks to identify and abstract different patterns and delicate spatial characteristics in the processed hand shapes that assisted in improving the proposed system's classification confidence and system reliability.

The experiments showed that the integration of segmentation based on HSV with convolutional networking reached a favorable degree of effectiveness with the highest test set average of 95.3%. Header 3 Linear and smooth within the section and the area's corresponding metrics of precision and recall reached values of over 93.5% and 92.8%, respectively.

Table 2~~.-~~: Performance evaluation of the proposed HSV + CNN model

Metric	**Accuracy**	**Precision**	**Recall**	**F1-score**
Value (%)	95.3	93.5	92.8	93.1

The results in Table 2 indicate that the proposed system is able to achieve a high recognition accuracy of 95.3%, demonstrating the reliability of combining HSV segmentation with CNN-based classification. Precision and recall values remain consistently above 92%, highlighting the system's

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com
مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

robustness in correctly classifying most Arabic letters while minimizing false positives and false negatives. The obtained F1-score of 93.1% further validates the balanced performance of the model.

Additionally, the CNN system demonstrated strong performance in recognizing visually distinct Arabic letters, such as "alif" and "nun," but struggled to distinguish between gestures with subtle differences and similarities (such as "dal" and "dhal"). This was attributed to the similarity of hand configurations in these letters and the strong similarity between finger movements. Table 2 presents the performance metrics of the proposed Arabic Sign Language recognition system based on HSV segmentation and CNN classification.

Despite these promising outcomes, it is important to note that system performance slightly decreased in scenarios involving variable illumination and complex backgrounds. This suggests that while the HSV color space is effective in controlled environments, further refinements (e.g., adaptive thresholding, hybrid color models) are necessary to enhance robustness under real-world conditions. The proposed work was compared with some classification techniques and Table 3 shows the results we obtained from applying the techniques on the same dataset.

To ensure the generalization process and eliminate overfitting, the data was divided equally into 70% for training and the 30% for testing and verification. Cross validation was used, as well as data augmentation shown in Figure 3, which helped eliminate overfitting. The total number of images for each sign is 311 in training and 67 images in the testing and validation phases.

Table 3: Comparative performance of the proposed HSV + CNN model against other classifiers.

Method	Segmentation Technique	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Limitations
HSV + CNN (Proposed)	HSV skin color thresholding	95.3	93.5	92.8	93.1	Sensitive to extreme lighting variations
HSV + SVM	HSV skin color thresholding	87.7	87.4	86.9	87.1	Lower generalization on complex signs
HSV + RF	HSV skin color thresholding	84.3	84.1	83.7	83.9	Prone to overfitting, less efficient
ResNet (No HSV)	Direct raw image input	95.5	94.2	93.9	92.0	High computational cost, needs large dataset

The proposed HSV + CNN system provides competitive results, as seen in Table 3, yielding 95.3% accuracy. Proposed methods incur less computation and smaller dataset requirements, especially given the needs of mobile and other real-time applications. When compared to ResNet, the proposed methods are much more efficient and adaptable to smaller datasets. For comparison purposes, Random Forest RF and Support Vector Machine SVM were also used, but they fell considerably short with 87.7% and 84.3% accuracies, respectively.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

The results do indicate that the implementation of the HSV color space as an early stage preprocessing step improves the effectiveness of the image segmentation process. Nevertheless, in terms of complexity and the diversity of Arabic sign language with the traditional classifiers particularly for images of the gestures, the limitations still remain.

Findings highlight the value of using HSV segmentation and CNN classification as a viable framework for Arabic sign language recognition. This method involves the application of HSV-based segmentation techniques for background removal and convolutional neural networks for classification. This method proves to be highly effective and is especially useful for real-time applications. Moving forward, this approach can be compared to more complex systems like ResNet.

Thanks to convolutional networks and techniques like HSV segmentation, we captured the hand gesture while reducing noise, and we achieved accurate classification without excessive computation, making the approach appropriate for mobile and embedded applications. Conversely, ResNet's complex architecture is the result of its residual connections, which allow training of much deeper networks, more than 60 and probably 100 layers. This facilitates the extraction of high-level, complex, and discriminating features, which enhances performance, particularly on large, complicated datasets. Still, the associated costs expand computation and necess larger datasets. In the use of ResNet for sign language recognition, while classification accuracy is improved, the associated costs include high-end computation, the potential for overfitting, and a large dataset to train the model.

When comparing CNN to ResNet, it is fundamentally a trade-off of efficiency versus accuracy. The proposed CNN model using HSV segmentation, is lite, interpretable, and scalable for real time systems. ResNet, however, is powerful and resource hungry with preference towards scenarios where speed, accuracy, and resource utilization is of less importance. Discussing both approaches in the Arabic Sign Language Recognition context confirms the proposed method's efficiency and relevance in comparison to leading-edge architectures.

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

Table 4: The actual comparison of the application of both CNN and ResNet with Arabic sign language

Aspect	CNN	ResNet
Architecture	Sequential layers (Conv → ReLU → Pooling → FC)	Residual blocks with skip connections (Conv → ReLU → Conv + Input addition)
Depth	Shallow to medium (5–20 layers)	Very deep (50–150+ layers possible)
Training Stability	May be affected by the vanishing phenomenon in deep networks	Avoids the problem of fading through skip connections
Feature Extraction	Focuses on general features (edges, shapes)	Picks up subtle, multi-level patterns
Computational Complexity	Faster and less resource-intensive, which is what the proposed system requires	Requires more memory and processing power(GPU/TPU), which burdens real-time signal recognition
Dataset Requirement	Good with medium-sized datasets	Requires large datasets for effective training

CONCLUSION:

Recognizing sign languages has always been a vital topic in societies, especially with the recent massive expansion of the deaf and mute community. Therefore, it is important to focus on easy and highly accurate techniques that can be relied upon with relatively available resources.

The proposed methodology is highly flexible and scalable, as the HSV color space boundaries can be modified or the CNN architecture optimized to suit research requirements and the characteristics of different datasets.

The experiments resulted in the development of an effective system for Arabic sign language recognition, based on hand segmentation according to the HSV space followed by a classifier based on convolutional neural networks (CNNs). The proposed method showed promising results, as the results of the used criteria ranged between Accuracy 95.3%, Precision 93.5%, Recall 92.8%, and F1-score 93.1%. Understanding the value of preprocessing steps in enhancing the quality of the input data and enabling the network to gain more discriminative feature extraction is fundamental. While there are difficulties due to changes in lighting and background clutter, the methodology has provided a clear ground to build more sophisticated systems, drawing from clear, large-scale, skin-region centered input images. The proposed methodology has also been contrasted with other approaches to showcase its efficacy and dominance in tackling the complexities and the vast range of variations surrounding the Arabic Sign Language.

A notable point from the findings is the effect of The selection of the HSV thresholds on the system operation. The thresholds set during the segmentation produced excellent results and contributed to high classification accuracy. The importance of setting a threshold to large objects is to ignore all the minute details in the background. Providing the network with only the hand significantly reduces the restrictions imposed by the use of HSV.

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652

info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

This research aims to develop an integrated framework for Arabic letter recognition using HSV and CNN. Test the accuracy and performance of the system under different imaging conditions. Perform the system's use in practical applications such as interactive education for deaf children. Improving the user experience in healthcare and communication environments. The expected outcomes are not limited to the academic field but extend to creating a direct societal impact that enhances the quality of life for the deaf and hard-of-hearing.

There is potential to develop the system to handle entire sentences rather than single letters. A combination of HSV and YCbCr or Deep Segmentation techniques could be used. Testing on real mobile devices is also being conducted for practical application.

Conflict of interests.

There are non-conflicts of interest.

References

[1] I. A. ADEYANJU, O. O. BELLO, M. A. ADEGBOYE, " Machine learning methods for sign language recognition: A critical review and analysis", *Intelligent Systems with Applications*, vol. 12, pp. 200056, 2021.

[2] N. C. CAMGOZ, et al., " Sign language transformers: Joint end-to-end sign language recognition and translation, " *in 2020 Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, p. 10023-10033, 2020.

[3] R. E. Rwelli, O.R. Shahin, A.I. Taloba, " Gesture based Arabic Sign Language Recognition for Impaired People based on Convolution Neural Network", *International Journal of Advanced Computer Science and Applications,* vol 12, 2021.

[4] N. Fadel, Noor, E. I. Abdul Kareem, "Detecting Hand Gestures Using Machine Learning Techniques ", *Ingenierie des Systemes d'Information,* 2022.

[5] M. M. Kamruzzaman, "Arabic sign language recognition and generating Arabic speech using convolutional neural network ", *Wireless Communications and Mobile Computing* , no.1,pp. 3685614, 2020.

[6] R. Daroya, D. Peralta, P. Naval, "Alphabet Sign Language Image Classification Using Deep Learning, " *in TENCON 2018, IEEE Region 10 Conference*, pp. 0646-0650, 2018.

[7] R. A. Alawwad, O. Bchir, M.M. Ismail,"Arabic sign language recognition using Faster R-CNN", *International Journal of Advanced Computer Science and Applications*, vol. 12, No. 3, 2021.

[8] A.Z.K. Matloob, M.I. Kareem, H.K. Alwan, "Machine learning-based classification models for efficient DDoS detection ", *International Journal of Computing*, vol. 17, no. 1, pp. 1-13, 2025.

[9] W. Al-Hameed, N. Fadel, "Fuzzy logic for defect detection of radiography images", *Journal of computational and theoretical nanoscience,* vol. 16, no. 3, pp. 1023- 1028, 2019.

[10] S. Suganyadevi,V. Seethalakshmi, K. Balasamy, " A review on deep learning in medical image analysis", *International Journal of Multimedia Information Retrieval* , Vol. 11, No. 1, pp. 19-38, 2022.

[11] S. E. Umbaugh, *Digital image processing and analysis: computer vision and image analysis*, CRC Press, 2023.

[12] A. Alabdulkarim, S. Alghowinem, and A. Alhothali, "A robust Arabic sign language recognition system using Kinect V2 and leap motion controllers, *" in 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC),* pp. 3109-3114, 2021. doi: 10.1109/SMC52423.2021.9658709.

[13] M. Mohandes, S. Aliyu, and M. Deriche, "Arabic sign language recognition through deep neural networks and fuzzy logic", *IEEE Access*, vol. 5, pp. 9512-9522, 2017.doi: 10.1109/ACCESS.2017.2701386.

[14] A. A. Alani and G. Cosma, "Real-time Arabic alphabet sign language recognition using convolutional neural networks, " *in 2019 International Conference on Computational Intelligence and Information Systems (CIIS)*, pp. 1-6, 2019. doi:10.1109/CIIS47705.2019.8947701.

[15] S. Aly and W. Aly, "Continuous Arabic sign language recognition using 3D body joint data and recurrent neural networks, " *in 2020 IEEE Winter Conference on Applications of Computer Vision Workshops (WACVW),* pp. 50-59, 2020. doi:10.1109/WACVW50321.2020.9096933.

[16] T.H. Noor, A. Noor, A.F. Alharbi, A. Faisal, R. Alrashidi, A.S. Alsaedi, G. Alharbi, T. Alsanoosy, A. Alsaeedi, " Real-Time Arabic Sign Language Recognition Using a Hybrid Deep Learning Model", *Sensors,*vol. 24, no. 11, pp. 3683,2024.

\

ISSN: 2312-8135 | Print ISSN: 1992-0652

info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

الخلاصة

يُعد فهم لغة الإشارة والتعرف عليها تلقائيًا أمرًا بالغ الأهمية لتحقيق تكافؤ الفرص للصم وضعاف السمع في المجتمع. في عالم التواصل بلغة الإشارة العربية، يتشارك الناس أفكارهم ومشاعرهم ويتفاعلون يوميًا بشكل هادف. مع تطور تقنيات الذكاء الاصطناعي والرؤية الحاسوبية، احتلت الدراسات التي ركزت على معالجة لغة الإشارة العربية الصدارة، ولا تزال تُمثل مجالًا بحثيًا رائدًا مقارنةً بلغات العالم الأخرى. اقترح هذا البحث منهجية للاستفادة من نموذج مساحة لون HSV للتجزئة الأولية لإزالة ضوضاء الخلفية وتحسين بيانات الإدخال. تُجزأ هذه العملية منطقة اليد أو منطقة الإشارة الدلالية. ثم تُمرر البيانات إلى شبكة عصبية تلافيفية (CNN) لتصنيف الإشارات بدقة. تكمن ميزة هذه الطريقة الجديدة في التفاعل المنسق لتبسيط البيانات التفاضلية من خلال معالجة مساحة اللون، والقوة التمثيلية العميقة لشبكة CNN لتحسين الأداء العام للنظام في التعرف على إشارات غطاء المحرك. يسعى البحث أيضًا إلى سد الفجوة القائمة في مجال معالجة الإشارات العربية من خلال تقديم نموذج متكامل، يفتح آفاقًا واسعة للتطبيقات في مجالات التعليم والرعاية الصحية وخدمات الحكومة الذكية. وقد أظهرت الطريقة المقترحة نتائج واعدة، حيث تراوحت نتائج المعايير المستخدمة بين الدقة 95.3%، والدقة 93.5%، والاستدعاء 92.8%، ودرجة F1 93.1%.

المقدمة:

لطالما كان التعرّف على لغات الإشارة موضوعًا حيويًا في المجتمعات، لا سيما مع التوسع الهائل الذي شهده مجتمع الصم والبكم مؤخرًا. فهي الوسيلة الأساسية للتواصل لأكثر من 70 مليون شخص من الصم وضعاف السمع حول العالم. لذلك، من المهم التركيز على تقنيات سهلة ودقيقة للغاية يمكن الاعتماد عليها بموارد متاحة نسبيًا. تُعد لغة الإشارة واللغة العربية من أهم لغات الإشارة نظرًا لانتشار استخدامها على نطاق واسع في منطقة الشرق الأوسط وشمال إفريقيا. ومع ذلك، تُمثل حواجز التواصل بين هذه الفئة ومجتمع الصم تحديًا مستمرًا يؤثر على الجوانب التعليمية والاجتماعية والمهنية. لذلك، يُمثل تطوير أنظمة حاسوبية قادرة على التعرّف تلقائيًا على لغة الإشارة العربية خطوة استراتيجية لتعزيز التكامل الاجتماعي وسد فجوة التواصل[1].

طرق العمل:

يعتمد نظام التعرف على إيماءات اليد المقترح على منهجية متعددة المراحل تهدف إلى معالجة صور عالية الدقة واستخلاص خصائص فعّالة. تبدأ المنهجية بمرحلة جمع البيانات، حيث تُجمع مجموعة من الصور ومقاطع الفيديو التي تُمثل الحروف والكلمات بلغة الإشارة العربية على مجموعة بيانات كاجل.

الاستنتاجات:

لطالما كان التعرّف على لغات الإشارة موضوعًا حيويًا في المجتمعات، لا سيما مع التوسع الهائل في أعداد الصم والبكم. لذلك، من المهم التركيز على تقنيات سهلة ودقيقة للغاية يُمكن الاعتماد عليها بموارد متاحة نسبيًا. تتميز المنهجية المقترحة بمرونة عالية وقابلية للتطوير، حيث يُمكن تعديل حدود مساحة ألوان HSV أو تحسين بنية CNN لتناسب متطلبات البحث وخصائص مجموعات البيانات المختلفة.

الكلمات المفتاحية:

لغة الإشارة، اللغة العربية لغة الإشارة، HSV، الشبكة العصبية التلافيفية.