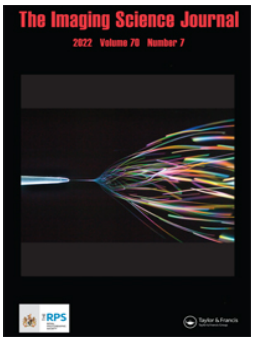

Taylor & Francis
Taylor & Francis Group

The Imaging Science Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/yims20

QSLRS-CNN: Qur'anic sign language recognition system based on convolutional neural networks

Hany A. AbdElghfar, Abdelmoty M. Ahmed, Ali A. Alani, Hammam M. AbdElaal, Belgacem Bouallegue, Mahmoud M. Khattab & Hassan A. Youness

To cite this article: Hany A. AbdElghfar, Abdelmoty M. Ahmed, Ali A. Alani, Hammam M. AbdElaal, Belgacem Bouallegue, Mahmoud M. Khattab & Hassan A. Youness (2023): QSLRS-CNN: Qur'anic sign language recognition system based on convolutional neural networks, The Imaging Science Journal, DOI: 10.1080/13682199.2023.2202576

To link to this article: https://doi.org/10.1080/13682199.2023.2202576

Published online: 29 Apr 2023.

THE IMAGING SCIENCE JOURNAL
https://doi.org/10.1080/13682199.2023.2202576

RESEARCH ARTICLE

QSLRS-CNN: Qur'anic sign language recognition system based on convolutional neural networks

Hany A. AbdElghfar[a,b], Abdelmoty M. Ahmed [c], Ali A. Alani[d], Hammam M. AbdElaal[e], Belgacem Bouallegue [c], Mahmoud M. Khattab[c] and Hassan A. Youness[a]

[a]Department of Computers and Systems Engineering, Faculty of Engineering, Minia University, Egypt; [b]Higher Thebes's institute of Engineering, Cairo, Egypt; [c]College of Computer Science, King Khalid University, Abha, Saudi Arabia; [d]Department of Computer Science, University of Diyala, Diyala, Iraq; [e]Department of Information Technology, Faculty of Computers and information, Luxor University, Egypt

ABSTRACT

Deaf and dumb Muslims face educational barriers. They can't read, recite, or comprehend the Holy Qur'an, hence they can't practise Islamic ceremonies. This study proposes a CNN-based Qur'anic sign language recognition methodology. First, photos are used to train for dynamic and static gesture recognition. Second, preparing images diversifies datasets. Finally, CNN-based deep learning models extract and classify features. To teach the deaf and dumb Islamic ceremonies, the programme recognises Arabic sign language hand motions referring to dashed Qur'anic letters. Only 24,137 photos of the Holy Qur'an's 14 dashed letters were used in the trials from ArSL2018, a huge Arabic sign language collection. SMOTE raises training and testing accuracy to 98.31% and 97.67%, respectively, whereas the proposed model reaches 98.05% and 97.13%. RMU obtains 98.66% and 97.52% training and testing accuracy, whereas RMO achieves 98.37% and 97.36%.

ARTICLE HISTORY
Received 15 October 2022
Accepted 11 April 2023

KEYWORDS
Holy Qur'an; Qur'anic sign language; RMO; convolutional neural network; SMOTE; sign language; deep Learning; feature extraction

Introduction

Deaf and dumb people use sign language to communicate with others in their daily lives. Sign language is uncommon outside the deaf community, and communication between deaf and hearing people is a major challenge. Some hearing parents have deaf children, which generates a language gap. Deaf children are difficult to raise, nurture, and teach Islamic customs [1,2]. Numerous Arabic sign languages (ArSLs) use the same alphabetic. Deaf and dumb Arabs use Egyptian, Jordanian, Tunisian, and Gulf sign languages. Lack of information, an inability to communicate, and an inability to perform religious ceremonies create a gap. All of ArSL's issues need machine translation to enable the deaf to attend school and get scientific knowledge in their native language [1,2]. Pattern recognition in human–computer interaction shifts to computer vision and machine learning. Therefore, deaf hand gestures are a major tool for recognizing Qur'anic alphabetical letters.

This paper aims to recognize the movements of the Arabic sign language by recognizing the hand gestures that refer to the dashed Qur'anic letters in order to help the deaf and dumb learn their Islamic rituals. This is due to the fact that deaf and dumb Muslims cannot reach advanced levels of education due to the impact of obstruction on their educational attainment, which leads to their inability to learn, recite, and understand the meanings and interpretations of the Holy Qur'an as easily as ordinary people, which also prevents them from applying Islamic rituals such as prayer that require learning and reading the Holy Qur'an.

In this paper, we focus on Qur'anic sign language recognition systems (QSLRS), which enable the deaf and dumb community to overcome communication challenges, learn the Islamic rituals, and learn the Arabic alphabet, which is the language of the Holy Qur'an. Through the use of QSLRS, they can identify the 29 Quranic Surah's with the dashed letters, which are 14 letters arranged in alphabetical order: 'ا ح ر س ص ط ع ق ك ل م ن ه ي'. However, according to previous studies, this remains a challenge for academics and researchers due to the fact that the automated detection algorithms for ArSLs are inaccurate and the recognition field is small. Therefore, we propose a new model for QSLRS based on convolutional neural networks (CNNs) through data preparation, preprocessing, feature extraction, and classification stages in order to help the deaf and dumb learn their Islamic rituals, overcome challenges of communication with others, and learn the Arabic alphabetical characters. The main contributions of this paper are as follows:

CONTACT Abdelmoty M. Ahmed abd2005moty@yahoo.com

i) Identification of fixed alphabetic signs for ArSL in order to help the deaf and dumb learn the Quranic Surah's that begin with dashed letters.
ii) Formation and construction of a new Qur'anic sign language (QSL) dataset built from a large ArSL dataset called ArSL2018, which consists of only 24137 samples representing 14 dashed letters from the beginnings of the Holy Qur'an Surahs to implement the proposed model.
iii) The use of image augmentation approaches makes the proposed QSLRS-CNN model work better in real-life scenarios and reduces overfitting.
iv) Previous research findings indicate that there are currently no studies published or available online on the subject of Qur'anic sign language recognition (QSLR). As a result, one of the purposes of this study is to advance the studies that are being conducted in this field and to stimulate the production of materials that can be used for further research in this field.

The remainder of this paper is structured as follows: The related works are presented in Section 2. The proposed methodology is presented in Section 3. In Section 4, the experimental results are presented and the main findings are discussed. Lastly, the conclusion is outlined in Section 5.

Related works

The literature has several strategies for human–computer sign language recognition. These devices increase communication by understanding gestures and signs. These techniques involve capture, preprocessing, gesture representation, feature extraction, and classification. This section explores movement recognition strategies in sign languages.

There are several research initiatives to create sign language recognition systems across the globe, including in Arab countries, which rely on vision or sensor gloves. This work focuses on vision-based systems that recognize ArSL alphabets.

This section examines studies on the identification of the ArSL alphabet as well as the size of the datasets used in these studies. The algorithms or methodologies used by researchers are also discussed.

In [3], Abdelmoty et al. propose an ArSL translation system. As for the Arabic text system (ATASAT), this system relies on building two datasets for Arabic alphabet gestures. They introduce a new manual detection technique that detects and extracts Arabic sign gestures from an image or video depending on the hand's coverage. They also use different statistical classifiers and compare the results to get a better classification.

In [4], they propose a machine-learning-based Arabic sign language alphabet recognition system. They evaluate 2,800 images and 28 alphabets, with 10 participants in each class. There are 100 images for each letter, for a total of $28 \times 100 = 2{,}800$ images. Feature extraction is performed using a hand shape-based description, where each hand image is characterized by a vector of 15 values indicating key point locations, while classification is performed with K-nearest neighbors (KNN) and multilayer perceptron (MLP) algorithms. Testing shows 97.548% accuracy.

Luqman and Mahmoud [5] study Fourier, Hartley, and Log-Gabor transforms for ArSL recognition (ArSLR). The Hartley transform detects ArSLR with 98.8% accuracy using the support vector machine (SVM) classifier. Alzohairi [6] can automatically recognize 63.5% of the movements of the Arabic alphabet by using a method that is based on images.

In 2020, Kamruzzan [7] develops a vision-based approach for identifying Arabic hand-sign-based characters and converting them into Arabic speech with a 90% recognition rate using CNN. Eibadawy et al. [8] propose a CNN-based framework for recognizing 25 ArSL signs. On the basis of the training and test data, this model has accuracy scores of 85 and 98 percent, respectively. Mohamed describes in [9] a computerized system that employs depth-measuring cameras and computer vision techniques to capture and segment images of facial expressions and hand gestures with a 90% recognition rate. Ghazanfar et al. [10] propose different CNN architectures using 54,049 sign images [11]. Their findings demonstrate the considerable influence that the size of the dataset has on the correctness of the model that is presented. The proposed model's test accuracy increases from 80.3 percent to 93.9 percent when the amount of the dataset is raised from 8,302 samples to 27,985 samples. A further improvement in the proposed model's test accuracy occurs when the amount of the dataset is increased from 33406 samples to 50,000 samples, resulting in a corresponding rise from 94.1 percent to 95.9 percent.

In [12], Alani and Cosma develop an Arabic sign recognition system based on the ArSL2018 dataset and a unique ArSL-CNN architecture. The accuracy of the suggested ArSL-CNN model during training is originally 98.80 percent, whereas the accuracy during testing is initially 96.59 percent. They decide to use a variety of resampling strategies on the dataset in order to mitigate the effect that imbalanced data has on the precision of the model. Based on the findings, the synthetic minority oversampling method (SMOTE) results in an improvement in overall test accuracy from 96.59 percent to 97.29 percent.

Yaser and Ghassan [13] utilize the ArSL2018 to improve the accuracy of recognizing 32 hand motions from the ArSL-CNN using transfer learning and fine tuning deep CNNs. To address the imbalance produced by class size disparity, the dataset is subjected to random under-sampling. The total number of images is reduced from 54,049–25,600. The generated model has a 99.4 percent validation accuracy for the visual geometry group (VGG-16) and a 99.6 percent validation accuracy for the ResNet-152.

Shahin and Almotairi [14] suggest a deep transfer learning-based strong identification approach for ArSL. To reduce overfitting and improve performance, they use transfer learning techniques based on fine-tuning and data augmentation. The proposed residual network ResNet101 system achieves maximum accuracy with a score of 99.52 percent.

Abeje et al. [15] offers a unique sign language recognition system that converts Ethiopian sign language (ETHSL) to Amharic alphabets using computer vision and a deep CNN. The system receives sign language graphics and outputs Amharic. Preprocessing, feature extraction, and recognition make up the suggested system. The methodology includes data gathering, preprocessing, backdrop normalization, picture scaling, ROI identification, noise reduction, brightness correction, and feature extraction. A deep CNN is utilized for end-to-end classification. The JPEG images were gathered under controlled conditions. images. Adjusting the image size and colour reduced the running time. In addition, the findings reveal better recognition accuracy. The model achieves 98.5% training, 95.59% validation, and 98.3% testing accuracy.

Tamiru et al. [16] discuss the construction of an autonomous Amharic sign language translator utilizing digital image processing and machine learning methods. Preprocessing, segmentation, feature extraction, and classification are the four key system steps. Thirty-four characteristics are retrieved from the form, motion, and colour of hand motions to depict Amharic sign symbols. Artificial neural networks (ANN) and multi-class SVM classification models are used. The recognition system can recognize Amharic alphabet signs with an average accuracy of 80.82 and 98.06 using ANN and SVM classifiers, respectively.

Despite recent advances in deep learning and the high precision of image categorization and prediction obtained with CNN, unbalanced data can have an impact on prediction model performance. Imbalanced data can have an influence on a model's capacity to learn and its ability to be used in real-time scenarios. It's also worth looking at how sign language movements are translated into other mediums, such as writing and voice.

In the most current literature reviews, there are various research publications pertaining to ArSLR. In Table 1, we provide a concise description of the ArSLR systems that have been used in the past. In [1,3,6,8,10], and [12–14,17–19], some proposed approaches and models for detecting the Arabic script have inadequate datasets. Latif et al. [11] presented the ArSL2018 dataset. This amounts to 54,049 samples.

Materials and methods

In this section, we present a broad explanation of the architecture of the QSLRS-CNN, which is created to categorize the motions used in the QSL. In addition, it provides a description of the QSL dataset as well as the preprocessing methods that are used on the dataset.

QSL dataset

In this study, a portion of a dataset called ArSL2018 [11] is used, which contains 24,137 images of the ArSL alphabet created by more than 40 people for 14 letters representing the openings of the Qur'anic Surah's, which are letters arranged alphabetically: ('ي ه ن م ل ك ق ع ط ص س ر ح ا '), and in another order: ('ن ص ح ك ي م ل ه س ر ق ا ط ع ') so that an explicit sentence is ('نص حكيم له سر قاطع '). Before being employed by the proposed model, the images are pre-processed, and the dataset is separated into three groups for training, testing, and validation. It is then made available to researchers in machine learning and deep learning, consisting of greyscale images with dimensions of 64*64. It is executed in various forms of images with specific lighting and context, as illustrated in Figure 1. As mentioned earlier, it consists of a total of 14 output classes, ranging from 0 to 13, each representing an ArSL gesture. Table 2 shows the various classes along with their labels and number of samples.

In this study, we also use a part of the other ArSL dataset to test the proposed QSLRS-CNN model that is collected previously [3]. This dataset consists of 350 colour images representing the gestures of 14 Arabic letters, with an average of 25 images per character gesture. As mentioned in [20], the dataset is taken in different ways, under different lighting conditions, and based on different signers with different hand sizes and wearing dark-coloured gloves, as it is shown in Figure 2.

In order to solve the under-sampling problem when the dataset consists of classes of different sizes in terms of data elements, the class imbalance problem occurs, which leads to bias towards the majority class and negatively affects classification accuracy. To rectify the imbalance and eliminate bias, sample number fixing is used. ArSL2018 data samples are shown in Table 2. Ha contains 1526 images, whereas Ain has just 2114. Before using QSLRS-CNN data, the number of samples is fixed.

Table 1. Comparison of the characteristics of recent ArSLR systems.

Ref/Year	Dataset Size and Type	Classifier	Accuracy in %
[1]/2020	100 Arabic signs from regular ArSL	Euclidian distance	95%
[3]/2016	Static alphabet two datasets (Dataset 1: 700 samples for 28 characters using bare hands; Dataset 2: 700 samples for 28 characters using coloured gloves)	C4.5(J48), MLP, K-NN(IBK) and Naïve-Bayesian classifiers	80.67, 88.66, 90.7%, and 84.4% for dataset 1 and 89.5, 94.11, 97.5, and 96.63 for dataset 2
[6]/2018	30 people shot actual images using Smartphones. Volunteers gesture the 30 ArSL alphabets. Each letter uses 30 of the 900 images	SVM	63.5%
[8]/2017	The dataset has 25 words from unified ArSL dictionary	3D CNN – soft-max layer	98% accuracy for observed data and 85% average accuracy for new data
[10]/2020	ArSL2018 is comprised of a total of 54,049 image for the 32 ArSL alphabets and signs, which are gathered from 40 different participants	CNN	97.6%
[12]/2021	(ArSL2018) which contains 54,049 images of 32 sign language gestures	CNN	97.29%
[13]/2020	ArSL2018 contains 54,049 images distributed around 32 classes of Arabic signs	deep CNN	99%
[14]/2019	ArSL2018 dataset consists of 54,049 images with 32 class	Transfer learning approach of deep CNN	99.52%
[17]/2018	450 coloured ArSL videos captured at a rate of 30 fps	Euclidean distance classifier	97%
[18]/2011	6,000 sign images are obtained from six gestures	Max-pooling CNN (MPCNN)	96%
[19]/2019	7869 images for recognizing 28 Arabic letters and numbers from 0 to 10	CNN	90.02%

Dataset problems

In machine learning, classification involves training a system using labelled datasets to classify an unknown dataset. In recent years, data has grown, but labelled data is scarce. Oversampling and under-sampling are methods for changing a dataset's class distribution (ratio of classes or categories). Most shallow machine learning approaches rely on target classes having the same number of training examples. But in many cases, this assumption is wrong. The models favour the majority class and exclude the minority class since almost all instances are identified with one class and few with the other. When datasets are imbalanced, model performance suffers, and a class imbalance exists. In this circumstance, we may have good accuracy but poor precision, recall, and F1-score [21].

Our dataset has an in-class imbalance. To balance the dataset, resampling is used. Resampling may under-sample or oversample the dataset. Under-sampling reduces the number of majority target samples. Oversampling involves developing new examples or repeating current ones while boosting minority class samples. Borderline-SMOTE [22] is an example of an oversampling approach. In this paper, the unbalanced dataset of ArSL2018 is utilized with multiple machine learning models to explore oversampling and under-sampling strategies and compare different evaluation measures. The next parts of this paper offer oversampling and under-sampling findings for our proposed machine learning classification models.

Data preparation and preprocessing

Before testing the dataset, all gesture images are converted to 64*64 greyscale images. This removes RGB gesture image over-processing. The image format has been changed from int8 to float32 for efficiency and speed of training, while the images may lose some information that can be retrieved [23,24]. Images are standardized using 0–1 pixel values. The dataset is adjusted to meet CNN's formatting requirements. The gesture images are randomly selected. The dataset is divided into a testing set (20%) and a training set (80%).

Proposed QSL-CNN model architecture

We propose a new model for Qur'anic sign language recognition based on convolutional neural networks (CNNs) where the CNN has recently attracted a lot of interest in several fields, including natural language processing (NLP) and image recognition. It is more efficient in determining the high-dimensionality of the input data. Convolution layers use a

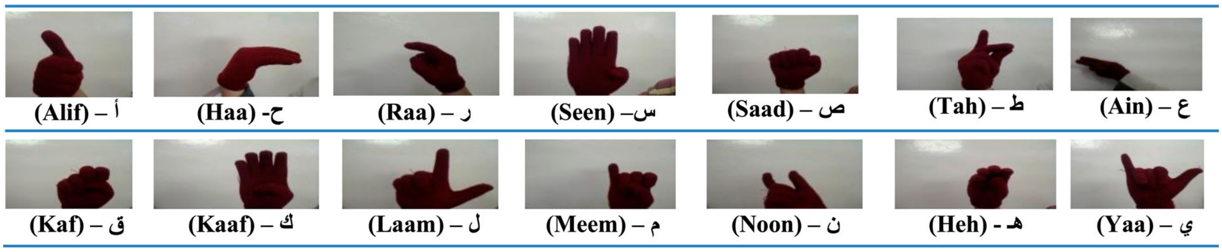

Figure 1. Samples of ArSL alphabet of our dataset.

Table 2. Number of images available in the ArSL2018 dataset for selected gestures.

	Gesture	Samples No.	Gesture	Samples No.	Gesture	Samples No.	Total
classes	Alif (أ – ألف)	1672	Tah (ط – طاء)	1816	Mim (م – ميم)	1838	5326
	Ha (ح – حاء)	1526	Ayn (ع – عين)	2114	Non (ن – نون)	1766	5406
	Ra (ر – راء)	1659	Qaf (ق – قاف)	1526	Haa (هـ – هاء)	1552	4737
	Sin (س – سين)	1638	Kaf (ك – كاف)	1672	Ya (ي – ياء)	1672	4982
	Sad (ص – صاد)	1895	Lam (ل – لام)	1791			3686
Total	**5 classes**	**8390**	**5 classes**	**8919**	**4 classes**	**6828**	**24137**

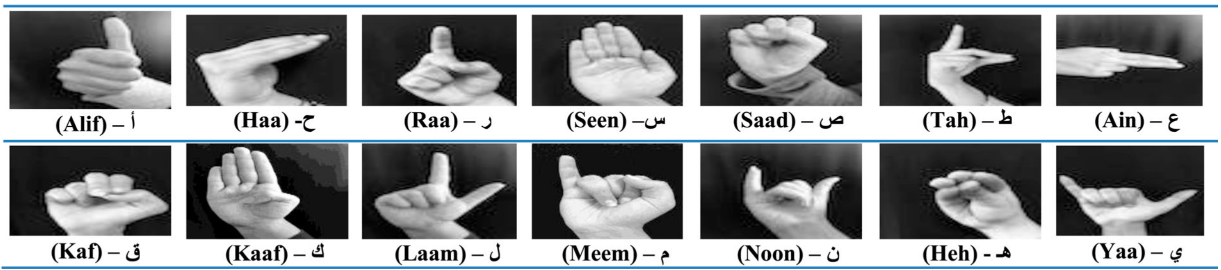

Figure 2. Samples of ArSL alphabet of ArSL2018 dataset.

series of filters, or 'kernel,' which are built on top of each other to learn the complicated features. Every layer has the ability to learn certain features of the input data. Edge detection is performed by the first layer, shape detection is performed by the next, and feature detection is performed by the last layer, which combines all of this data. A major benefit of the CNN over competing deep learning models is that it requires the fewest training parameters. It can speedily categorize network traffic while also cutting down on training time. The CNN may also be employed extensively for network packets due to the fact that the structure of network packets is comparable to the structure of words and traffic is comparable to sentences and articles.

The significance of deep learning and machine learning is rising rapidly in today's world. The dataset is gathered from numerous sources throughout the process of creating the analytical model utilizing deep learning or machine learning. However, the data gathered cannot be utilized immediately to conduct the necessary analysis. In order to maximize the benefits of the machine learning and deep learning models, we need to make sure the data is in the appropriate format. Therefore, data preparation and preprocessing are done to solve this problem.

The general framework of the proposed CNNs-based model consists of data preparation, data preprocessing, feature extraction, and classification stages as shown in Figure 3. The first stage is to identify dynamic and static gestures. Since deep learning models require data for training, gathering images to form a viable training set is the first step through data preparation stage. Data preprocessing is the next stage to perform transformations on each image for further diversification of each dataset. The final stage is to choose CNN-based deep learning model for feature extraction and classification on training and testing dataset.

CNNs, a deep learning approach, are good at image classification. Convolutional, pooling, and fully connected layers are utilized to develop CNN model architectures. Several of these layers are stacked to form a CNN. The structure that is mentioned below is the one that we employ in the proposed CNN model. Convolutional layers make up the first two layers of the neural network. Each of these convolutional layers has 32 feature maps, each of which has a kernel size of 3*3 pixels and uses the rectified linear units (ReLU) activation function. A batch normalization (BN) layer follows these two layers to normalize inputs and produce consistent activation value distributions via training.

Next, we define a Max pooling layer with a 2*2 pool size. Next is a dropout regularization layer. Randomly excluding 0.1 neurons reduces overfitting. The following hidden layer is composed of two convolutional layers with 64 feature maps and 3*3 kernels. Both layers employ ReLU activation. Then come another BN layer, a pooling layer, and a dropout layer that randomly excludes 0.1 neurons. The next hidden layer has two convolutional layers with 128 feature maps with ReLU activation and 3*3 kernels. These two layers are followed by a BN layer, a pooling layer, and a dropout layer that randomly excludes 0.3 neurons. The final hidden layer is a convolutional network with 256 feature mappings and a 3*3 kernel size. This layer is followed by another BN layer, a pooling layer, and a dropout layer that excludes 0.3 neurons randomly. After that, we get a layer that is called the 'flatten layer.' This layer takes the data from the two-dimensional matrix and turns it into a vector. This allows the final output to be processed by typical fully connected layers, allowing us to move on to the next layer. The subsequent layer of the fatten network is a fully linked one, and it has 512 neurons that use the ReLU activation function. In the conclusion, we conclude the QSLRS-CNN proposed model with the output layer, which presents the ultimate classification result. This layer is comprised of 14 neurons, one for each of the 14 classes, and uses a soft-max activation function. The graphical depiction of the model can be seen in Figure 4, and Table 3 provides information about the model's parameters.

Results and discussion

The experiments use Keras packages and Python Tensor Flow. QSLRS-CNN is trained on a system with an NVIDIA K80 GPU, 12 GB of RAM, and a 100 GB SSD. To eliminate bias, the training dataset is scrambled before being given to the network. The proposed QSLRS-CNN model is trained and tested using 14 classes from the original ArSL2018 dataset; the model is then trained and tested using alternative resampling strategies to solve class imbalance. Accuracy is used to evaluate the QSLRS-CNN technique. A signifies accuracy; TC and FC represent properly and wrongly categorized cases, respectively. Multiplying the computed amount by 100 gives a percentage at Equation (1).

$$A = \mathrm{TC}/(\mathrm{TC} + \mathrm{FC}) * 100 \quad (1)$$

For a class, the accuracy can be determined using Equation (2).

$$\mathrm{Ac} = \mathrm{TCc} \ / \ (\mathrm{TCc} + \mathrm{FCc}) * 100 \quad (2)$$

TCc is the number of properly categorized examples from class c, while FCc is the number of wrongly classified instances. The final number is multiplied by 100 to determine each class' accuracy.

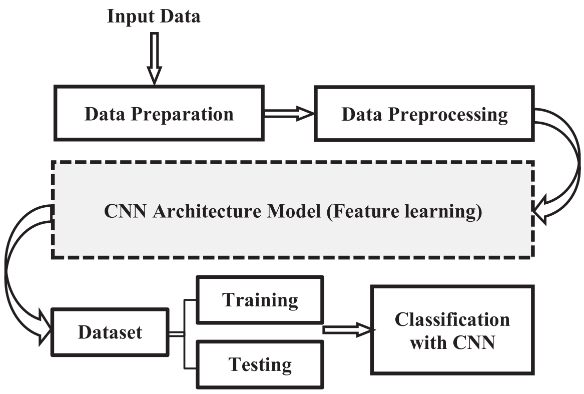

Figure 3. The general framework of the proposed CNNs-based model.

QSLRS-CNN performance evaluation

Table 4 shows the performance of the proposed QSLRS-CNN model on 14 ArSL2018 classes. The training dataset includes 24,137 images in 14 ArSL gesture groupings. Each training batch includes 128 samples. Each of the input and output layers contains 4,096 neurons. QSLRS-CNN is trained across numerous epochs. QSLRS-CNN obtains 97.13% accuracy after 100 learning epochs. Figure 5 shows the model's accuracy after 100 epochs of training. Training and testing performances are similar across epochs, indicating no overfitting. The learning rate of 0.1 is a traditionally common default value.

Figure 6 shows the QSLRS-CNN model's training and testing accuracy curves after 100 epochs. Different epochs have similar training and testing performances.

Figure 6 shows QSLRS-100-epoch CNN's confusion matrix. Off-diagonal entries in the confusion matrix represent mislabeled images. The sum of the confusion matrix's diagonal values indicates classification accuracy.

QSLRS-CNN performance evaluation while oversampling and under-sampling

In the previous section, QSLRS-CNN model findings are produced without data sampling. This study employs resampling approaches (oversampling and under-sampling) to eliminate bias and address data imbalance in class distribution, which involves modifying the previous distribution for the minority and majority classes.

A. Oversampling

Oversampling generates synthetic samples from minority samples to correct class imbalances. This improves classification performance by increasing the

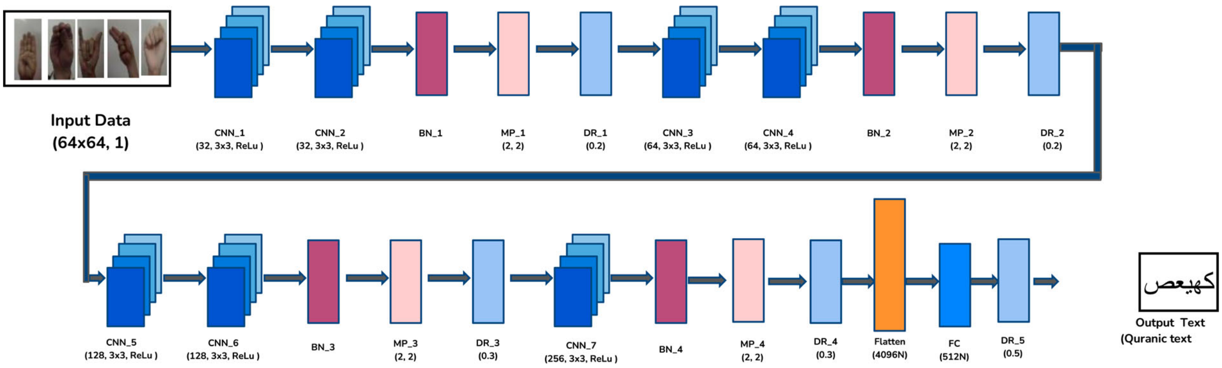

Figure 4. QSLRS-CNN architecture.

quantity of minority class samples. Increasing minority class samples lengthens training. Random minority oversampling (RMO) randomly repeats samples from minority classes. The second is SMOTE, which solves class imbalance by interpolating neighbouring data points. Table 5 shows the QSLRS-CNN model's findings on the ArSL2018 dataset after RMO and SMOTE. Applying oversampling methods boosts the QSLRS-CNN model's efficiency. Using RMO, the proposed model achieves 98.37% training accuracy and 97.36% testing accuracy. Using SMOTE oversampling, the proposed model achieves 98.31% training accuracy and 97.67% testing accuracy.

Figure 7 shows model accuracy and training loss while using the SMOTE oversampling strategy. This graph demonstrates that the training and testing performances are close over various training and testing epochs, which indicates that the QSLRS-CNN model isn't overfitting the data.

Table 3. QSLRS-CNN parameters.

Layers Operation	Filters	Pool Size	Kernel Size	Dropout Size	Parameters No.
Convolution_1	32	–	3 × 3	–	320
Convolution_2	32	–	3 × 3	–	9248
BN_1	32	–	–	–	128
Max-pooling_1	32	2 × 2	–	–	–
Dropout_1	32	–		0.1	–
Convolution_3	64	–	3 × 3	–	18496
Convolution_4	64	–	3 × 3	–	36928
BN_2	64	–	–	–	256
Max-pooling_2	64	2 × 2	–	–	–
Dropout_2	64	–		0.1	–
Convolution_5	128	–	3 × 3	–	73856
Convolution_6	128	–	3 × 3	–	147584
BN_3	128	–	–	–	512
Max-pooling_3	128	2 × 2	–	–	–
Dropout_3	128	–		0.3	–
Convolution_7	256	–	3 × 3	–	295168
BN_4	256	–	–	–	1024
Max-pooling_4	256	2 × 2	–	–	–
Dropout_4	256	–		0.3	–
Flatten layer	4096	–	–	–	–
Fully connected	512	–	–	–	2097664
Dropout_5	512	–	–	0.5	
Fully connected	14	–	–	–	7182

Figure 8 shows the confusion matrices of a 100-epoch QSLRS-CNN model trained using SMOTE oversampling. Overall, classification performance is good.

B. Under-sampling

Under-sampling of minorities at random is the second strategy for changing the distribution of samples across all of the classes in the ArSL2018 dataset with random minority under-sampling (RMU). The dataset is not balanced until RMU excludes samples at random from classes with the majority membership. However, this may result in the loss of valuable information if there are fewer samples taken from members of the minority classes. Table 6 summarizes the findings that are obtained via the application of the QSLRS-CNN model to the ArSL2018 dataset after RMU. Through the use of the under-sampling strategy, the proposed model is able to attain training and testing accuracy of 98.66 and 97.52 percent, respectively.

When the RMU approach is used, the model accuracy and training loss are shown in Figure 9.

Figure 10 displays the confusion matrices of the QSLRS-CNN model, which is trained with a total of 100 epochs and makes use of the RMU oversampling strategy.

Comparative analysis

The accuracy comparison of the proposed model with current state-of-the-art methods on the ArSL2018 dataset is provided in Table 7. The results demonstrate that the proposed QSLRS-CNN model outperforms current state-of-the-art methods in terms of accuracy and training time when performing RMO, SMOTE, and RMU resampling on the dataset. The original CNN obtains an accuracy of 98.05% and

Table 4. Accuracy result from 14 derived ArSL2018 classes.

No. Epochs	Training Acc. (%)	Testing Acc. (%)	Training Time (mins)
100	98.05	97.13	20.61

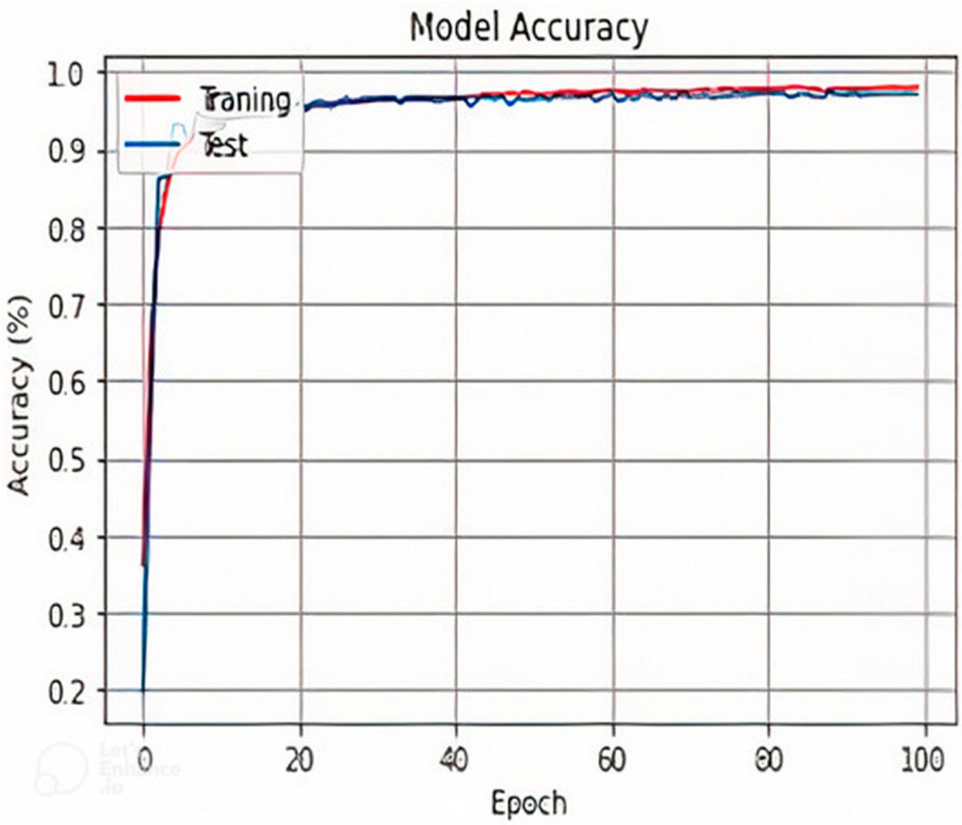

Figure 5. Test, train accuracy of 14-class QSLRS-CNN model.

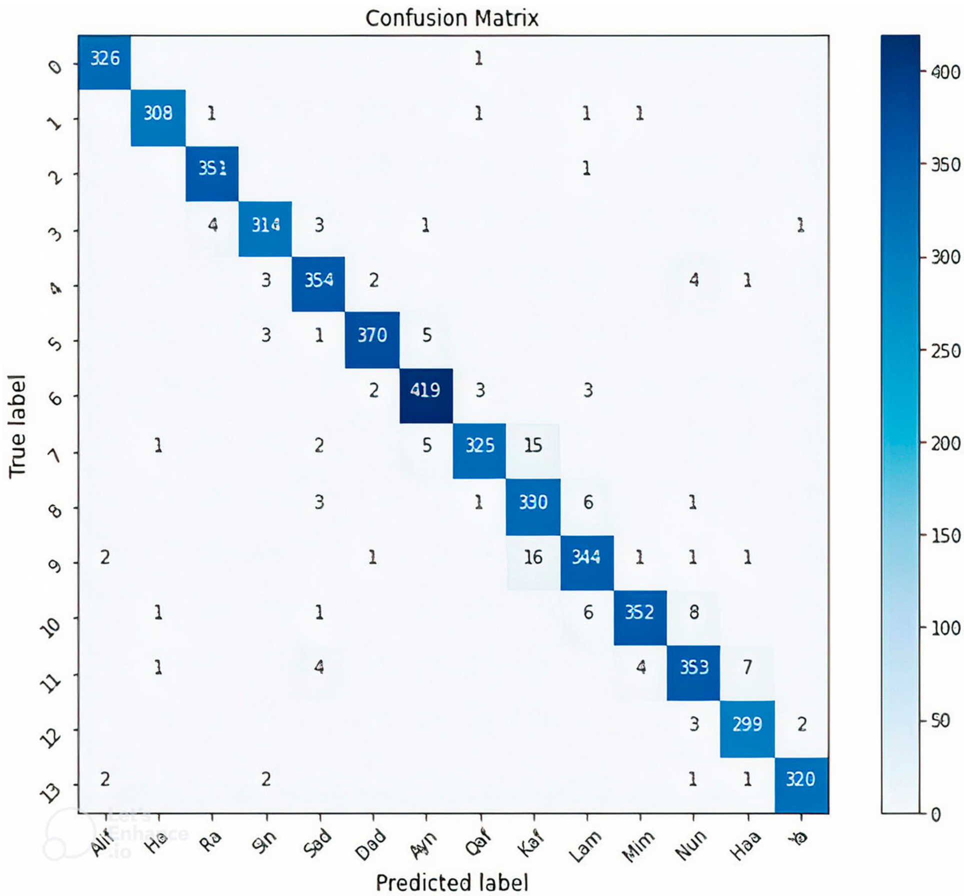

Figure 6. QSLRS-CNN confusion matrix.

Table 5. QSLRS-CNN test data accuracy after RMO and SMOTE.

Resampling Technique	No. Epochs	Training Acc. (%)	Testing Acc. (%)	Training Time (mins)
RMO	100	98.37	97.36	25. 1
SMOTE	100	98.31	97.67	30.56

97.13% for training and testing, respectively on a dataset of 24,137 samples where it takes the least training time of 20.16 min. Ghazanfar et al. [10] obtain a training and testing accuracy of 97.6% and 97.1%, respectively on a dataset of 50000 samples with a very large training time of 486 min, while Alani and Cosma [12] perform CNN to get a training and testing accuracy of 98.80% and 97.29%, respectively on a dataset of 54049 samples with a training time of 141.9 min.

On the other hand, the proposed models have been conducted on a part of a large Arabic sign language dataset called ArSL2018, which contains only 24,137 images for the 14 dashed letters in the Holy Qur'an. The proposed model using SMOTE increases training and testing accuracy to 98.31% and 97.67%, respectively with a training time of 30.56 min. In addition, the proposed model using RMU achieves 98.66% and 97.52% training and testing accuracy respectively with a training time of 21.18 min, while it achieves 98.37% and 97.36% using RMO with a training time of 25.1 min. Therefore, in comparison, the proposed models achieve a superior level of accuracy and training time, which highlight the relevance of providing an appropriate amount of samples to improve the generalization efficacy of CNN while training deep learning models, and also achieve better performance than the other existing models.

Discussion and recommendations

Within the scope of this investigation, we develop an Arabic sign recognition system by making use of a novel QSLRS-CNN architectural framework. Experiments are carried out using the ArSL2018 dataset serving as the basis. The dataset initially had 24137 images gathered from a total of 40 users, which are then divided up into 14 categories after being organized. The proposed QSLRS-CNN model obtains training and testing accuracies of 98.05% and 97.13%, respectively. The findings illustrate the difficulties associated with dealing with unbalanced data and, as a consequence, highlight the need to supply a suitable amount of samples from each class in order to adequately test and train deep learning models.

The findings also show how the accuracy of the model is affected by the presence of unbalanced data. The dataset undergoes a series of tests, during which time it is subjected to many resampling strategies. According to the findings, using SMOTE elevates the total test accuracy from 97.13% to 97.67%, resulting in an increase in test accuracy that is statistically significant. The QSL-CNN model that is presented may be trained on a range of ArSL, which can help to eliminate the communication hurdles that are experienced by deaf people in nations that speak Arabic. The results provide additional evidence that the SMOTE oversampling strategy is successful when

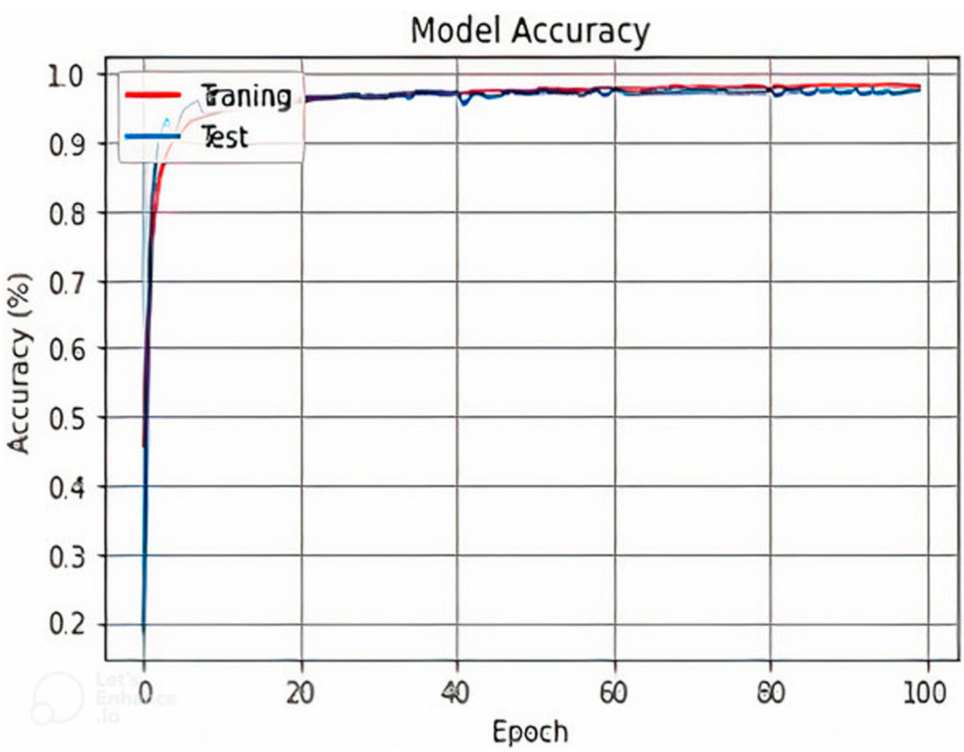

Figure 7. Test and train accuracy of the proposed QSLRS-CNN model using SMOTE.

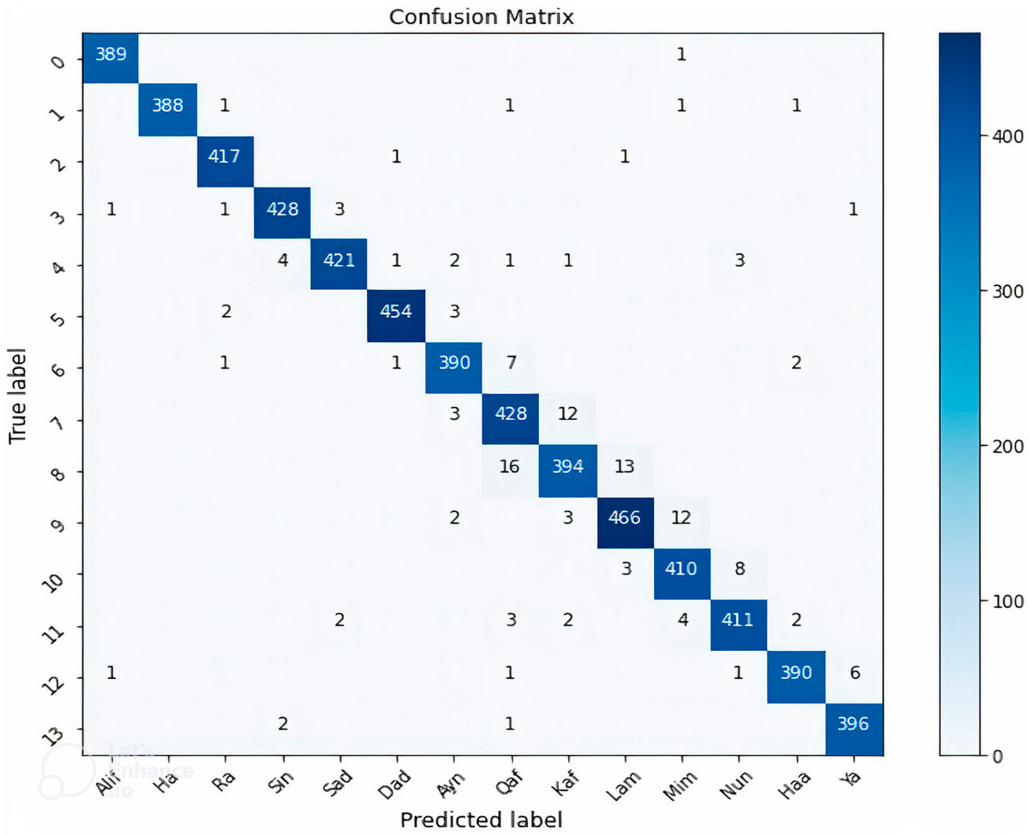

Figure 8. Confusion matrix of the proposed QSLRS-CNN model using SMOTE.

Table 6. QSLRS-CNN tests data accuracy after RMU.

Resampling Technique	No. Epochs	Training Acc. (%)	Testing Acc. (%)	Training Time (mins)
RMU	100	98.66	97.52	21. 18

using the ArSL2018 dataset. As far as we know, our research is the first of its kind to analyze the class imbalance present in the ArSL2018 dataset which focus on 14 letters representing the openings of the

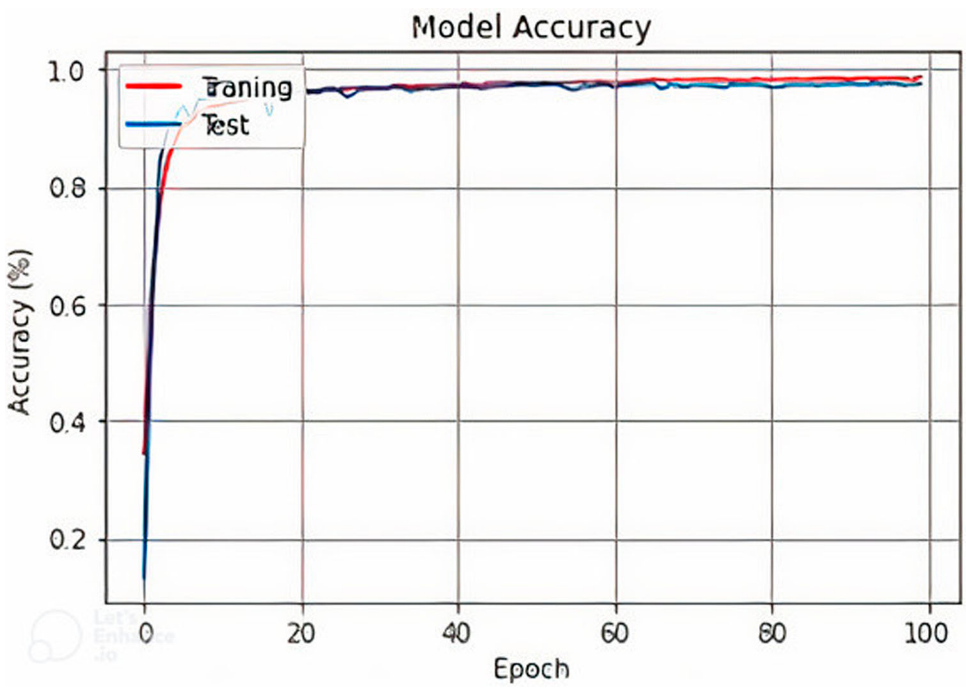

Figure 9. Test and train accuracy of the proposed QSLRS-CNN model using RMU.

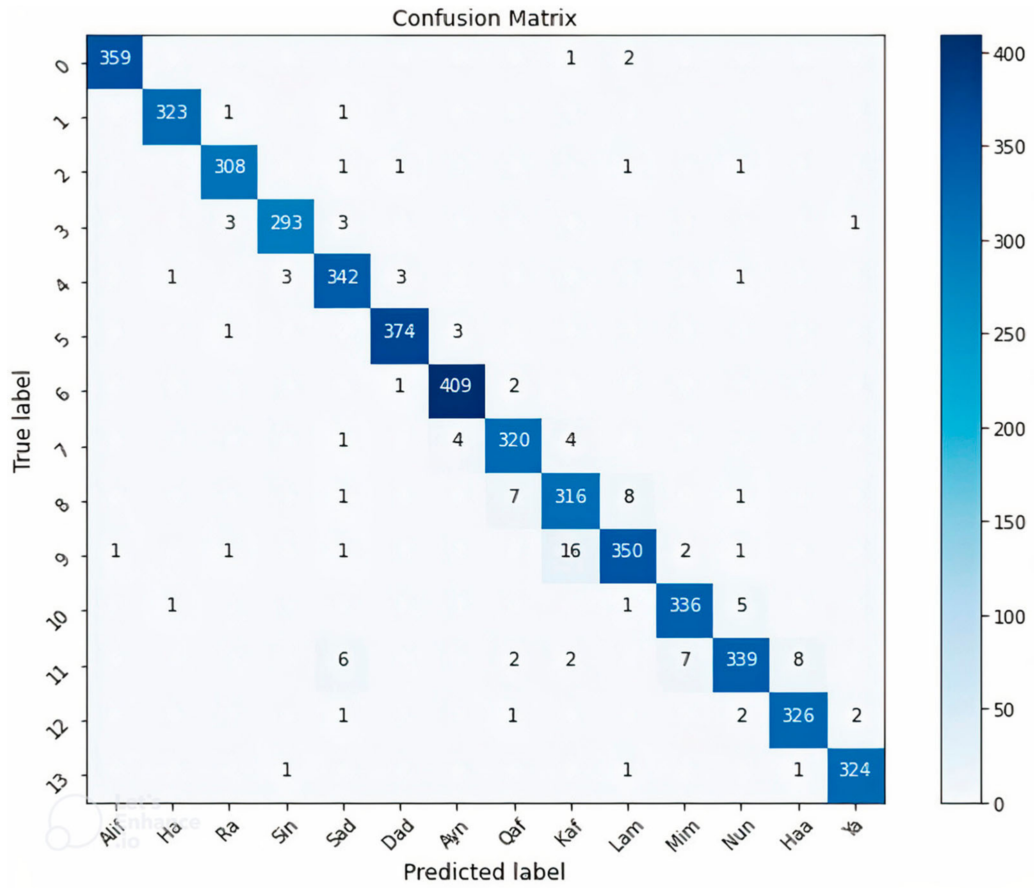

Figure 10. Confusion matrix of the proposed QSLRS-CNN model using RMU.

Qur'anic Surahs via the use of the SMOTE oversampling method. The QSLRS-CNN model achieves an accuracy of 98.66% using RMU. However, the proposed model is limited to images of fixed gestures that represent the discontinuous letters at the beginning of the Qur'anic Surahs. In addition, the treatment of isolated Qur'anic words or texts is not addressed.

In our future work, we are going to concentrate on testing the QSLRS-CNN on other datasets and looking into how well recurrent neural networks (RNN) and long short-term memory (LSTM) perform for the job. There are several ArSLs that employ the same alphabets, and Arabic-speaking nations are home to these languages. These kinds of differences in ArSLs may serve as a barrier to communication. Therefore, the future study will also involve using transfer learning in order to construct an improved ArSL deep learning model that is compatible with ArSL's variants. Those who communicate in ArSL may benefit from the use of this approach in order to circumvent the difficulties they have in doing so. This technique could make it easier for those who communicate using ArSL; we concentrate on the categorization of the alphabet letters that represent the first Surahs of the Quran in QSL. Using deep learning methods, work should be done to construct a model that can be used to convert the meanings of the Holy Quran into sign language so that it may be accessed by deaf people.

Table 7. Comparison of the results obtained by the proposed model and other previous models.

Methods		Accuracy Training (%)	Accuracy Testing (%)	No. of Samples	Training Time (min)
Ghazanfar et al. [10]		97.6	97.1	50,000	486.0
Alani and Cosma [12]		98.80	97.29	54,049	141.9
Proposed	CNN	98.05	97.13	24,137	20.16
Model	RMO	98.37	97.36	24,137	25.1
	RMU	98.66	97.52		21.18
	SMOTE	98.31	97.67		30.56

Conclusions

The QSLRS-CNN model is proposed based on the Arabic sign recognition system, which uses a portion of the ArSL2018 dataset. Our dataset contains 24,137 images of the ArSL alphabet created by more than 40 people, with 14 letters representing the openings of the Qur'anic Surahs. The QSLRS-CNN model has 98.05 percent training accuracy and 97.13 percent testing accuracy. Unbalanced data and the necessity for adequate samples from each class to train and assess deep learning models are underlined. Results reveal SMOTE's efficacy on the ArSL2018 dataset to increase the total training and testing accuracy to 98.31% and 97.67%, respectively. In addition, the proposed model achieves training and testing accuracy of

98.66% and 97.52% using the random minority undersampling (RMU) and 98.37% and 97.36% using the random minority oversampling (RMO), respectively. Therefore, the proposed model aims to recognize the movements of the QSL by recognizing the hand gestures that refer to the dashed Qur'anic letters in order to help the deaf and dumb learn their Islamic rituals. As a result, the experimental results demonstrate that the proposed model performs better than the other existing models.

Acknowledgment

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University.

Disclosure statement

No potential conflict of interest was reported by the author (s).

Funding

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through large group Research Project under grant number [RGP2/246/44].

Notes on contributors

Hany A. AbdElghfar received his B.Sc. and M.Sc. degrees from Systems and Computers Engineering, Faculty of Engineering, Al- Azhar University in Cairo, Egypt. His research interests include, Artificial intelligent, pattern recognition, machine learning, Deep Learning, E-Learning, Intelligence Systems and Computer Vision IOT systems. He is having 15 years of teaching and research experience at various reputed Universities of Egypt.

Abdelmoty M. Ahmed received his B.Sc., M.Sc. and PhD degrees. His research interests include Digital image processing, Artificial intelligent, pattern recognition, Human Computer Interaction, Computer Graphics, machine learning, Deep Learning, E-Learning, Intelligence Systems Engineering, Computer Vision and IOT systems, he is senior lecturer in computer engineering department at College of Computer Science, King Khalid University, Abha, Saudi Arabia, he is also interested in researching the technical fields that serve deaf and dumb and also works in the automatic translation of the Arabic Sign Language. He is having 20 years of teaching and research experience at various reputed Universities of Egypt and Saudi Arabia. He has published more research articles in reputed SCI and scopus indexed journals and conferences.

Ali A. Alani received his B.Sc. Degree in computer sciences from Diyala University, Diyala, Iraq in 2006 and M.Sc. Degree in Information Technology from Universiti Tenaga Nasional, Selangor, Malaysia in 2014. Recently, he is working as Assistant Lecturer in Department of computer sciences in university of Diyala, Diyala, Iraq. His research interests include Big data, Machine learning, Deep Learning and Computer vision.

Hammam M. AbdElaal received his B.Sc. and M.Sc. degrees in computers & systems engineering from faculty of engineering, Al-Azhar University, Cairo, in 2005 and 2016, respectively, and the Ph.D. degree in computer engineering from the Computer Engineering Department, Faculty of Engineering, Minia University, Egypt in 2020. He has been a Doctor (Lecturer) at faculty of computer and information, Luxor University. His main areas of research interest are Machine learning Techniques, supervised Learning algorithms, Natural language processing, and Data Mining.

Belgacem Bouallegue received his B.Sc. and M.Sc. degrees from University of Monastir, Tunisia, and Ph.D. from Graduate School of Engineering Science and Technology, University of Southern Brittany in in Lorient, France with the cooperation of University of Monastir, Tunisia. He is currently an Assistant Professor in Department of Computer Engineering at College of Computer Science, King Khalid University, Saudi Arabia. His research interests include Integrated System Design, Fault Tolerance, HW/SW Co-design, Parallel Computers, Embedded Systems and IoT, Network on Chip NoC, AI, IPs and MPSoCs, Machine Learning, Deep Learning, Wireless Sensor Networks Security, and Cryptography. He is working in collaboration with Lab-STICC Laboratory, Lorient, France and LIP6, Computer Science Research Laboratory, PARIS Cedex 05, France.

Mahmoud M. Khattab received his B.Sc. (2005) and M.Sc. (2009) degrees in computer science from faculty of computers and information, Menofiya University, Egypt. He earned his Ph.D. (2022) degree in computer science from kulliyyah (faculty) of information and communication technology, International Islamic University Malaysia (IIUM), Kuala Lumpur, Malaysia. The area of his research interest lies in super-resolution, image processing, pattern recognition, artificial intelligent, and computer vision. He is a lecturer in computer science department at King Khalid University (KKU), Saudi Arabia.

Hassan A. Youness received the B.Sc. and M.Sc. degrees from Assiut University, Assiut, Egypt, and the Ph.D. degree from the Graduate School of Information Science and Technology, Osaka University, Japan, with the cooperation of Ain Shams University, Egypt. He worked for IBM Company and Mentor Graphics, Egypt. He is currently Professor with Minia University, and also the Chairman of the Computers and Systems Engineering Department. His research interests include integrated system design, fault tolerance, HW/SW co-design, parallel computers, embedded systems, GPGPU, APU and MPSoCs, and homogeneous/heterogeneous systems.

ORCID

Abdelmoty M. Ahmed http://orcid.org/0000-0002-3379-7314
Belgacem Bouallegue http://orcid.org/0000-0001-7292-6345

References

[1] Ahmed AM, Abo Alez R, Tharwat G, et al. Arabic sign language intelligent translator. The Imaging Sci J. 2020;68:11–23.
[2] Mohandes M. Arabic sign language recognition. in In International conference of imaging science, systems, and technology, Las Vegas, Nevada, USA, 2001, pp. 753–762.

[3] Ahmed AM, Alez RA, Taha M, et al. Automatic translation of Arabic sign to Arabic text (ATASAT) system. J Comput Sci Inf Technol. 2016;6:109–122.

[4] Tharwat G, Ahmed AM, Bouallegue B. Arabic sign language recognition system for alphabets using machine learning techniques. J Electr Comput Eng. 2021;2021:1–17.

[5] Luqman H, Mahmoud SA. Transform-based Arabic sign language recognition. Procedia Comput Sci. 2017;117:2–9.

[6] Alzohairi R, Alghonaim R, Alshehri W, et al. Image based arabic sign language recognition system. International Journal of Advanced Computer Science and Applications (IJACSA). 2018;9:185–194.

[7] Kamruzzaman M. Arabic sign language recognition and generating Arabic speech using convolutional neural network. Wirel Commun Mob Comput. 2020;2020:1–9.

[8] ElBadawy M, Elons A, Shedeed HA, et al. Arabic sign language recognition with 3d convolutional neural networks. in 2017 Eighth international conference on intelligent computing and information systems (ICICIS), 2017, pp. 66–71.

[9] Mohamed MM. Automatic system for Arabic sign language recognition and translation to spoken one. Int J. 2020;9:7140–7148.

[10] Latif G, Mohammad N, AlKhalaf R, et al. An automatic Arabic sign language recognition system based on deep CNN: an assistive system for the deaf and hard of hearing. Int J Comput Digital Syst. 2020;9:715–724.

[11] Latif G, Mohammad N, Alghazo J, et al. Arasl: Arabic alphabets sign language dataset. Data Brief. 2019;23:103777.

[12] Alani AA, Cosma G. ArSL-CNN: a convolutional neural network for Arabic sign language gesture recognition. Indonesian J Electr Eng Comput Sci. 2021;22:1096–1107.

[13] Saleh Y, Issa G. Arabic sign language recognition through deep neural networks fine-tuning. International Journal of Online and Biomedical Engineering (iJOE). 2020;16:71–83.

[14] Shahin A, Almotairi S. Automated Arabic sign language recognition system based on deep transfer learning. IJCSNS Int J Comput Sci Netw Secur. 2019;19:144–152.

[15] Abeje BT, Salau AO, Mengistu AD, et al. Ethiopian sign language recognition using deep convolutional neural network. Multimed Tools Appl. 2022;81:29027–29043.

[16] Tamiru NK, Tekeba M, Salau AO. Recognition of Amharic sign language with Amharic alphabet signs using ANN and SVM. Vis Comput. 2022;38:1703–1718.

[17] Ibrahim NB, Selim MM, Zayed HH. An automatic arabic sign language recognition system (ArSLRS). J King Saud University-Computer Inf Sci. 2018;30:470–477.

[18] Nagi J, Ducatelle F, Di Caro GA, et al. Max-pooling convolutional neural networks for vision-based hand gesture recognition. in 2011 IEEE international conference on signal and image processing applications (ICSIPA), Kuala Lumpur, Malaysia, 2011, pp. 342–347.

[19] Hayani S, Benaddy M, El Meslouhi O, et al. Arab sign language recognition with convolutional neural networks. in 2019 International Conference of Computer Science and Renewable Energies (ICCSRE), Agadir, Morocco, 2019, pp. 1–4.

[20] Ahmad NB, Ishak MK, Alias UF, et al. An approach for e-learning data analytics using SOM clustering. Int J Adv Soft Comput Its Appl. 2015;7:94–112.

[21] Abdelaal HM, Ahmed AM, Ghribi W, et al. Knowledge discovery in the Hadith according to the reliability and memory of the reporters using machine learning techniques. IEEE Access. 2019;7:157741–157755.

[22] Smiti S, Soui M. Bankruptcy prediction using deep learning approach based on borderline SMOTE. Inf Syst Front. 2020;22:1067–1083.

[23] Khattab MM, Zeki AM, Alwan AA, et al. Regularized multiframe Super-Resolution image reconstruction using linear and nonlinear filters. J Electr Comput Eng. 2021;2021:1–16.

[24] Khattab MM, Zeki AM, Alwan AA, et al. A hybrid regularization-Based multi-Frame super-Resolution using Bayesian framework. Comput Syst Sci Eng. 2022;44:35–54.