This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2022.Doi Number

Advancing Multilingual Handwritten Numeral Recognition with Attention-driven Transfer Learning

AMIRREZA FATEH[1], REZA TAHMASBI BIRGANI[2], MANSOOR FATEH[2], AND VAHID ABOLGHASEMI[3]

[1] School of Computer Engineering of Iran University of Science and Technology (IUST), Tehran, Iran
[2] Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran
[3] School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, UK

Corresponding author: Vahid Abolghasemi (e-mail: v.abolghasemi@essex.ac.uk), Mansoor Fateh (e-mail: mansoor_fateh@shahroodut.ac.ir).

ABSTRACT As deep learning continues to evolve, we have observed huge breakthroughs in the fields of medical imaging, video and frame generation, optical character recognition (OCR), and other domains. In the field of data analysis and document processing, the recognition of handwritten numerals plays a crucial role. This work has led to remarkable changes in OCR, historical handwritten document analysis, and postal automation. In this study, we present a novel framework to overcome this challenge, going beyond digit recognition in only one language. Unlike common methods that focus on a limited set of languages, our method provides a comprehensive solution for recognition of handwritten digit images in 12 different languages. These specific languages are chosen because most of them have fairly distant representations in latent space. We utilize transfer learning, as it reduces the computational cost and maintains the quality of enhanced images and the models' recognition accuracy. Another strength of our approach is the innovative attention-based module called the MRA module. Our experiments confirm that by applying this module, major progress is made in both image quality and the accuracy of handwritten digit recognition. Notably, we reached high precisions, surpassing nearly 2% improvement in specific languages compared to earlier techniques. In this work, we present a robust and cost-effective approach that handles multilingual handwritten numeral recognition across a wide range of languages. The code and further implementation details are available at https://github.com/CVLab-SHUT/HandWrittenDigitRecognition

INDEX TERMS Deep learning, Transfer learning, Multilingual, Handwritten numeral recognition

I. INTRODUCTION

Handwritten numeral recognition plays a crucial role in the fields of image processing and computer vision, making significant contributions to diverse applications. This technology is widely used in a wide range of domains, including bank check processing, postal code recognition, and the analysis of historical handwritten documents [1-5]. Accurate recognition of handwritten numerals is crucial in delicate systems like medical records and biometric identification, where mistakes can lead to serious consequences [6-10]. Nevertheless, handwritten digit recognition poses several challenges due to the diversity in shapes, sizes, and variations present in different handwriting styles. These differences can be further influenced by factors such as the writing instrument used and the speed at which the individual writes. In addition, the presence of diverse numeric systems across languages, as seen in applications such as postal codes, introduces a greater level of complexity. By automating the recognition process, it eliminates the need for manual involvement, resulting in cost-effective and time-efficient performance. This paper introduces innovative techniques within an end-to-end framework to enhance the precision of handwritten numeral recognition. It addresses the issues associated with diverse writing styles and the multilingual nature of the problem, along with variations in numerical systems, where dealing with 120 unique classes (12 languages each one with 10-digit signs), poses a considerable challenge.

Researchers have conducted extensive studies in machine learning with a focus on the recognition and classification of handwritten numerals. Various machine learning algorithms, including SVM (Supported Vector Machine) [11-14], KNN (K Nearest Neighbor) [15], RF (Random Forest) [16], Fuzzy models, and FDDL [15, 17, 18], have been employed in the

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

past for recognizing handwritten numerals, demonstrating favorable results in terms of recognition accuracy. For instance, Abhishek Sethy proposed a subsequential method that combined LS-SVM and RF classifiers, achieving an impressive overall accuracy of 99.01% on handwritten Odia character dataset [16]. In recent years, the advent of deep learning has led to a renewed focus on improving the effectiveness of systems for recognizing handwritten scripts, particularly through the use of RNNs (Recurrent Neural Networks) and CNNs (Convolutional Neural Networks) [2, 19, 20]. Notably, Abu Sufian introduced an end-to-end approach utilizing a densely connected convolutional neural network (BDNet) and achieved a remarkable accuracy of 99.78% on ISI Bengali handwritten numerals [21]. It is worth noting that most of these methods are validated only in a single language and lack the versatility to handle multilingual handwritten numeral recognition. Numerous studies have focused on identifying handwritten digits in single languages like English or Chinese [22]. However, the field of multi-lingual digit recognition has seen relatively fewer research efforts [6, 13, 23]. This task presents unique challenges, including variations in size, thickness, stroke direction, order, and language-specific features such as spacing, loops, and hooks. This diversity makes it challenging to develop an integrated model for accurate identification across all languages, therefore a robust and adaptable architecture is required to generalize to new languages without extensive retraining [23]. Additionally, creating a well-balanced dataset is complicated due to varying data availability and quality across languages [13]. One feasible approach is the adoption of language-specific models trained on distinct datasets. However, this methodology necessitates the development of individual digit recognition models for each language, thereby consuming more resources and introducing complications in model integration and processing procedures [23].

Our methodology tackles the challenge of accurate recognition per language through a combination of techniques. Firstly, we employ convolutional neural networks (CNNs) to extract relevant features from digit images and improve the performance of our models. To tackle the challenges of multi-class classification, we leverage a language recognition model. The model's output flows through digit classifiers, each associated with a predicted language, allowing us to identify the numeral represented in the given image. Secondly, we introduce the MRA (Multi-Resolution Attention) UNet model for the super-resolution task. This unique approach involves training the model with digitally generated numeral images. These images are made of thousands of fonts from various languages. We further utilized this model to enhance the quality and resolution of handwritten digits without the need for fine-tuning handwritten data. Furthermore, we apply transfer learning to convey the weights from the language recognition model to the digit recognition models. By utilizing the pre-trained knowledge and features, we optimize the performance of the digit recognition models while significantly reducing training time and computational resources. The integration of transfer learning and the MRA UNet model shows effective promises in improving accuracy and resolving the challenges of multi-lingual handwritten digit recognition.

To summarize, we present five novel approaches that have improved the performance of the whole framework. These contributions can be outlined as follows:

1) Multilingual Categorization: We introduced an innovative language recognition model capable of distinguishing between handwritten numerals from 12 distinct languages.
2) Attention-wise Module: We have designed the MRA module based on an attention-wise block, which is incorporated into various models of this work, such as UNet, language recognition, and digit recognition models. This module plays a crucial role in enhancing model performance by selectively attending to relevant features and improving the overall representation learning process.
3) Digitally Generated Dataset: We have generated a dataset of 100,000-digit images from 12 languages. These images are carefully examined to resemble the handwritten digit images closely. We have ensured that the dataset maintains balance among different digits and languages. Then, the generated dataset is used to train the UNet-based model.
4) Image Enhancement: We have developed a UNet-based model to improve the quality of input images. This approach includes two novel aspects. The first is the incorporation of the MRA module at the skip connection, and the second is the training of this model using the digitally generated dataset. This model is then used to predict handwritten digit images. It demonstrated high accuracy without requiring additional fine-tuning.
5) Digits Recognition Using Transfer Learning: In our digit recognition models, we have adopted the transfer learning approach. This method uses the knowledge obtained from the language classifier which simultaneously decreases both the training duration and computational expense.

The following sections of this paper are organized as follows. In Section 2, we review the state-of-the-art works related to our research. Section 3 provides details about our proposed method, explaining its design and how it works. In Section 4, we evaluate the performance of our models, presenting experimental results that show the effectiveness of our approach. Finally, in Section 5, we conclude by summarizing the key findings and suggesting possible directions for future research.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

II. Related Work

This section presents a comprehensive overview of related works in the field of handwritten numeral recognition. The review encompasses various methodologies, including those using handcrafted features, traditional classifiers, deep neural networks, sparse representation, transfer learning, and combinations of these. By reviewing the literature on these different approaches, we aim to establish a strong foundation for the proposed advancements in our research.

A. Classic Approach

The support vector machine (SVM) is a well-known and widely used classification technique in pattern recognition. The study [11] proposed a Devanagari character classification method that utilized support vector machines (SVM) to recognize and classify documents written in Hindi, Sanskrit, and Marathi, both in printed and handwritten forms. The method begins by preprocessing and segmenting the image through projection profiles, removing shirorekha, and extracting features. The SVM algorithm is then applied to classify characters without the shirorekha into predetermined categories. The study demonstrates the effectiveness of the proposed SVM-based method, achieving high classification accuracies of 99.54% for printed images and 98.35% for handwritten images. These results outperform other Devanagari OCR-based techniques.

In [17], a technique was suggested for the recognition of handwritten Hindi and English numerals, utilizing exponential membership functions as a fuzzy model. The fuzzy sets are generated by calculating normalized distances using the Box approach. The membership function undergoes modification through the optimization of two structural parameters, achieved by maximizing entropy while ensuring the membership function attains unity. The recognition rates are 95% for Hindi numerals and 98.4% for English numerals. This method has potential applications in character recognition and can contribute to developing efficient recognition systems.

The research [12] proposes a hybrid MLP-SVM method for recognizing unconstrained handwritten digits. The method utilizes specialized Support Vector Machines (SVMs) to improve the MLP's performance in local areas around the separating surfaces between each pair of digit classes. The hybrid architecture achieves a recognition rate of 98.01% for real mail zip code digits recognition tasks. The method introduces a rejection mechanism based on the distances provided by the local SVMs, which improves the error-reject trade-off performance. The proposed method can be applied to other scenarios where an MLP network demonstrates strong performance.

B. Deep learning Approach

Deep learning is a transformative term, that has revolutionized the industries, outperformed the classical approaches, and showed promising results. Deep learning techniques have a wide range of applications in various fields, from image and video generation, and transferability across multiple 3D models to medical diagnostics and OCR [24-27].

The BDNet model based on densely connected convolutional neural networks is proposed for Bengali handwritten numeral digit recognition, achieving a test accuracy of 99.78% and reducing error by 47.62% In comparison to prior state-of-the-art models [21]. The training of the model is conducted utilizing the ISI Bengali handwritten numeral dataset with unconventional pre-processing and augmentation techniques. A dataset of 1000 Bengali handwritten numeral images is created to test the model, showing promising results.

To improve the recognition accuracy of handwritten Kannada numerals, [28] introduces a method that takes a document image containing numerals written in diverse styles by various users as input. The method involves a series of pre-processing steps, including noise removal and attribute extraction techniques like Drift Length Count and DWT, followed by a deep convolution neural network (DCNN) classifier for classification. This approach leads to an impressive accuracy rate of 96%.

The study [29] presents a historical handwritten digit dataset called DIDA, containing single and multi-digit images from Swedish handwritten documents. The paper introduces DIGITNET, a deep learning framework consisting of DIGITNET-dect and DIGITNET-rec for digit detection and recognition, respectively. DIGITNET-dect outperforms previous methods, Specifically, DIGITNET-rec sets a new benchmark in detecting and recognizing historical Swedish handwritten documents, achieving an impressive accuracy of 97.12%. It demonstrates the effectiveness and efficiency of the proposed framework.

Study [30] presents an EfficientDet-D4 model for recognizing handwritten digits. EfficientDet-D4 is based on EfficientNet-B4, which is a highly efficient convolutional neural network that balances depth, width, and resolution using compound scaling and advanced techniques, achieving state-of-the-art accuracy on various computer vision tasks. The deep learning-based model can accurately detect and classify digits with an accuracy of 99.83% on the MNIST dataset and is robust to post-processing attacks with an accuracy of 99.10% on the USPS dataset. The model is a reliable solution for digit recognition and has the potential for use in automated number plate recognition and optical character recognition applications.

C. Sparse learning Approach

In paper [31], a novel technique for dictionary learning is proposed, called labeled projective dictionary pair learning, which uses a synthesis-analysis dictionary pair to simplify the calculation of sparse representation. A robust pattern recognition model is achieved by incorporating HOG features and class labels as a penalty term. The proposed method is

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

evaluated on various databases, including Chinese, Arabic, and English handwritten number datasets, and compared with state-of-the-art methods. The results demonstrate the effectiveness of the proposed technique (~98%), which requires only eight parameters to be fine-tuned and operates on regular computers without relying on cloud servers or GPUs, functioning locally. However, dictionary learning may not be as effective as deep learning approaches in handling complex and high-dimensional data contributing to the classification task.

The research study [22] introduced a novel approach to improve dictionary pair learning classification accuracy by adding an incoherence penalty term. This research offers a new dataset for benchmarking pattern recognition algorithms using the InDPL algorithm and cross-validation methods. The InDPL algorithm achieved superior results on a new Chinese number database, especially with limited training samples. The resulting accuracy with the InDPL algorithm, close to 97%, demonstrates better performance than the DPL and KNN methods.

In [15], the authors proposed the FDDL method for image classification. FDDL learns a structured dictionary with discriminative coefficients and uses reconstruction residual and representation coefficients for classification. FDDL outperformed numerous cutting-edge approaches based on dictionary learning methods on various image recognition tasks, with a reported recognition error rate of 2.89% on the USPS English handwritten database.

D. Transfer learning Approach

The research study [32] aimed to develop effective deep-learning models for recognizing Arabic (Indian) digits and texts using transfer-learning approaches. The investigation utilized two well-known transfer models, namely GoogleNet + LSTM and VGG-16 + LSTM, with modifications that included the integration of LSTM layers to enhance the models' recognition capabilities. The LSTM layers functioned as a recurrent neural network, persisting the extracted features from the convolutional neural network (CNN) part. The research outcomes demonstrated the efficacy of using LSTM layers in the transfer learning models to learn long-term dependencies which achieve impressive accuracy levels, reaching up to 99%, accompanied by noteworthy recall and precision values when classifying the ten digits. The proposed models were shown to be comparable to, or even more effective than, state-of-the-art techniques, with accuracy ranging from approximately 98% up to 99.3%.

The study [6] proposes a new recognition algorithm for a pattern recognition system using deep convolutional neural networks (CNNs) to improve the recognition performance of handwritten digits. The approach involves developing several deep CNNs to maximize accuracy for each class, and the models are chosen based on standard deviation and median. To enlarge the MNIST dataset, standard augmentation techniques were employed. The dataset was then divided, with 75% allocated for training purposes and the remaining 25% for testing. This division allowed for a more accurate and comprehensive analysis to be conducted. Experimental results demonstrate high accuracy in recognizing digits, ranging from 97.82% to 99.72% within a few epochs. The proposed method enhances the intra-class correlation. Improving classification accuracy for each class, and demonstrating superior recognition ability compared to the state of the art in several cases.

The study [2] explores the categorization of handwritten Gujarati numerals ranging from zero to nine using deep transfer learning techniques. The investigation utilized ten pre-existing CNN architectures, namely LeNet, VGG16, InceptionV3, ResNet50, Xception, ResNet101, MobileNet, MobileNetV2, DenseNet169, and EfficientNetV2S, to identify the most suitable model through the fine-tuning of weight parameters. The implementation of the pre-trained models was carried out Utilizing a self-curated dataset of handwritten Gujarati digits, consisting of 8000 images of zero to nine along with data augmentation techniques. Multiple experiments were conducted using diverse performance evaluation matrices. The results revealed that the EfficientNetV2S model, considering all models, including three scenarios of transfer learning, showed encouraging results, achieving a training accuracy of 98.39%, testing accuracy of 97.92%, f1-score of 97.69%, and AUC of 97.15%.

E. Multilingual Approach

Recognizing and classifying multilingual handwritten digits is challenging due to several factors. One of the primary challenges is the variations in writing styles and digit shapes across different languages. For example, the shapes of digits in Arabic and Chinese languages are significantly different from those in English and other Latin-based languages. This results in a need for large and diverse datasets to train and evaluate the models. Another challenge is the varying sizes of digits, as writers tend to have different writing sizes and styles. Furthermore, developing a model that can recognize and classify digits from multiple languages is more challenging than developing a model for a single language, as the model must handle the variations across multiple languages while maintaining high accuracy. Overcoming these challenges requires the development of robust and adaptable models that can handle diverse writing styles, shapes, sizes, and image qualities, which is an ongoing area of research in multilingual character recognition. [33].

The study [13] presents a novel approach to address the challenges in multilingual numeral recognition systems. The authors conducted extensive experiments on datasets containing numerical digits of 8 languages, including both Indic and non-Indic scripts. The proposed method achieved an impressive accuracy rate of 96.23% for all eight scripts combined, demonstrating the effectiveness of convolutional neural networks (CNN) in multilingual handwritten

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

numeral recognition. The system stands out for its distinctive approach of employing a ten-class recognition system for multiple languages without the need for prior numeral identification. Additionally, it is the first attempt to develop a fusion-free approach for recognizing handwritten numerals in eight different languages, independent of the script used in each language, challenging the notion of multilingualism in handwritten character recognition.

The research work [19] developed a knowledgeable framework for Handwritten Character Recognition (HCR) using Neural Networks, which can accurately identify specific type-format characters. The proposed methodology involves the utilization of both a machine learning model and a character recognition MATLAB model to recognize and identify handwritten digits accurately. A translator using MATLAB is also designed to overcome language barriers. The proposed technique can be employed to convert English, Marathi, and Gujarati text into spoken English using a text-to-speech conversion approach. The experiments were performed on several datasets using MATLAB and ANACONDA software systems, including Gujarati, Hindi, and English literature. The accuracy of the digit recognition model was evaluated using existing datasets like MNIST and custom-made CSV files.

In [23] a CNN-based model that is not specific to any particular language is proposed to address the challenge of recognizing numerals in six different languages. The model includes language recognition and digit recognition components to handle multi-script images. Transfer learning is used to enhance image quality and recognition performance. Extensive experiments are conducted to verify the effectiveness of the model for recognizing the recognition of digits associated with various languages. Testing the model with six different languages shows an average accuracy of up to 99.8%. The model's resilience and the procedure employed in its design make it a cost-efficient solution for recognizing handwritten numeric symbols in various languages.

III. Proposed Method

As shown in Figure 1, the proposed method consists of five parts:

Firstly, the input data goes through a preprocessing step. Handwritten digit images in the original datasets are usually in a smaller size, typically 28x28 pixels. However, merely Up-sampling these images can lead to a decline in image quality, negatively impacting the classifier's performance. To overcome this challenge, as our second step, we have developed a robust and novel UNet-based model named MRA UNet, which utilizes transfer learning techniques to enhance the quality of the images. Thirdly, we integrate a language recognition model to identify the language of each processed image. Following this, as our final step, we use a digit recognition model that specifically classifies digits associated with the identified language. The digit recognition model utilizes the transfer learning approach from the language recognition model and is fine-tuned for digit recognition. Detailed implementation aspects of this system will be provided in subsequent sections.

A. Preprocessing

In this research, our focus was on classifying 12 different languages. However, high-quality handwritten images in the size of 128x128 pixels were not readily available for training the MRA-UNet model. To address this, a unique approach was adopted, involving the collection of a diverse set of more than 1000 fonts encompassing all the languages. Data augmentation techniques were subsequently applied to ensure a balanced dataset with diverse examples, especially for languages with limited font availability. For the input of the super-resolution model, low-quality images were needed. As shown in Figure 1, to generate these inputs, we initially created high-quality images and then applied both down-sampling and up-sampling operations. This process allowed us to obtain low-resolution versions of each digit image, which served as suitable inputs for training the super-resolution model.

As previously mentioned, to overcome the limitations of data availability in some languages, we employed data augmentation techniques. Specifically, we utilized horizontal shift and rotation functions to increase the number of images by several folds. Data augmentation offers several benefits. Firstly, it addresses the issue of limited data by expanding the dataset, particularly in languages with insufficient sample sizes. By generating additional augmented samples, we enrich the model with

A wider range of training examples. Secondly, data augmentation acts as a regularization technique by introducing variations into the training data. This prevents the model from memorizing specific instances and encourages it to learn more robust and generalized features that can be applied to unseen data. Additionally, data augmentation ensures a more balanced representation, preventing the model from being biased towards dominant classes and improving its performance in accurately classifying all classes.

B. Dataset

In this paper, a novel 100,000 digitally generated dataset of digit images is introduced. We obsessively selected fonts from 12 distinct languages to construct this dataset using a Python script. Table 1 shows sample images of 12 different handwritten digits datasets.

The training data for handwritten digits utilized in this study includes twelve datasets. The MNIST-MIX is chosen due to the extensive experiments conducted on this dataset, making it an excellent benchmark for this study. It consists of Persian, Bengali-Lekha, Tibetan, Urdu, ISI Bangla, ARDIS (Swedish), languages, and lastly Kannada, a challenging dataset with low

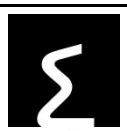
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

accuracy in inference time [33]. Chinese handwritten numbers dataset is another contender which brings additional complexity to the investigation as it has a distant representation in latent space compared to others. It includes samples from 100 individuals with varying handwriting styles, collected by the Newcastle University research team in the UK from Chinese nationals. Additionally, a comprehensive set of English handwritten digits from USPS was compiled [22]. We collected datasets containing Gurmukhi and Gujarati images from GitHub repositories as well [34, 35]. Furthermore, English (USPS [20]) and Arabic datasets (MADBase [36]) are considered to assess the proposed method. By training our models on digit images from 12 languages, we improve the generalization ability of models on other language numerals. Table 2 presents representative examples from the datasets employed in this study. On the other hand, Table 3 provides a brief summary of the dataset characteristics, comprising the number of samples and their sizes.

TABLE 1

DIGIT FOUR IN LANGUAGES OF DIGITALLY GENERATED DATASET

 Arabic	ARDIS	Chinese	Farsi and Urdu	Gujarati
ISI Bangla and BanglaLekha	Kannada	Tibetan	English	Gurmukhi

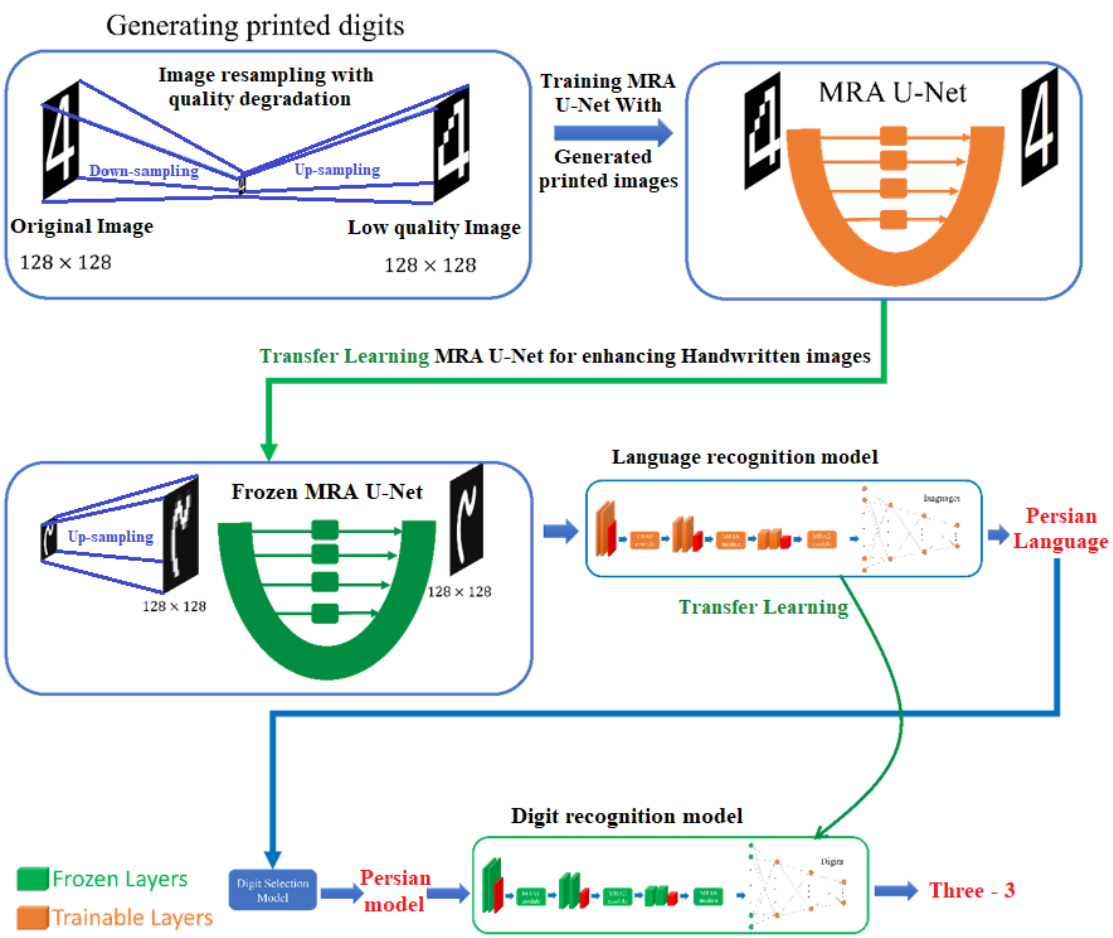


Figure 1. The main block diagram of the proposed model. Recognition of Persian digit '3' image is demonstrated as an example.

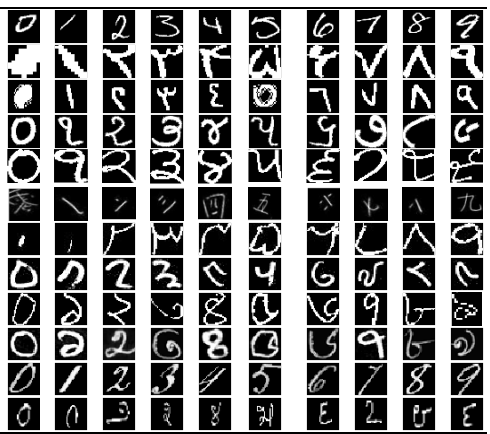
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

One of the major challenges we ran into, was to collect a suitable and adequate range of fonts for each language. We took great care to collect and validate each font, ensuring they correctly produced the intended digit with the right size and positioning within the image frame. We found that some fonts created completely black or unrelated images, or irrelevant images which we then removed from our collection. For languages like Gujarati, Gurmukhi, and Kannada, there were fewer fonts available, resulting in less varied and fewer overall images compared to other datasets. We managed to balance this out by using image augmentation techniques.

TABLE 2

SAMPLE IMAGES OF 12 DIFFERENT HANDWRITTEN DIGITS DATASETS.

Language / digits	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
English										
Persian										
Arabic										
Gujarati										
Gurmukhi										
Chinese										
Urdu										
Tibetan										
Bangla Lekha										
ISI Bangla										
ARDIS										
Kannada										

C. Super Resolution

In our research, the need to improve the quality of high-resolution handwritten digit images posed a significant challenge. To overcome this challenge, we looked for a robust and high-capacity model with strong learning capabilities. In light of this objective, we chose a model based on the well-established UNet architecture. The proposed model comprises three main components, as depicted in Figure 2.

The encoder section of our model consists of multiple blocks, each composed of two 3x3 convolution layers. The filter sizes for these layers are set to 64, 128, 256, and 512, respectively, for the first, second, third, and fourth blocks. At the end of each block, a 2x2 max-pooling operation is applied. The outputs from each encoder block are divided into two branches. One branch proceeds to the next encoder block, while the other branch enters the specially designed MRA module within the skip connection part of the UNet.

The bottleneck section consists of four 3x3 convolution layers arranged sequentially. The outputs from the bottleneck section are then fed into the decoder section. The decoder section, which is the third part of our model, comprises four blocks. Each block includes a transposed convolution layer. The first input layer is obtained from the bottleneck section, followed by concatenation with the corresponding feature maps from the skip connection part at the same level in the encoder. Subsequently, two convolution layers are applied. The filter size of each convolution section is reversed compared to the encoder blocks, with filter sizes of 512, 256, 128, and 64, respectively. Similarly, the output from each decoder section enters the higher-level decoder section, and this process continues.

The choice of hyperparameters has been carefully made to enhance the model's effectiveness. The optimizer used in this model is Adam (Adaptive Moment Estimation). Adam is a popular choice due to its computational efficiency and has little memory requirement. The learning rate is set to 0.001. This small learning rate ensures that the model does not skip over any potential solutions to the optimization problem. The last layer uses a sigmoid function, typically used in binary classification models to predict class probabilities. This is suitable as our output images consist of pixel values of either 0 or 1. The model uses categorical cross-entropy as the loss function, ideal for multi-class classification where each pixel is a category.

We have introduced a novel module architecture called the MRA module, which is integrated into the skip connection part of the UNet. This module operates on the convolutional feature maps that are passed from each decoder block. As illustrated in Figure 3 The module first applies different sizes of convolution filters (1x1, 3x3, and 5x5) to the input layer. Each convolution operation extracts different spatial features from the input image. The outputs of these convolutions are then added together, which means the module combines the features extracted by different sizes of filters. The combined features then go through another 3x3 convolution operation, and the output of this operation is used to calculate the channel-wise attention. The channel-wise attention mechanism allows the model to focus on more informative channels and suppress less useful ones. It does this by applying a Global Average Pooling operation, followed by two 1x1 convolutions and a sigmoid activation function. The output of this process is a set of attention scores, one for each channel. These attention scores are then used to weight the output of the 3x3 convolution operation. This means that the channels that the model finds more informative will have a greater influence on the final output of the module.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

TABLE 3

SUMMARY OF HANDWRITTEN DIGITS DATASETS FOR TRAINING.

Dataset	*Size*	*Amount*	*Language*
Custom	128x128	100,000	All Languages
USPS[20]	variable	20000	English
MNIST-MIX[33]	28x28	60000	Persian
MADBase[36]	28x28	60000	Arabic
Gujarati [34]	256x256	5600	Gujarati
Gurmukhi [35]	32x32	1000	Gurmukhi
Chinese [22]	64x64	60000	Chinese
MNIST-MIX [33]	28x28	6606	Urdu
MNIST-MIX [33]	28x28	14214	Tibetan
MNIST-MIX [33]	28x28	15798	BanglaLekha
MNIST-MIX [33]	28x28	19392	ISI Bangla
MNIST-MIX [33]	28x28	60000	ARDIS
MNIST-MIX [33]	28x28	60000	Kannada

The MRA module is designed to extract and combine a wide range of spatial features from the input image, while also allowing the model to focus on the most informative channels. This can help the model to capture complex patterns in the image data and improve its performance on tasks such as image classification or object detection.

For training this model, we utilized the transfer learning technique. To create a balanced dataset, we collected 10,000 low-quality and high-quality digit images for each language, resulting in a total of 100,000 images across all languages. (Considering the languages BanglaLekha, ISI Bangla, and Farsi, Urdu, they are perceived as alike). This thoughtfully assembled dataset was then used to train our robust model.

Following the training process, the obtained model was utilized for the prediction and generation of handwritten digit images without requiring any additional retraining. The results accomplished from this approach showed excellent outcomes.

D. Language Recognition

Given the multilingual nature of our datasets, which encompass 12 languages with 10 numerical characters each (excluding the Chinese dataset with 15 classes), we face the challenge of recognizing a total of 125 distinct number classes. Handling such a large number of classes with a single model poses challenges. Additionally, treating all similar numbers across languages as the same class can slow down the model's ability to learn effectively, as there may be substantial variations in how the same numbers are represented across different languages. Moreover, adding a new language would require retraining the entire model or fine-tuning.

To address these challenges, we propose an approach that involves using a language detection model to determine the language of the input image. Based on the detected language, a corresponding classifier model is selected to recognize the specific type of number. By separating the language detection process from the digit recognition model, we achieve a more streamlined approach for incorporating new languages into our system. This separation allows us to develop new number recognition models for additional languages and integrate them smoothly into the existing system, without disrupting the language detection and other digit detection components. This flexibility ensures scalability as the number of supported languages continues to grow.

To address the task of language detection for our system, we looked for a model capable of robustly extracting and learning essential features. In Figure 4 we introduce our novel "language classifier based on MRA" model, which consists of three blocks. Each block is composed of two convolutional layers including a 3x3 filter, a max-pooling layer, and an MRA module. Notably, the convolutional layers in these blocks are thoughtfully configured with 64, 128, and 256 filters, respectively.

Following the three feature extraction blocks, we incorporate three fully connected layers, including 512, 256, and 128 units. The final layer of the model consists of 12 neurons, aligning with the number of classes (languages), and uses the Softmax activation function to generate probability distributions over the language classes. Remarkably, the model is trained using a balanced dataset containing 10,000 samples for each language, producing outstanding recognition accuracy.

We have chosen the Softmax activation function for the final layer. The loss function we have used is categorical cross-entropy, and we have selected the Adam optimizer with a learning rate of 0.001. One of the significant advantages of this proposed language recognition model lies in its simplicity and efficient structure, which contributes to a small number of parameters, rendering it computationally lightweight. Furthermore, due to its architecture, this model can also be effortlessly integrated into the digit recognition step, thereby providing a unified framework for the entire handwritten digit recognition system.

We encountered an additional challenge due to the resemblance in the shapes of various numbers in different languages. To better illustrate this point, Table 4 provides a comprehensive overview. It's crucial for our language classifier to accurately recognize the correct class in such situations.

TABLE 4

SIMILAR IMAGE IN DIFFERENT LANGUAGES

Number one in Gujarati	Number one in Gurmukhi	Number seven in BanglaLekha	Number seven in ISI Bangla	Number nine in Urdu	Number nine in English	Number nine in Arabic	Number nine in Persian

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

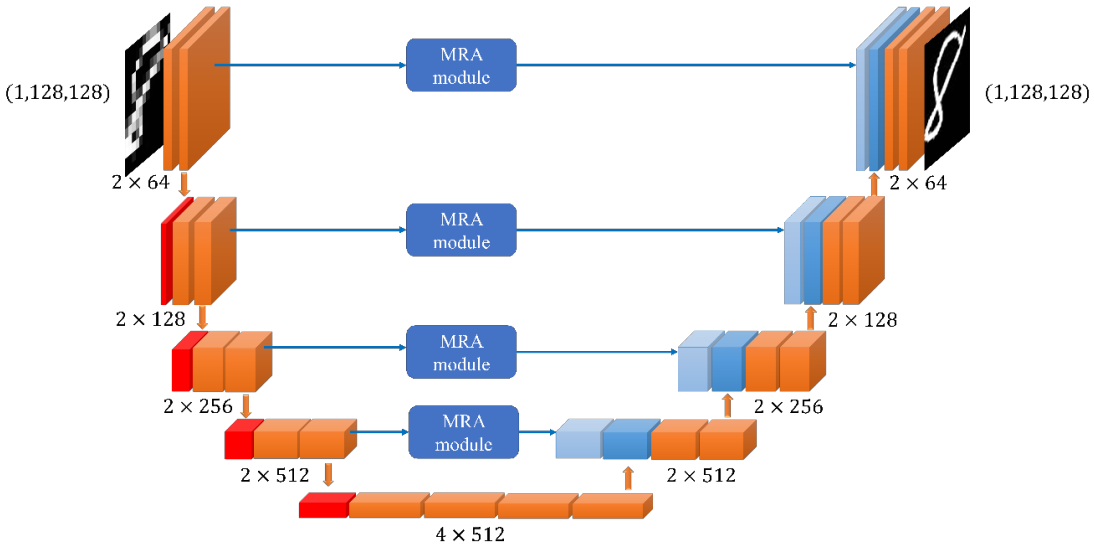


Figure 2 Diagram of the proposed UNet model using (MRA) module

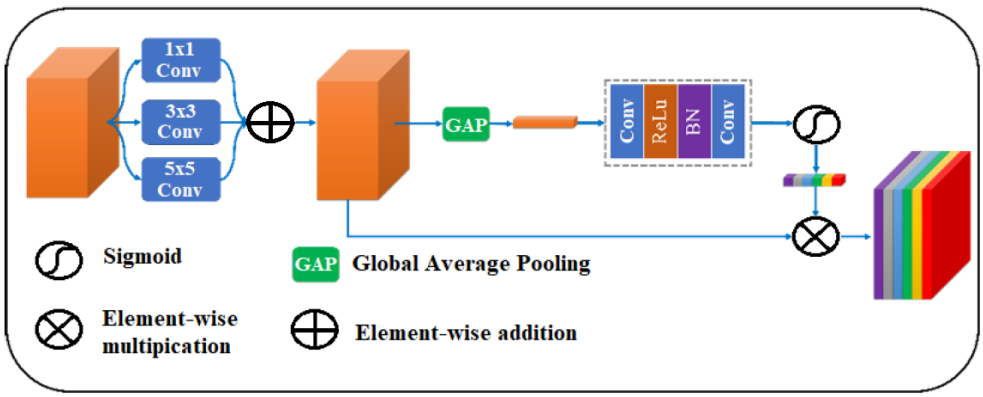


Figure 3 Multi-Resolution Attention (MRA) module

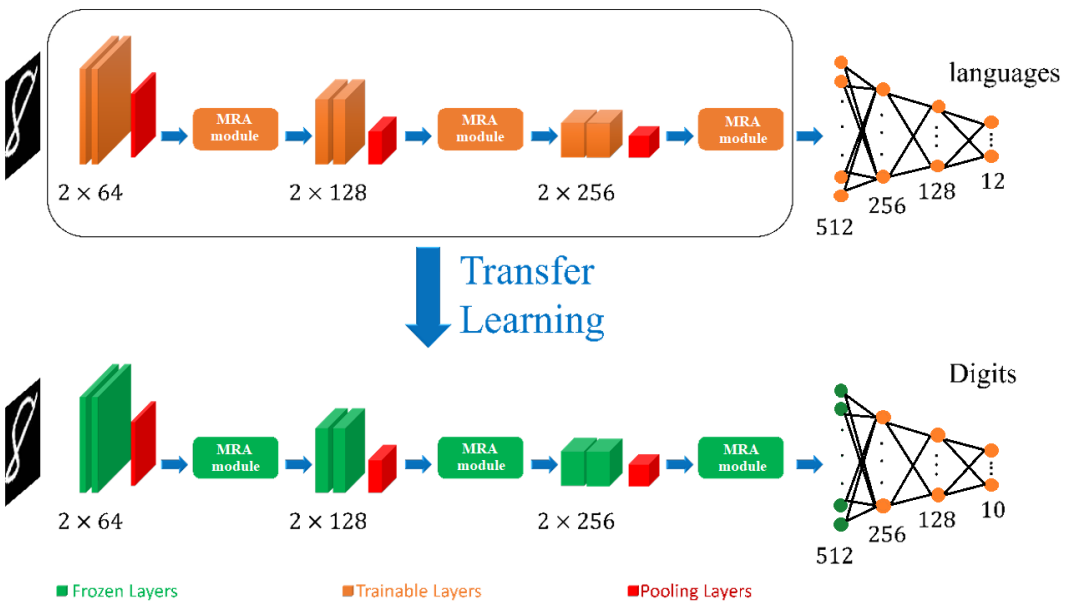


Figure 4 Architecture of Language recognition and Digit recognition model using Transfer Learning based on (MRA) module

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

E. Digit Recognition

In the digit recognition stage, we utilize the language-specific classifier models to achieve accurate number recognition, resulting in 12 distinct models, each corresponding to a particular language. To automate this process and simplify the recognition pipeline, we have developed a complex module that seamlessly integrates the language detection model with the corresponding number recognition model. This integration enables us to efficiently direct the input image to the appropriate number recognition model based on the identified language, thereby enhancing the overall efficiency of the system's output.

The structure of the digit recognition model closely resembles that of the language detection model, utilizing transfer learning techniques and fine-tuning to optimize performance. Within the last fully connected layer, we utilize 10 neurons to facilitate the classification of the ten classes, representing the numbers from 0 to 9. It is important to note that the final model structure is specialized for each language.

The incorporation of transfer learning in the digit recognition process offers several significant advantages during model training. As illustrated in Figure 4 this approach significantly reduces the required training time and computational resources compared to training a model from scratch. We improve the accuracy and performance of the digit recognition model, leading to enhanced generalization and overall system efficiency.

IV. Experimental Results

In this section, we present the experimental results and analyses of our proposed approach for multilingual handwritten digit recognition. To evaluate the performance of our model, we compare it with existing methodologies. Additionally, we investigate the effects of transfer learning on the model's performance and conduct an ablation study to assess the impact of individual components within the proposed model. Notably, the implementation and training of our models were conducted on Google Colab Pro, providing a reliable and scalable computational environment for our experiments.

A. Comparative performance

In this subsection, we performed a thorough comparison of the performance between our proposed model and other relevant methods. As shown in *TABLE 6* The datasets of various languages, including English, Chinese, Arabic, Persian, Urdu, Gujarati, Gurmukhi, Tibetan, BanglaLekha, ARDIS, Kannada, and ISI Bangla, were trained using different approaches. Among the tested methods, Gupta CNN-Based [13] achieved a commendable accuracy of 96.70% in recognizing ISI Bangla digits. In comparison, Fateh CNN-Based [23] displayed competitive accuracy in recognizing Chinese and Arabic digits, with respective accuracies of 99.26% and 99.18%. InceptionV3, a widely recognized method, demonstrated strong performance across all languages, attaining an impressive accuracy of 99.00% in recognizing ARDIS digits. This is while our proposed mothod achieved the accuracy of 98.70%.

As indicated by the results in Table 5, among the various popular model architectures, Lenet-5 is considered to be the most suitable Neural Network architecture. However, through a comparative analysis of the obtained results, it is evident that the proposed method has consistently outperformed the Lenet-5, VGG16, ResNet50, and MNIST-MIX methods across various languages. While certain limitations may arise when comparing specific languages to alternative methods, such as the Chinese dataset where Fateh CNN-Based achieved a higher accuracy of 99.26 % on the testing data, the inclusion of MRA UNet addressed this limitation and increased the accuracy from 98.38 to 99.30%, surpassing Fateh CNN-Based. Furthermore, the proposed method enhanced by UNet exhibited substantial improvements, surpassing other methods in recognizing Gujarati, Arabic, Persian, and English digits, with exceptional accuracy rates of 99.23%, 99.10%, 98.70%, and 99.75%, respectively.

In our study, we compared the performance of the proposed model with other relevant techniques, particularly dictionary learning methods. We evaluated six dictionary learning methods (SRC, DPL, DLSI, InDPL, LC-KSVD1, and LC-KSVD2). we conducted language measurements for English, Persian, Arabic, Chinese, and Urdu. The results showed that among the various dictionary learning methods, DPL stood out as the most effective. Additionally, our proposed method demonstrated superior performance compared to the other dictionary learning methods.

In a particular study that utilizes the MNIST MIX dataset, 14 languages are examined. The languages under consideration are listed in a table. When the numbers presented are compared with the results from our model, which does not use super resolution, it is found that our model's accuracy surpasses that of all other reported languages. For instance, we take the Kannada dataset. Based on [23], [33] and [37] papers, accuracies of 88.01, 85.70, and 86.69 have been reported. Nonetheless, our approach demonstrates its superiority by attaining an accuracy of 88.50% under basic usage, and this accuracy increases to 90.28% when the super resolution method is applied.

B. Transfer Learning Effects

The language recognition model has achieved an average accuracy of 99.27%. The trained weights from this model have been applied to digit recognition models, leading to a reduction in the number of parameters, as shown in Table 5.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

TABLE 5

COMPARING NUMBER OF PARAMETERS USING TRANSFER LEARNING

Models	*number of parameters for LR model*	*number of parameters for each DR model*	*Total number of parameters for all languages*
Approach 1	10.7 m	10.7 m	139 m
Approach 2	$\mathbf{10.7\ m}$	$\mathbf{0.17\ m}$	$\mathbf{2.2\ m}$

TABLE 6

EXPERIMENTAL RESULTS ON 12 DIFFERENT DATASETS.

Accuracy for digit recognition model per Language		**English**	**Persian**	**Arabic**	**Gujarati**	**Gurmukhi**	**Chinese**	**Urdu**	**Tibetan**	**Bangla Lekha**	**ISI Bangla**	**ARDIS**	**Kannada**
Dictionary learning-based methods	**LC-KSVD1 [38]**	91.25%	91.15%	96.49%	-	-	95.23%	87.48%	-	-	-	-	-
	LC-KSVD2 [39]	91.10%	91.15%	96.49%	-	-	95.24%	87.74%	-	-	-	-	-
	DLSI [40]	96.10%	97.30%	97.62%	-	-	97.80%	89.03%	-	-	-	-	-
	DPL [41]	96.68%	98.46%	98.21%	-	-	98.36%	95.23%	-	-	-	-	-
	SRC [42]	81.81%	82.69%	90.97%	-	-	90.97%	85.32%	-	-	-	-	-
Deep learning-based methods	**Gupta CNN–Based [13]**	99.68%	-	96.53%	99.22%	-	-	-	-	-	96.70%	-	-
	Fateh CNN-Based [23]	97.33%	98.99%	99.18%	-	-	99.26%	98.23%	-	-	-	-	88.01%
	InceptionV3 [43]	98.74%	99.00%	97.20%	97.67%	94.38%	98.64%	98.10%	97.41%	97.67%	97.46%	98.48%	86.30%
	ResNet-50 [44]	90.33%	97.10%	96.89%	95.29%	93.82%	92.57%	96.46%	96.87%	92.58%	91.69%	92.19	80.14%
	VGG-16 [45]	96.33%	96.95%	98.58%	97.41%	99.23%	96.67%	96.67%	96.31%	93.74%	94.45%	92.79%	79.49%
	LeNet-5 [46]	96.86%	98.40%	98.18%	98.58%	96.62%	96.81%	96.81%	98.14%	96.50%	96.20%	98.40%	84.66%
	Advanced MNIST-MIX[37]	98.57%	94.10%	95.54%	-	-	-	95.95%	94.53%	87.14%	94.86%	98.39%	86.69%
	MNIST-MIX [33]	-	98.18%	-	-	-	-	97.31%	98.28%	94.86%	97.05%	98.20%	85.70%
Proposed method	**Ours**	98.86%	98.52%	98.12%	98.86%	97.87%	98.38%	98.54%	99.14%	99.24%	98.84%	99.10%	88.50%
	Ours^{SR}	**99.75%**	**98.70%**	**99.25%**	**99.23%**	**99.09%**	**99.30%**	**98.86%**	**99.55%**	**99.75%**	**99.48%**	**99.65%**	**90.28%**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

As previously stated, we developed the MRA UNet model and utilized transfer learning to enhance the resolution of handwritten digit images. We trained our model on a dataset consisting of 100,000 handpicked digital images generated. This dataset includes both high-quality and low-quality images, with a standardized resolution of 128x128 pixels. Remarkably, our model exhibited exceptional performance in enhancing images from all 12 languages, even though it has not been specifically fine-tuned for the handwritten data. *TABLE 6* shows the improved results obtained from nearly all datasets.

In the context of training digit recognition models, the considerable number of language-specific models results in a high computational cost. To reduce this cost, we employed the transfer learning approach, leveraging the knowledge acquired from the language recognition model for each digit recognition task. As shown in Figure 4, by freezing the convolutional and MRA modules' trainable layers, including the first fully connected layer, and subsequently fine-tuning the fully-connected layers, we adapted the model to the specific digit recognition dataset.

In

Table 5, we compare two distinct approaches: Approach 1 involves training each digit recognition model from scratch, while Approach 2 incorporates transfer learning to efficiently reduce the number of parameters. Notably, the results showcased in

Table 5 demonstrate that Approach 2 effectively reduces the number of parameters from 139 million to 2.2 million, underscoring the efficiency of deep learning models in handling complex patterns and variations within handwritten digits.

C. Ablation Study

Ablation study in AI refers to a systematic experimentation process where specific components or modules of a model or algorithm are individually removed or modified to understand their impact on the overall performance. The goal is to isolate and evaluate the contribution of each component to the system's effectiveness.

In this subsection, we conducted an ablation study on our proposed method using a dataset of Chinese handwritten characters as the baseline. Initially, we employed a simple approach without incorporating the MRA module and UNet enhancement. Then, we introduced and experimented with different parts of our method by adding these components to understand the specific contributions and effects of the MRA module and UNet enhancement on the performance of our system.

TABLE 7

EFFECTS OF MRA AND UNET SUPER-RESOLUTION

Base line	*MRA*	*UNet*	*Accuracy*
✓	×	×	95.06%
✓	×	✓	98.89%
✓	✓	×	98.38%
✓	✓	✓	99.30%

Table 7 reveals that the basic model, which doesn't include MRA and UNet, still managed to reach a respectable accuracy of 95.06%. This highlights the significance of these elements. When we added just the MRA module, the accuracy jumped to 98.38%, showing the impact it can have on its own. In the same vein, when we used UNet with its super-resolution features, the accuracy climbed even higher to 98.89%. Comparatively, the combination of the MRA module and UNet proved to be the most effective, demonstrating the highest and most remarkable accuracy of 99.30%, surpassing all other variations tested. This suggests that the integration of these components enhances the model's performance, leading to outstanding accuracy in recognizing Chinese handwritten digit images. we further evaluated the performance of the combination approach by drawing the confusion matrix and calculating precision and recall metrics, Table 8.

The provided confusion matrix demonstrates the classification results for digit images in the Chinese dataset. Although the training model for the Chinese dataset consists of 15 classes, the confusion matrix displays numbers ranging from 0 to 9 to resemble to other language datasets.

Upon analysis, it is evident that a significant number of digits are correctly classified, as indicated by the high values along or near the diagonal. However, there are instances where confusion arises among similar numbers. It is worth noting that there are eight instances where the number 9 is incorrectly identified as 7, and there are also cases where the numbers 8 and 3 are mistakenly associated with 1 and 2. This indicates that the classifier sometimes assigns incorrect labels to instances of the digits 9, 8, and 3, mistaking them for 7, 1, and 2. The observation highlights the possibility of visual similarities among these digits within the Chinese dataset.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

TABLE 8

CONFUSION MATRIX FOR CHINESE DATASET BASED ON PROPOSED METHOD

	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*recall*
0	1000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0
1	0	1000	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0
2	0	2	998	0	0	0	0	0	0	0	0	0	0	0	0	0.998
3	0	0	6	993	0	1	0	0	0	0	0	0	0	0	0	0.993
4	0	0	0	2	992	1	0	0	0	0	0	3	0	0	2	0.992
5	0	0	0	8	0	922	0	0	0	0	0	0	0	0	0	0.992
6	0	0	1	0	0	0	998	0	9	0	1	0	0	1	0	0.988
7	0	1	0	0	0	0	0	1000	0	1	0	0	0	0	0	1.0
8	0	13	0	0	0	0	0	0	985	0	1	1	0	0	0	0.985
9	0	0	0	0	0	0	0	5	0	993	0	0	0	0	2	0.993
10	0	0	0	0	0	0	0	2	0	0	993	0	5	0	0	0.993
11	0	0	0	0	0	0	0	0	0	0	0	994	0	6	0	0.994
12	0	0	0	0	0	0	0	0	1	0	7	0	992	0	0	0.992
13	0	0	0	0	0	1	0	0	0	0	0	0	1	997	1	0.997
14	0	0	0	0	0	0	0	2	0	1	0	0	0	2	995	0.995
Percison	1.0	0.985	0.993	0.990	1.00	0.997	1.00	0.991	0.990	0.999	0.991	0.996	0.994	0.991	0.995	

TABLE 9

CONFUSION MATRIX FOR ARABIC DATASET BASED ON PAPER [23]

	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*recall*
0	983	5	1	0	0	9	2	0	0	0	0.983
1	10	987	0	0	0	0	0	1	1	1	0.987
2	3	2	991	1	1	1	0	0	1	0	0.991
3	1	1	5	987	1	0	3	2	0	0	0.987
4	1	5	5	0	989	0	0	0	0	0	0.989
5	7	0	4	0	4	977	0	4	0	4	0.992
6	0	1	0	0	1	0	996	0	1	1	0.977
7	0	0	0	0	0	1	0	999	0	0	0.996
8	2	0	1	0	0	1	0	0	996	0	0.999
9	0	0	1	0	1	2	2	0	2	992	0.992
Percison	0.9761	0.9860	0.9831	0.999	0.9919	0.9858	0.9930	0.9930	0.9950	0.9940	

TABLE 10

CONFUSION MATRIX FOR ARABIC DATASET BASED ON PROPOSED METHOD

	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*recall*
0	984	2	2	1	0	5	2	2	1	1	0.984
1	16	984	0	0	0	0	0	1	1	1	0.984
2	0	0	999	1	0	0	0	0	0	0	0.999
3	0	1	4	994	0	0	1	0	0	0	0.994
4	0	1	4	1	993	0	0	0	0	1	0.993
5	7	0	2	0	0	987	0	1	1	2	0.987
6	1	1	0	0	1	0	997	0	0	0	0.997
7	0	0	0	0	0	0	0	1000	0	0	1.00
8	1	0	1	0	0	0	0	0	996	2	0.996
9	0	0	2	0	0	0	1	0	0	997	0.997
Percision	0.984	0.994	0.995	0.994	0.993	0.995	0.997	1.00	0.996	0.997	

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

The comparison of confusion matrices for Arabic datasets is presented in Table 9 and Table 10. Table 9 displays the results from the [23], while Table 10 shows the outcomes of our proposed method. Our method introduces slightly improved results when compared to the two studies. We utilized the macro-average of the following metrics to conduct a comparison between these two confusion matrices.

For a better grasp of evaluation metrics, we have proposed some definitions in the subsequent section.

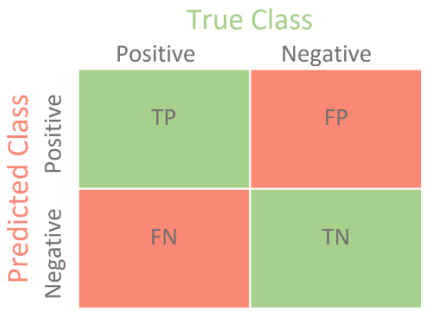


Figure 5. The confusion matrix for performance measurement

Figure 5 presents a confusion matrix, a tool utilized for assessing the performance of classification. For the sake of simplicity, we have depicted a binary class confusion matrix. This matrix introduces four key terms:

- TP (True Positive): Represents the instances where the model correctly predicted the positive class.
- FP (False Positive): Refers to the instances where the model incorrectly predicted the positive class.
- FN (False Negative): Denotes the instances where the model incorrectly predicted the negative class.
- TN (True Negative): Signifies the instances where the model correctly predicted the negative class.

These terms are utilized in precision, recall, and F1-score metrics calculations.

Percision also known as the positive predictive value, is the proportion of relevant instances out of the total retrieved instances.

$$Percision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives(FP)} \quad (1)$$

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negetives(FP)} \quad (2)$$

Recall is the proportion of relevant instances that have been correctly identified. It quantifies the percentage of true positives that were accurately recognized.

The F1 score is a metric that evaluates a model's accuracy on a dataset. It's specifically used for binary classification systems that label examples as either 'positive' or 'negative'. The F1 score is the harmonic mean of precision and recall, effectively merging these two metrics into one single value.

$$F1 = 2\,x\,\frac{Percision\ x\ Recall\ (TP)}{Percision\ + Recall} \quad (3)$$

Macro-average is a method that first computes the metric independently for each class and then takes the average. Hence, it treats all classes equally and evaluates the overall performance of the classifier against the most frequent class labels. Table 11 presents the calculated macro-average, comparing the the results of Fateh CNN-Based [23] and the proposed method on Arabic dataset.

TABLE 11

MACRO AVERAGE FOR ARABIC DATASET

Digit recognition	Macro-average Precision	Macro-average Recall	Macro-average F1 Score
Fateh CNN-Based [23]	0.9948	0.9948	0.9947
Proposed method	0.9952	0.9951	0.9951

In the second subsection, the application of various optimizers in our Chinese digit classifiers is discussed.

TABLE 12

THE ACCURACY OF THE FINAL MODEL BY DIFFERENT OPTIMIZERS ON CHINESE LANGUAGE

Optimizer	Accuracy
Adam	98.38
Adamax	96.75
RMSprop	94.91
SGD	97.45

As shown in Table 12, the Adam optimizer has attained the highest accuracy. These accuracy values pertain to our proposed method without the Super-Resolution (SR) component.

V. Conclusion

This research paper introduces a novel multi-lingual approach for recognizing handwritten digit images of 12 different languages. The proposed system consists of three key models: a UNet-based, a language recognition model, and a digit recognition model, all based on a CNN architecture. A specially designed module called MRA was introduced and significantly improved image quality and recognition accuracy. The system first enhances the quality of input images and then determines the language of the image. The identified image is then processed by the corresponding language model to ultimately determine the final number. Additionally, transfer learning is utilized to ensure consistent performance across different datasets, resulting in reduction in

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

computational costs and parameters. Extensive experiments were conducted to optimize parameters, leading to superior accuracy compared to other techniques, including CNN-based approaches. The proposed framework has demonstrated exceptional precision during the prediction phase, achieving nearly a 2% improvement in certain languages compared to previous techniques. For future work, we aspire to develop a model that demonstrates proficiency across a broader spectrum of languages and numeric systems. Our aim is to utilize transfer learning more extensively, suggesting the extraction and application of insights gained from handwritten digit datasets to various fields, including the analysis of medical images. Maintaining the decrease in computational complexity and the quantity of parameters is another goal for our future outlook. An additional proposal is to broaden the application of this framework by integrating it with digit detection and localization methods. The present constraint of this system is its time-intensive nature, and it's engineered to support only 12 languages.

REFERENCES

1. Rajpal, D. and A.R. Garg, *Deep Learning Model for Recognition of Handwritten Devanagari Numerals with Low Computational Complexity and Space Requirements.* IEEE Access, 2023.
2. Goel, P. and A. Ganatra, *Handwritten Gujarati Numerals Classification Based on Deep Convolution Neural Networks Using Transfer Learning Scenarios.* IEEE Access, 2023. **11**: p. 20202-20215.
3. Rahman, A.A., et al., *Two decades of bengali handwritten digit recognition: A survey.* IEEE Access, 2022. **10**: p. 92597-92632.
4. Fateh, A., M. Fateh, and V. Abolghasemi, *Enhancing optical character recognition: Efficient techniques for document layout analysis and text line detection.* Engineering Reports, 2023: p. e12832.
5. Parhami, P., et al., *A comparison of deep neural network models for cluster cancer patients through somatic point mutations.* Journal of Ambient Intelligence and Humanized Computing, 2023. **14**(8): p. 10883-10898.
6. Azawi, N., *Handwritten digits recognition using transfer learning.* Computers and Electrical Engineering, 2023. **106**: p. 108604.
7. Rasheed, A., et al., *Handwritten Urdu characters and digits recognition using transfer learning and augmentation with AlexNet.* IEEE Access, 2022. **10**: p. 102629-102645.
8. Malhotra, R. and M.T. Addis, *End-to-End Historical Handwritten Ethiopic Text Recognition using Deep Learning.* IEEE Access, 2023.
9. Fateh, A., et al., *Persian printed text line detection based on font size.* Multimedia Tools and Applications, 2023. **82**(2): p. 2393-2418.
10. Fateh, A., et al., *Providing a Voting-Based Method for Combining Deep Neural Network Outputs to Layout Analysis of Printed Documents.* Journal of Machine Vision and Image Processing, 2022. **9**(1): p. 47-64.
11. Puri, S. and S.P. Singh, *An efficient Devanagari character classification in printed and handwritten documents using SVM.* Procedia Computer Science, 2019. **152**: p. 111-121.
12. Bellili, A., M. Gilloux, and P. Gallinari, *An MLP-SVM combination architecture for offline handwritten digit recognition: Reduction of recognition errors by Support Vector Machines rejection mechanisms.* Document Analysis and Recognition, 2003. **5**(4): p. 244-252.
13. Gupta, D. and S. Bag, *CNN-based multilingual handwritten numeral recognition: A fusion-free approach.* Expert Systems with Applications, 2021. **165**: p. 113784.
14. Aly, S. and A. Mohamed, *Unknown-length handwritten numeral string recognition using cascade of PCA-SVMNet classifiers.* IEEE Access, 2019. **7**: p. 52024-52034.
15. Yang, M., et al., *Sparse representation based fisher discrimination dictionary learning for image classification.* International Journal of Computer Vision, 2014. **109**: p. 209-232.
16. Sethy, A., P.K. Patra, and S.R. Nayak, *A hybrid system for handwritten character recognition with high robustness.* Traitement du Signal, 2022. **39**(2): p. 567.
17. Hanmandlu, M. and O.R. Murthy, *Fuzzy model based recognition of handwritten numerals.* pattern recognition, 2007. **40**(6): p. 1840-1854.
18. Memiş, S., S. Enginoğlu, and U. Erkan, *Numerical data classification via distance-based similarity measures of fuzzy parameterized fuzzy soft matrices.* IEEE Access, 2021. **9**: p. 88583-88601.
19. Vidhale, B., et al. *Multilingual Text & Handwritten Digit Recognition and Conversion of Regional languages into Universal Language Using Neural Networks*. in *2021 6th International Conference for Convergence in Technology (I2CT)*. 2021. IEEE.
20. Hull, J.J., *A database for handwritten text recognition research.* IEEE Transactions on pattern analysis and machine intelligence, 1994. **16**(5): p. 550-554.
21. Sufian, A., et al., *Bdnet: bengali handwritten numeral digit recognition based on densely connected convolutional neural networks.* Journal of King Saud University-Computer and Information Sciences, 2022. **34**(6): p. 2610-2620.
22. Abolghasemi, V., et al., *Incoherent dictionary pair learning: Application to a novel open-source database of chinese numbers.* IEEE Signal Processing Letters, 2018. **25**(4): p. 472-476.
23. Fateh, A., M. Fateh, and V. Abolghasemi, *Multilingual handwritten numeral recognition using a robust deep network joint with transfer learning.* Information Sciences, 2021. **581**: p. 479-494.
24. Elakkiya, R., P. Vijayakumar, and M. Karuppiah, *COVID_SCREENET: COVID-19 Screening in Chest Radiography Images Using Deep Transfer Stacking.* Information Systems Frontiers, 2021. **23**(6): p. 1369-1383.
25. B, N., E. R, and M. Prasad, *Sentence2SignGesture: a hybrid neural machine translation network for sign language video generation.* Journal of Ambient Intelligence and Humanized Computing, 2022. **14**: p. 1-15.
26. Zhang, J., et al., *Improving transferability of 3D adversarial attacks with scale and shear transformations.* Information Sciences, 2024. **662**: p. 120245.
27. Vinotheni, C. and S.L. Pandian, *End-To-End Deep-Learning-Based Tamil Handwritten Document Recognition and Classification Model.* IEEE Access, 2023. **11**: p. 43195-43204.
28. Hallur, V.C. and R. Hegadi, *Handwritten Kannada numerals recognition using deep learning convolution neural network (DCNN) classifier.* CSI Transactions on ICT, 2020. **8**: p. 295-309.
29. Kusetogullari, H., et al., *DIGITNET: A deep handwritten digit detection and recognition method using a new historical handwritten digit dataset.* Big Data Research, 2021. **23**: p. 100182.
30. Ahmed, S.S., et al., *A novel technique for handwritten digit recognition using deep learning.* Journal of Sensors, 2023. **2023**.
31. Ameri, R., et al., *Labeled projective dictionary pair learning: application to handwritten numbers recognition.* Information Sciences, 2022. **609**: p. 489-506.
32. Alkhawaldeh, R.S., *Arabic (Indian) digit handwritten recognition using recurrent transfer deep architecture.* Soft Computing, 2021. **25**(4): p. 3131-3141.
33. Jiang, W., *MNIST-MIX: a multi-language handwritten digit recognition dataset.* IOP SciNotes, 2020. **1**(2): p. 025002.
34. Gandhi, M., *Gujarati-Dataset*. 2020: https://github.com/MikitaGandhi/Gujarati-Database-.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3378598

35. Pramanik, S., *Gurmukhi-Dataset*. 2023: https://github.com/siddharthapramanik771/Gurmukhi-Handwritten-Digit-Classification.
36. Abdelazeem, S. and E. El-Sherif, *Modified Arabic Digits Database*. 2008, School of Science and Engineering, Department of Electronics Engineering ….
37. Jabde, M., et al. *Offline Handwritten Multilingual Numeral Recognition Using CNN*. in *International Conference on Information Science and Applications*. 2023. Springer.
38. Jiang, Z., Z. Lin, and L.S. Davis. *Learning a discriminative dictionary for sparse coding via label consistent K-SVD*. in *CVPR 2011*. 2011. IEEE.
39. Jiang, Z., Z. Lin, and L.S. Davis, *Label consistent K-SVD: Learning a discriminative dictionary for recognition.* IEEE transactions on pattern analysis and machine intelligence, 2013. **35**(11): p. 2651-2664.
40. Ramirez, I., P. Sprechmann, and G. Sapiro. *Classification and clustering via dictionary learning with structured incoherence and shared features*. in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2010. IEEE.
41. Gu, S., et al., *Projective dictionary pair learning for pattern classification.* Advances in neural information processing systems, 2014. **27**.
42. Wright, J., et al., *Robust face recognition via sparse representation.* IEEE transactions on pattern analysis and machine intelligence, 2008. **31**(2): p. 210-227.
43. Szegedy, C., et al. *Rethinking the inception architecture for computer vision*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
44. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
45. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition.* arXiv preprint arXiv:1409.1556, 2014.
46. El-Sawy, A., H. El-Bakry, and M. Loey. *CNN for handwritten arabic digits recognition based on LeNet-5*. in *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016 2*. 2017. Springer.

Amirreza Fateh is a PhD student in AI at IUST, specializing in Few-Shot Segmentation and Transfer Learning. With a strong background in image processing, he earned the 2nd rank in his bachelor's and the 1st rank in his master's program. Amirreza is a senior researcher dedicated to advancing the frontiers of AI, particularly in innovative applications for image processing.

Reza Tahmasbi Birgani is an undergraduate student at the University of Shahrood, pursuing a degree in computer engineering. He is also a teaching assistant at the university and a member of the research lab. In his free time, Reza enjoys reading about the latest advancements in artificial intelligence and machine learning, and he is always eager to learn more about the field.

Mansoor Fateh received the M.S. degree in Biomedical Engineering from Tarbiat Modares University, Tehran, Iran, and the PhD from Tarbiat Modares University, Tehran, Iran. He is a faculty member in the Faculty of Computer Engineering, Shahrood University of Technology, Iran. His research interests include machine learning and image processing.

Vahid Abolghasemi (Senior Member, IEEE) received the Ph.D. degree in signal processing from the University of Surrey, Guildford, U.K., in 2011. He is currently an Associate Professor with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, U.K. During 2011–2019, he held academic positions with Brunel University, Uxbridge, U.K., and Shahrood University of Technology, Shahrood, Iran. His main research interests include signal and image processing, compressive sensing, and machine learning.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4