JKAU: Comp. IT. Sci., **Vol. 13** No. 2, pp: 58 – 66 (2024 A.D.)
DOI: 10.4197/Comp.13-2.4

Investigating Active Learning based on Dynamic Data Selection techniques for Image Classification

Salma Kammoun Jarraya
Department of Computer Science, King Abdulaziz University
Jeddah, KSA
smohamad1@kau.edu.sa

***Abstract*— This paper explores the efficacy of Active Learning (AL) techniques, specifically focusing on Dynamic Data Selection (DDS), for improving image classification tasks. AL is a machine learning paradigm that enables the automatic selection of the most informative data samples for annotation, thereby reducing the annotation burden and enhancing model performance. In this study, we investigate the integration of DDS techniques with AL strategies to iteratively select the most informative image samples for model training. We use a fine-tuned VGG16 as the underlying classification model due to their effectiveness in image analysis tasks. Our experimental evaluation involves comparing the performance of fine-tuned VGG16 trained with three AL-based DDS techniques on Arabic sign language dataset. We analyze various DDS strategies, including Random selection, Entropy-based selection, and margin selection to determine their impact on model accuracy and annotation efficiency. The results of our study demonstrate the effectiveness of margin selection method-based AL approach in improving the performance of recognition of 32 hand gestures for Arabic sign language (95.3 %) while minimizing the annotation effort.**

Keywords—Computer Vision, Active Learning, CNN

I. INTRODUCTION

In recent years, in response to the exponential proliferation of images across a variety of fields, the importance of image classification approaches that are both efficient and accurate has been brought to light. Traditional methods frequently rely on huge amounts of labeled data, which can be challenging to gather and require significant time and resources. Active learning is a subset of machine learning that offers a promising solution to this challenge. It allows for the selection of the data samples with the most relevant information for annotation, reducing the amount of labeling effort required while maintaining classification performance. This study aims to investigate the use of active learning in the context of image classification, with a particular emphasis on dynamic data selection strategies. The selection of data samples to label in active learning methods is based on the expected enhancement of those samples to increase the classifier's performance. The active learning algorithm can iteratively refine its understanding of the data distribution and prioritize the annotation of samples, which is most beneficial for improving classification accuracy. Dynamic data selection techniques enhance this process by adapting the selection criteria over time. This allows the algorithm to improve classification accuracy. The primary purpose of this study is to explore the efficacy of dynamic data selection strategies based on their effectiveness. When picking informative data samples for annotation, we aim to evaluate and contrast the effectiveness of various active learning procedures, such as margin selection, entropy-based selection, and random selection. This study provides insights into the strengths and limitations of dynamic data selection approaches in active learning for hand gesture recognition. These insights will be provided through experiments done on an Arabic sign language dataset. A few sign language (SL) recognition algorithms recognize Arabic sign language by utilizing deep learning techniques. Two ways that can be utilized with SL recognition systems are those based on images and those based on sensors. In order to recognize hand gestures utilizing sensorbased techniques, the user must wear instrument gloves fitted with sensors throughout their entire body. This technology requires the interface of a large number of sensors with a glove to collect gestures through sensor data, which is then processed for tasks including gesture recognition and translation. Several difficulties are associated with sensor-based techniques, even though they are reliable and accurate. One of these drawbacks is that the signer's gloves may be uncomfortable when they contain sensors, wires, and other materials [1]. Consequently, researchers made a great effort to use a dataset that relied heavily on images. In general, this study contributes to the research that has been done on active learning and image classification by highlighting the potential benefits of dynamic data selection strategies in terms of improving classification performance and decreasing labeling costs.

This paper is structured as follows: Section II provides a background and Literature Review on Active learning and sign language (SL)recognition, Section III describes the proposed approach, Section IV presents experimental results, and Section V discusses the findings, conclude the paper and suggest future work.

II. BACKGROUND AND LITERATURE REVIEW

Training high performance neural networks models require large amount of data, this has many challenges such as the difficulty to gather huge datasets while ensuring its quality. Also, another challenge is the time and resources complexity, since the time of training increases as the size of data increases, as well as the needed resources. Active learning tries to solve this problem

by applying algorithms that can chooses the most informative unlabelled data samples (to be labelled) in order to reach the most possible optimal solution. After that, these samples are provided to an oracle, such as a human, so they can label these data. There are many approaches for extracting informative samples, one of them is entropy-based approach proposed by [2].

The approach proposed in [2] consists of four components: (1). A set of labelled examples used for training. (2). Unlabelled examples. (3). An oracle, such as a human, to label the unlabelled examples. (4). A methodology used to choose the informative examples to request the correct labels for them. This is an iterative process so every time the chosen unlabelled data (called Most Informative Unlabeled Point (MIUP)) are labelled, new examples are provided to the oracle [2].

For step 4, there are many methods that can be used such as choosing the most confused samples based on specific measurements, for example the point that is closest to the surface of decision separating two classes. Another possible method is calculating the similarities between the sample images [2].

SL is the official language used by deaf and hard hearing people to communicate with others, this language consists of alphabets, numbers, and words. Just like spoken languages, we have many types of sign languages such as Arabic sign language (ArSL). Any sign language may consists of two components: manual and nonmanual components, the first one is related to the hands' positions and movements, while the second one is related to more complex aspects such as the movements of body and the facial expressions [3, 4].

Sign Language Recognition (SLR) is the system that recognizes sign languages automatically to make communication between deaf and non-deaf people much easier. Arabic Sign Language Recognition (ArSLR) is the process of automatic recognition of Arabic Sign Language [5].

The problem of ArSLR can be categorized into three different levels: gestures levels (alphabets level), isolated gestures levels (words level), and continues gestures (sentences level) which is the most advanced one [6].

In the perspective of signs data sources, there are mainly two approaches: sensor-based and image-based, the first one focuses on capturing the gestures from sensors such as wearable devices, e.g. gloves. On the other hand, image-based is more simple as it focuses on capturing gestures using cameras(photos and/or videos)[5].

There are many techniques that are used in this domain, most of them are the same ones used in Speech Recognition (SR) since they are two similar problems. However, nowadays new techniques were improved specially for this type of problem. The types of used techniques depend on the chosen level and approach, for example, sensors-based approach needs different methods for data collection and features extraction from image-based approach. In the following sub-section, some of the different methods used in the different stages of SLR will be discusses [5].

For sensors-based approach, one of the used methods for features extraction is window-based statistical, where a window of size w, the means and standard deviations are calculated.

For image-based approach, assuming the usage of videos instead of separated images, motion detection technique is needed, a possible method that can be used is pixel-based difference. After that, features can be extracted using different methods such as 2D Discrete Cosine Transform (DCT), using a set of features depending on some criteria such as the shape of the gesture, or using a neural network approach such as R-CNN. For images datasets, the process is much simpler since there is no need to detect motion nor splitting the videos into multiple images, instead, features extraction can be immediately be implemented [5, 4, 6].

After features extraction phase, classification step takes place. Therearealsomanymethodsusedforitsuchask-nearestneighbors (KNN), improved KNN [7] and HMM, or using a deep learning based approach such as R-CNN or CNN [5, 4]

An end-to-end deep learning model can be used for both features extractions and classification at the same time. One popular neural network architecture that can be used is (Convolutional Neural Networks) which works very well with images. CNN can be considered that it is a form of Artificial Neural Network (ANN). ANN is inspired by humans' neurons, it consists of a large set of computational nodes that receive inputs then learn from them in an interconnected and distributed manner to reach an optimal output. Each node receives input from the layer before it then perform specific operations depending on the application, where the goal is to update the network weights to minimize the loss function which is placed in the last layer and calculated using the input's classes and the network's predictions [8].

Learning can be done using two different approaches, namely: supervised and unsupervised learning. In supervised learning, the input are associated with labels defined as vectors, where the goal is to reduce the error of classification in mapping between the input and output labels, in the case of SLR, this methodology is used. Unsupervised learning's input don't have any labels associated with them, thus, the main focus is reducing the cost function [8].

CNN differs from traditional ANN in which it is usually used in images applications, such as SLR, especially when the images have a large dimensions and multiple channels. Any CNN architecture consists mainly of three possible layers: convolutional layers, fully connected layers, and pooling layers. . In general, each layer performs different operations [8].

The convolutional layer uses kernels for learning, this kernel passes through the whole image including the depth, then the output will be a set of features called "activation map", the scalar product of every value in the kernel is calculated. Multiple kernels can be found then all their outputs, which are activation maps, are stacked together. To avoid the high complexity of the neural network, every node, i.e., neuron, in the convolutional layer will be connected to only a part of the volume of the input, this is called "receptive field size", for example, if a given image is the input and it has a size of 64*64*3, the receptive field size might be set to 6*6*3. Different hyperparameters are also included in the convolutional layer which are: depth, stride, and paddding [8].

Pooling layers are another type of layer in CNN. Its main goal is reducing the representation's dimensionality to decrease the complexity. There are three types of pooling layers: max pooling, general pooling, and the average pooling. Given the activation maps produced by the convolutional layers, each type applies different functions on every activation map: max function,

common operations, e.g., L1/L2-normalisation, and average function [8].

The last layer is fully-connected layer, as its name suggests, it is a type of layers where all neurons are connected in a direct manner to the two adjust, i.e, before and after, layers [8].

There is no specific way to construct a good CNN architecture. However, observing and reading literature work leads to understanding some common ways of constructing them so they produce good results. For example, in a traditional architecture, every similar layers are stacked together starting from convolutional layers, pooling layers and then fully connected layers. Another common method is to set the input layers to be divisible by number two, for example 96*96, 224*224...etc, it is recommended that images sizes don't exceed 128*128 to reduce the complexity [8].

Arabic sign language recognition has recently gained popularity as an active area of study. Systems that recognize sign language to transform gestures into text or speech can aid in communication with those who are deaf. Approaches for recognizing sign languages rely on one of the two methods for identifying gestures, theseare sensor-basedandimage-based recognitionsystems. Fig. 1. shows sign language recognition approaches [1].

To recognize sign language gestures, image-based systems use images or videos combined with machine-learning algorithms. These systems can be divided into two categories. The first relies on utilizing gloves with visual markers, such as colored gloves, to recognize hand gestures. The second category is dependent on images that capture sign-language hand gestures (bare hand) [1].

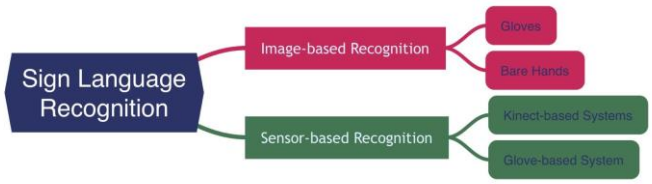

Fig. 1. Classification of Sign language recognition approaches. [1]

The authors of[9] provided an Arabic sign language dataset containing 8K videos of 20 signs performed by different participants. They then proposed a new approach for video classification and recognition that uses a combination of CNN and Recurrent Neural Network (RNN). The core concept of the proposed approach is to train two independent CNNs on different sets of data using the same architecture. The overall prediction was produced using RNN, which was also utilized to determine the relationship between the sequences of images.

In order to further the fields of related studies, such as sign language to sound approaches, translation of Arabic sign language to other languages, and many more areas, the work [10] introduced a model using fine-tuning with deep learning to recognize Arabic Sign Language. The concept of fine-tuning these images is to avoid the requirement of collecting a large dataset of images for training. The proposed model that has been presented uses CNN for image recognition and classification while requiring a smaller training dataset and achieving a higher level of accuracy. The work presented starts with the most recent pre-trained network models, Resnet152 and VGG-16, and then applies fine-tuning using the ArSL dataset. To reduce the imbalance caused by inconsistent class sizes, random under-sampling was applied to the dataset, which resulted in a decrease in dataset size from 54K to 25K.

A 3D Convolutional Neural Network (CNN) was employed by [11] to develop an ArSL recognition system based on 25 sign images. The system uses a video stream and receives a normalized depth input. From the input, the architecture extracts image pixel features. The softmax layer then functioned as a feature classifier. The accuracy of these findings is 85%. This accuracy can be improved by including more training samples in the dataset to account for the diversity of the signer and environment. The results of the experiments demonstrated the effectiveness of the 3D deep architecture.

In [12], author presented a visual SLRs that automatically converted isolated Arabic word signs into text. The proposed method is a signer-independent system that uses only one camera and does not require a signer to use gloves or markers. Hand segmentation was performed using a dynamic skin detector based on the color of the face. Segmented skin blobs were then employed by the head to recognize and track the hands. As 83% of the words in the sample had distinct occlusion states, the system demonstrated its ability to perform well in all occlusion conditions. The classification stage involves a Euclidean distance classifier.

In sensor-based systems, the appearance of the hand is detected by sensors, which allows for the detection of sign language. Two systems are considered for this system: glove-based and kinectbased systems. Electromechanical devices are used in glove-based systems for identifying hand gestures. Kinect sensors are input devices for Xbox games to interact with video games, and are utilized to recognize sign language gestures[1].

Authors in [5] compared two different recognition methods for continuous ArSLR, including a modified k-nearest neighbor (KNN) method that works with sequential data and hidden Markov model (HMM) methods based on two toolkits. Two new ArSL datasets comprising 40 Arabic sentences were gathered using a camera and Polhemus G4 motion tracker. The experiment demonstratedthattheclassificationaccuraciesforsignsentencesrecorded using a motion tracker and sensor gloves were similar. When comparing the computational time required for classification, the modified KNN solution was less efficient than the HMMs.

Authors in [13] developed a novel technology for ArSLR utilizing two Leap Motion Controllers (LMC). A Gaussian Mixture Model (GMM) and a Bayesian classifier were used to analyze the extracted features of the two LMCs. An evidence-based method, namely, the Dempster Shafer (DS) technique, was used to integrate the output from each separate Bayesian classifier. The suggested approach demonstrates a variety of benefits, including flexibility in the number of GMM components, the type of evidence utilized for the combination framework, and the types of features that may be retrieved from and applied to LMCs. In many sign and gesture detection contexts, this study provides new insights and research prospects for merging numerous interfaces.

Author in[1] works on automatic ArSL alphabet recognition using an sensor-based technique. Different visual descriptors were investigated in order to construct an accurate ArSL alphabet

detector. A one-versus-all soft-margin SVM is used in the proposed approach to extract the Histogram of Oriented Gradients (HOG) descriptor. The results showed that the HOG descriptor performed better than the other descriptors. The proposed study implemented an ArSL gesture system trained by a One-Versus-All SVM using HOG descriptors. Ultimately, 63.5% of Arabic alphabet gestures were successfully recognized by the system.

Table I summarizes related works by focusing in data collection method and test accuracy.

Based on the information provided in Table I , we can compare the approaches used for recognizing ArSL based on several factors: the dataset used in the research is an important factor to consider when comparing different approaches. In Table I, we can see that some researchers built their dataset from scratch, while others used existing datasets, such as ArSL. Building a dataset from scratch can provide more control over the data and ensure that it is relevant to the research question, but it can be time-consuming and requires a significant amount of effort. Using an existing dataset can save time and effort.

The dataset size is an important factor in machine learning and can affect the performance of the model. In Table I, we can see that the dataset size ranges from 200 to 25,000 samples. A larger dataset can provide more information for the model to learn, but it can also require more resources and time for training. The spatial or pixel features used for recognition are also important factors to consider. Researchers have used a variety of features, such as image pixels, FFT, and geometric parameters. Each feature has its own strengths and weaknesses, and the choice of feature may depend on the specific research question and the available resources.

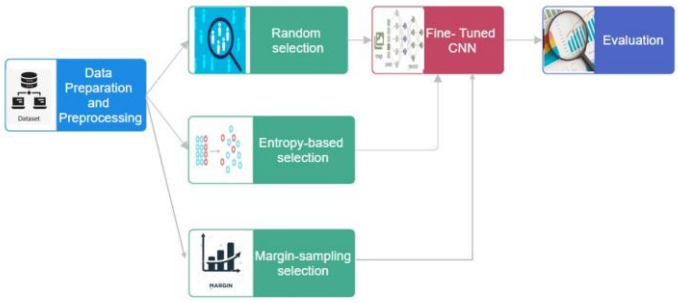

Fig. 2. Proposed Approach

Overall, each approach has its own strengths and weaknesses, and the choice of approach may depend on the specific research question and the available resources. In general, it appears that the use of pre-trained models and larger datasets can lead to higher accuracy levels; therefore, we consider these two factors in this study.

III. PROPOSED APPROACH

The proposed approach (Fig. 2) focuses on recognizing Arabic Sign Language using fine tuned CNN and active learning based on three different methods for extracting the most uncertain unlabelled images.

A. Data Preparation and Preprocessing

All images are RGB images and set to a size of 64*64, then they are divided into two sets: training and testing with a ratio of 7:3, so the total number of training images is 43239, and the total number of testing set is 10810, during training the validation set will be defined and constructed based on some conditions, more details will be discusses in Section 4.

B. Active Learning

Three methods for choosing the unlabelled samples are implemented in this research which are: random selection, entropybased selection, and margin sampling selection.

Random selection is a very simple method for selecting samples, given a set of samples x, selects k number of samples randomly

TABLE I. SUMMARY OF RELATED WORK

Ref	Dataset	Method of Collecting Dataset	Dataset size	Features	Classifier	Accuracy
[9]	Built from scratch.	Bare hand	8K	Spatial features	CNN and RNN	92%
[10]	ArSl	Bare hand	25K	Images pixel	Pre- trained VGG Pre-trained Resnet152	99.4% 99.6%
[11]	Built from scratch	Bare hand	200	Images pixels	3D CNN	85%

[12]	ArSl	Bare hand	450	Segmented image	Euclidean distance	83%
[5]	Bulit from scratch	Kinect sensor	2K	Geometric parameter	- Bayesian	88%
[13]	two datasets collected from different device	Gloves sensor	800 for each dataset	Threshold of image difference	Modified KNN Hidden Markov (HMM)	94.5%
[1]	ArSl	Glove sensor	Not mentioned	FFT (Fast Fourier transform)	SVM and HOG	63.5%

$$s(\theta,x)= -\sum_{y\in Y} p_\theta(y \mid x)\log p_\theta(y \mid x) \quad (1)$$

Entropy-based selection selects the most uncertain samples based on the highest values of entropy using the entropy Equation 1 [14].

$$f(q_j)= \sum_{i=1}^{m} \alpha_i y_i K(x_i,q_j)+b \quad (2)$$

Margin-sampling selection selects samples based on the properties of SVM, since the distance between each data point (sample) and the separable hyperplane indicates the confidence of the classifier in classifying that data point, i.e., if the point is very near to the hyperplane it means the classifier is least confident about it. In margin sampling, samples are selected which have the minimum distance to the hyperplane, since they are considered the most uncertain ones. This can be shown in two equations given a binary classification problem, where Equation 2 indicates the distance between any data point and the hyperplane, and Equation 3 indicates the k samples selected for the training [15].

$$k' = \underset{q_j\in U}{arg\ min}|f(q_j)|. \quad (3)$$

C. Model Architecture

An image is fed into a convolutional neural network, which then processes it through a series of layers that include convolution, pooling, and other fully connected layers to produce an output that represents only one of the images' possible classes. Well known CNN architectures that were used in our experiment on the Arabic Sign Language dataset are briefly introduced in this section. The VGGNet model architecture is one of the previously released models with training parameters that has received a lot of attention in recent years. The VGGNet model architecture was able to have fewer layers than the published state-of-the-art architecture, while still producing remarkably excellent results with an error of 7.3%. The used (16 layers) model's structure is shown in Fig. 3. The convolutional layer parameters are written as "conv (receptive field size)–(number of channels)" [10].

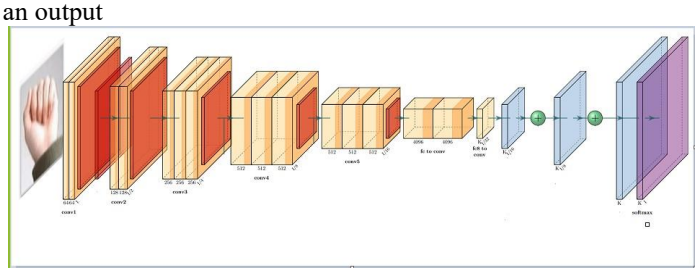

Fig. 3. Model structure

The proposed approach loads the pre-trained VGG16 model from Keras library. The weight parameter was set to 'imagenet' to use the pre-trained weights for the model. We excluded fully connected layers at the top of the network, which were replaced by our own layers. The next step was to freeze the layers in the VGG16 model to a certain point. This was performed to prevent the weights in these layers from being updated during the training process. For that, we freeze all layers in the VGG16 model, except for the last five layers.

Subsequently, new layers were added on top of the pre-trained VGG16 model. This was performed using the sequential function from Keras. We added a pre-trained VGG16 model to the new model. Then, we added a dropout layer with a dropout rate of 0.2 to prevent overfitting. A batch normalization layer was added to normalize the inputs to the next layer. Finally, we added a flattened layer to flatten the output of the VGG16 model, a dense layer with 32 units, and a softmax activation function to output the final classification probabilities for the input image. The final step is to compile the model. This is performed using the compile method, which considers the loss function, optimizer, and metrics as the arguments. In our approach, we used the categorical cross-entropy loss function, Adam optimizer, and accuracy metric to evaluate the performance of the model during training.

Overall, we fine-tuned a pre-trained VGG16 model on a new classification task. The pre-trained weights of the VGG16 model were used as a starting point, and new layers were added on top of the model to adapt it to the new task. The layers in the pre-trained model were frozen to a certain point to prevent the weights from being updated, and the new layers were trained on the new task. Finally, the model was compiled using a loss function, optimizer, and metrics to train the model and evaluate its performance.

IV. EXPERIMENTAL RESULTS

There are many scenarios of how to implement active learning, the one chosen in this research is pool-bases sampling, it works by selecting k random number of samples of the full training set, this will be the set to be fed into the model during training. After that, the validation set is constructed by copying all of the full training samples except the chosen k samples. In every iteration, new k number of samples are added to the training set, and therefor deleted from the validation set, this is done until the number of chosen samples reach a pre-defined parameter called max required samples, in this experiment, k issetto50andmaxrequiredsamples is set to 2700. The number of epochs is set to be 100, the batch size is 16, the optimizer is Adam, the loss function is categorical cross entropy, for classification, SoftMax activation function is used.

The used dataset is Arabic Alphabets Sign Language Dataset (ArASL) is named (ArSL2018) proposed in [16]. It consists of 54,040 images of 32 different signs and alphabets of Arabic sign languages, 40 individuals of different ages participated in creating this dataset, it is an open-source dataset available at https://data.mendeley.com/datasets/y7pckrw6z2/1. A sample of each alphabet of the dataset is shown in Fig. 4.

As mentioned in Section B, three different approaches for active learning are used: random selection, entropy-based selection, and margin-sampling selection. The results of each one of them are shown in Fig. 5.

Fig. 4. Samples of ArASL dataset

The plot of the loss function for over 2700 samples look like a wave (goes up and down). The horizontal axis of the plot represents the number of samples, and the vertical axis represents the value of the loss function. The peaks in the plot represent instances where the model predictions were far from the actual values, resulting in a high value of the loss function (in the case of entropy selection). The fact that the plot goes up and down indicates that the machine learning model adjusts its parameters during training to improve its predictions. As the model continues to learn from the data, the loss function decreases. Margin selection outperforms entropy and random selection in obtaining a lower value in the loss function.

As the model continued to learn from the data, the accuracy percentage gradually increased over time owing to variations in the data and the model's performance. The three selection methods showed an increase in performance. Margin selection and random selection have approximately similar performance but at the end the margin selection outperforms random selection by 2%. entropy selection performed the worst with 61.3% accuracy while margin sampling get 95.3% and random get 93.3%

Fig. 6 and Fig. 7 show samples of the dataset along with their real and predicted classes, indifferent background, using the model that is trained using margin sampling.

Using active learning in training CNN shows promising results in both efficiency of training and good results. However, from the presented experimental results , it is clear that different approaches used for uncertain samples selection play an important role in the results. On the other hand, margin-sampling selection and random selection produced good results, with a highest accuracy of 95.3% using the margin-sampling.

A strong point that this proposed approach has is that it uses only a sample of the dataset instead of the whole dataset, unlike most of the related papers. It takes advantages of two important ideas which are transfer learning and active learning to reach the best possible results in a reasonable time and used resources.

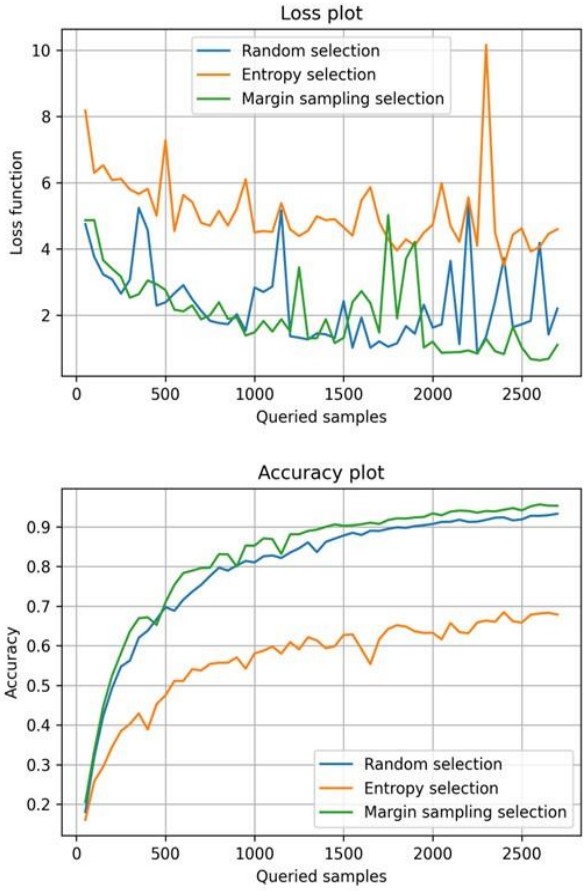

Fig. 5. Loss and accuracy results

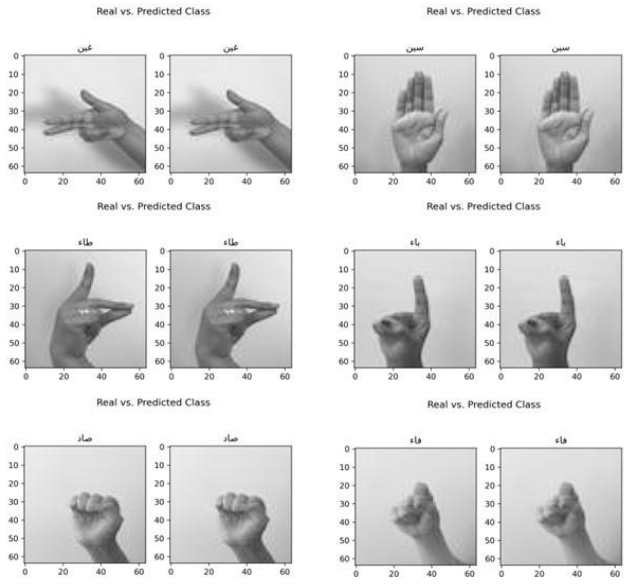

Fig. 6. Samples from the margin sampling's model prediction

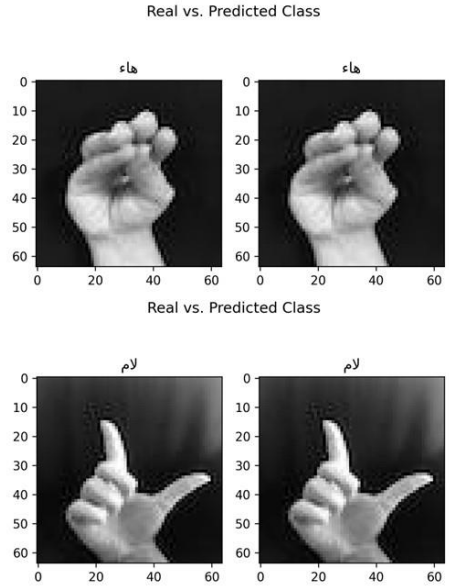

Fig. 7. Samples from the margin sampling's model prediction in different background

V. CONCLUSION

In conclusion, the proposed approach of recognizing Arabic Sign Language using CNN and active learning demonstrated promising results. By leveraging transfer learning and modifying the pretrained VGG16 model such as freezing some layers and adding additional ones, we were able to achieve optimal performance in sign language recognition. In addition, the use of active learning and uncertainty sampling techniques, specifically margin sampling and random sampling, proved to be effective in improving the accuracy of the model.

The results of the experiments showed that margin sampling selection provided the highest accuracy of 95.3%, followed closely by random selection, with an accuracy of 93.3%. However, entropy-based selection produced relatively poor results, with an accuracy of 63.3%. These results suggest that the margin sampling technique is particularly effective in selecting the most uncertain samples during training, leading to a better performance of the model in recognizing sign language. The use of CNNs and active learning techniques in sign language recognition has great potential to improve the accessibility of deaf and hard-of-hearing communities. With further research and development, this approach could lead to more accurate and efficient sign language recognition systems, which could greatly benefit the lives of individuals with Fig. 6. Samples from the margin sampling's model predict hearing impairment. In summary, the proposed approach of recognizing Arabic Sign Language using CNN and active learning, specifically by leveraging transfer learning and using uncertainty sampling

techniques, has shown promising results in improving the accuracy of the model.

Several research directions could lead to improvements in the performance and results of the model. One potential direction is to investigate the use of different deep learning architectures, such as ResNet or Inception, as alternatives to the model currently being used. By exploring alternative architectures, we could potentially find one that is better and that could improve the overall accuracy of the model. Another direction to consider is to explore different active learning strategies, such as uncertainty sampling or query by-committee. These strategies could help the model select the most informative samples from the training data and lead to better performance in the long run. The third direction is to implement the model in a real-time application, such as an app that can be used to assist deaf people in their daily lives. This requires the model to be optimized for speed and accuracy, which could lead to improvements in its overall performance. Finally, we examined how the VGG16 model could be used to identify sentences in the Arabic sign language. This would involve modifying the current model to handle more complex language structures, which could significantly impact the overall performance and accuracy of the model.

REFERENCES

[1] R. Alzohairi, R. Alghonaim, W. Alshehri, and S. Aloqeely, "Image based arabic sign language recognition system," International Journal of Advanced Computer Science and Applications, vol. 9, no. 3, 2018.

[2] A. Holub, P. Perona, and M. C. Burl, "Entropy-based active learning for object recognition," in 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 1–8, IEEE, 2008.

[3] H. Luqman and E.-S. M. El-Alfy, "Towards hybrid multimodal manual and non-manual arabic sign language recognition: Marsl database and pilot study," Electronics, vol. 10, no. 14, p. 1739, 2021.

[4] R. A. Alawwad, O. Bchir, and M. M. B. Ismail, "Arabic sign language recognition using faster r-cnn," International Journal of Advanced Computer Science and Applications, vol. 12, no. 3, 2021.

[5] M. Hassan, K. Assaleh, and T. Shanableh, "Multiple proposals for continuous arabic sign language recognition," Sensing and Imaging, vol. 20, pp. 1–23, 2019.

[6] G. Tharwat, A. M. Ahmed, and B. Bouallegue, "Arabic sign language recognition system for alphabets using machine learning techniques," Journal of Electrical and Computer Engineering, vol. 2021, pp. 1–17, 2021.

[7] N. Tubaiz, T. Shanableh, and K. Assaleh, "Glove-based continuous arabic sign language recognition in user-dependent

[8] mode," IEEE Transactions on Human-Machine Systems, vol. 45, no. 4, pp. 526–533, 2015.

[9] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," arXiv preprint arXiv:1511.08458, 2015.

[10] M. M. Balaha, S. El-Kady, H. M. Balaha, M. Salama, E. Emad, M. Hassan, and M. M. Saafan, "A vision-based deep learning approach for independent-users arabic sign languageinterpretation," MultimediaToolsandApplications, pp. 1–20, 2022.

[11] Y. Saleh and G. Issa, "Arabic sign language recognition through deep neural networks fine-tuning," 2020.

[12] M. ElBadawy, A. Elons, H. A. Shedeed, and M. Tolba, "Arabic sign language recognition with 3d convolutional neural networks," in2017Eighthinternationalconferenceonintelligentcomputingandinformation systems(ICICIS),pp.66–71, IEEE, 2017.

[13] N. B. Ibrahim, M. M. Selim, and H. H. Zayed, "An automatic arabic sign language recognition system (arslrs)," Journal of King Saud University-Computer and Information Sciences, vol. 30, no. 4, pp. 470–477, 2018.

[14] M. Deriche, S. O. Aliyu, and M. Mohandes, "An intelligent arabic sign language recognition system using a pair of lmcs with gmm based classification," IEEE Sensors Journal, vol. 19, no. 18, pp. 8067–8078, 2019.

[15] V.-L. Nguyen, M. H. Shaker, and E. Hu¨llermeier, "How to measure uncertainty in uncertainty sampling for active learning," Machine Learning, vol. 111, no. 1, pp. 89–122, 2022.

[16] J. Zhou and S. Sun, "Improved margin sampling for active learning," in Pattern Recognition: 6th Chinese Conference, CCPR 2014, Changsha, China, November 17-19, 2014. Proceedings, Part I 6, pp. 120–129, Springer, 2014.

[17] G. Latif, N. Mohammad, J. Alghazo, R. AlKhalaf, and R. AlKhalaf, "Arasl: Arabic alphabets sign language dataset," Data in brief, vol. 23, p. 103777, 2019.

التحقيق في التعلم النشط بناءً على تقنيات اختيار البيانات الديناميكية في تصنيف الصور

سلمى كمون الجراية[1]

قسم علوم الحاسبات، كلية الحاسبات وتقنية المعلومات، جامعة الملك عبد العزيز، جدة، المملكة العربية السعودية[1]

Smohamad1@kau.edu.sa

المستخلص. يستكشف هذا البحث فعالية تقنيات التعلم النشط(AL) ، مع التركيز بشكل خاص على اختيار البيانات الديناميكي (DDS)، لتحسين مهام تصنيف الصور. يعد التعلم النشط نموذجًا من نماذج التعلم الآلي الذي يمكّن من اختيار العينات الأكثر إفادة للتعليق بشكل تلقائي، مما يقلل من عبء التعليق ويعزز أداء النموذج. في هذه الدراسة، نقوم بتحقيق دمج تكامل تقنيات DDS مع استراتيجيات التعلم النشط لاختيار العينات الأكثر إفادة من الصور بشكل تكراري لتدريب النموذج. نستخدم نموذج VGG16 المعدل كالنموذج الأساسي للتصنيف بسبب فعاليته في مهام تحليل الصور. يتضمن تقييمنا التجريبي مقارنة أداء نموذج VGG16 المعدل باستخدام ثلاث تقنيات DDS قائمة على التعلم النشط على مجموعة بيانات لغة الإشارة العربية. نقوم بتحليل استراتيجيات DDS المختلفة، بما في ذلك الاختيار العشوائي، والاختيار المعتمد على الإنتروبيا، واختيار الهامش لتحديد تأثيرها على دقة النموذج وكفاءة التعليق. تظهر نتائج دراستنا فعالية نهج التعلم النشط المعتمد على طريقة اختيار الهامش في تحسين أداء التعرف على 32 إشارة يد في لغة الإشارة العربية (٩٥.٣%) مع تقليل جهد التعليق.

الكلمات المفتاحيةـ الرؤية بالحاسوب، التعلم النشط، الشبكات العصبية التلافيفية