

Received 7 September 2022, accepted 18 September 2022, date of publication 22 September 2022, date of current version 30 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3208959

RESEARCH ARTICLE

Handwritten Urdu Characters and Digits Recognition Using Transfer Learning and Augmentation With AlexNet

AQSA RASHEED^{1,2}, NOUMAN ALI¹, BUSHRA ZAFAR³, AMSA SHABBIR^{1,4}, MUHAMMAD SAJID⁵, AND MUHAMMAD TARIQ MAHMOOD⁶, (Senior Member, IEEE)

¹Department of Software Engineering, Mirpur University of Science and Technology, Mirpur-AJK 10250, Pakistan

²Department of Computer Science, University of South Asia, Lahore 54000, Pakistan

³Department of Computer Science, Government College University, Faisalabad 38000, Pakistan

⁴Department of Software Engineering, Capital University of Science and Technology (CUST), Islamabad 45750, Pakistan

⁵Department of Electrical Engineering, Mirpur University of Science and Technology, Mirpur-AJK 10250, Pakistan

⁶Future Convergence Engineering, Korea University of Technology and Education, Cheonan 31253, Republic of Korea

Corresponding author: Muhammad Tariq Mahmood (tariq@koreatech.ac.kr)

This work was supported by the Education and Research Promotion Program of KoreaTech (2022).

ABSTRACT Automated recognition of handwritten characters and digits is a challenging task. Although a significant amount of literature exists for automatic recognition of handwritten characters of English and other major languages in the world, there exists a wide research gap due to lack of research for recognition of Urdu language. The variations in writing style, shape and size of individual characters and similarities with other characters add to the complexity for accurate classification of handwritten characters. Deep neural networks have emerged as a powerful technology for automated classification of character patterns and object images. Although deep networks are known to provide remarkable results on large-scale datasets with millions of images, however the use of deep networks for small image datasets is still challenging. The purpose of this research is to present a classification framework for automatic recognition of handwritten Urdu character and digits with higher recognition accuracy by utilizing theory of transfer learning and pre-trained Convolution Neural Networks (CNN). The performance of transfer learning is evaluated in different ways: by using pre-trained AlexNet CNN model with Support Vector Machine (SVM) classifier, and fine-tuned AlexNet for extracting features and classification. We have fine-tuned AlexNet hyper-parameters to achieve higher accuracy and data augmentation is performed to avoid over-fitting. Experimental results and the quantitative comparisons demonstrate the effectiveness of the proposed research for recognition of handwritten characters and digits using fine-tuned AlexNet. The proposed research based on fine-tuned AlexNet outperforms the related state-of-the-art research thereby achieving a classification accuracy of 97.08%, 98.21%, 94.92% for Urdu characters, digits and hybrid datasets respectively. The presented methods can be applied for research on Urdu characters and in diverse domains such as handwritten text image retrieval, reading postal addresses, bank's cheque processing, preserving and digitization of manuscripts from old ages.

INDEX TERMS Automated recognition, Urdu HCR systems, CNN, transfer learning, AlexNet, SVM, optical character recognition.

I. INTRODUCTION

Digital image processing plays a vital role in different computer vision based applications such as image retrieval, medical image analysis, face recognition, decision support

The associate editor coordinating the review of this manuscript and approving it for publication was Jeon Gwangil.

systems with industrial applications, object recognition and image annotation [1], [2], [3], [4], [5], [6]. The recent massive growth in the application of mobile and computing devices, has increased the implications of Character Recognition (CR) [7]. Recognition of handwritten text is problematic because of the fact that writing styles differ from individual to individual.

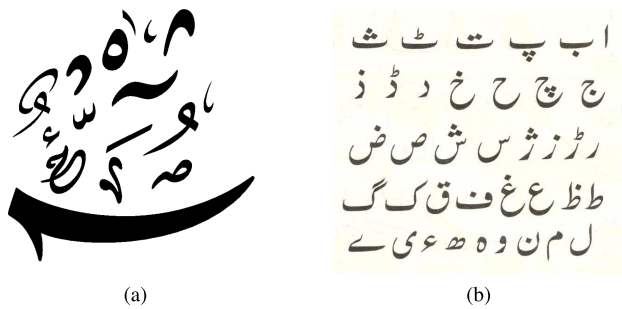


FIGURE 1. Example images from (a) The basic characters of Urdu language (b) Diacritics: The secondary characters of Urdu.

The Urdu language is the national language of Pakistan, and is extremely important as it is one of the largest languages of the world and is spoken by more than 60 million people (and more than 329 million people are much the same in spoken form when paired with Hindi). Other than this it is a mixture of Turkish, Persian and Arabic languages. Urdu character recognition has still problems due to its language complexity. The Figure 1(a) shows the 38 basic characters of Urdu language. Despite being one of the world's largest languages, unfortunately, a little amount of work has been done for recognition of handwritten Urdu characters. The presence of diacritics make Urdu CR more challenging as compared to English and many other languages. The complexity of Urdu languages is due to following things [8], [9]:

- **Cursive:** By nature Urdu is cursive language. The writing style which combines the words together makes it more complex.
- **Diacritics:** These are the secondary characters of Urdu and used above or below main characters such as dots, diagonals, madaaaa, hamza etc. The Figure 1(b) shows the diacritics used in Urdu language.
- **Script:** There are 12 different scripts in Urdu language. Mostly recognition techniques are script specific which means it will not work for recognition of other scripts [9].
- **Writing direction:** Mostly languages in the world are unidirectional but Urdu is bidirectional language.
- **Strokes:** According to the rule in Urdu there should be zero or one main stroke or one secondary stroke.

The above mentioned characteristics are some of the basics things that make the character recognition of Urdu language more difficult as compared to other languages.

The task of the Handwritten Character Recognition (HCR) is intuitive, that inputs a digitized image and a desired character is given as output. An automated CR system improves the efficiency as compared to manual recognition by human workers. As compared to Optical Character Recognition (OCR), handwritten CR is more challenging since the writing styles vary among different individuals. Recognition of handwritten text is an interesting task due to its vast uses and application in image processing, such as

translating handwritten records into a digital format, OCR (Optical Character Recognition), Urdu machine transliteration, integration with other languages, image restoration, automatically reading postal address, house numbers and robotics [7], [10], [11], [12], [13].

One of the main issues in the classification of handwritten characters is the massive variety in the types of handwriting by various peoples in distinct languages, that the recognition system has to deal with. The variations in writing style, shape and size of individual character and similarities with other characters makes the hand written recognition add to the complexity. Machine learning and deep Neural Network techniques have been widely used for automatic recognition of characters and digits of different languages, and in various classification-based problems [14], [15], [16], [17].

This research aims to apply pre-trained Convolution Neural Network (CNN) approach for recognition of handwritten Urdu characters, since a little amount of work has been done in literature so far in this direction. The pioneer dataset used in this research was introduced in 2020 [1], in which the authors applied unsupervised algorithm called autoencoder and CNN for recognition of Urdu handwritten characters. AlexNet is one of the simplest deep learning model and has shown commendable performance in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), in the past few years. The distinguishing characteristics of AlexNet as compared to other deep learning models are: having much more filters in each layer, pooling layer in addition to stacked convolutional layers, faster computing time and limited hardware dependency [18]. In the proposed research, we have proposed two frameworks for classification of hand-written characters and digits using pre-trained AlexNet neural network. The performance of the proposed research is evaluated on recently introduced dataset for handwritten characters and digits of Urdu [19].

The main contributions of this research are as follows:

- First we applied the pre-trained CNN AlexNet as the basic transfer learning model and used the extracted features to train the SVM classifier. We tested different transfer configurations to obtain the optimal classification performance for Urdu characters and digits recognition.
- Second, fine-tuning of the pre-trained CNN AlexNet hyper-parameters and data augmentation is applied to memorize the exact details of training images and avoid overfitting. Transfer learning is applied to transfer the layers to the new classification task.
- Quantitative comparison between the classification performance for the pre-trained CNN AlexNet using the SVM classifier and transfer learning from the fine-tuned AlexNet for feature extraction and classification is presented.

The rest of the article is organized as follows: literature review covering the current state-of-the-art is presented in Section II. Section III describes the architectural details of the pre-trained CNN AlexNet model and section IV presents

the details of the research methodology. Section V covers the details of the dataset used and the experimental results and finally; Section VI concludes the paper.

II. RELATED WORK

This section gives a short review of the related work, thereby covering the handwritten character recognition approaches based on machine-learning and deep-learning architectures. Machine learning models differ from deep neural networks in that they are considered as “shallow” models that attempt to learn patterns from data. Different machine learning models have been used in literature for recognition and classification of digital images [20], [21]. In [22], k-nearest neighbor classifier has been used, that uses a hybrid feature vector of both statistical and structural features of MNIST achieving an accuracy of 98.42%. In [23], the authors have proposed to combine multiple SVM classifiers in order to enhance the accuracy of a single SVM classifier.

Deep neural networks have been used extensively in literature for character recognition. The authors in [24], used neural networks for recognition of characters and concluded that varying hand writing patterns of humans can't be fully recognized with one network. Authors used different sized grids to try to recognise printed and handwritten characters (5×7 , 7×11 , 9×13). Despite differences in character orientation, size, and location, the network maintained a precision of 60% [24]. Xiao *et al.* [25] developed the algorithm, which may cut the network's computational burden by nine times and condense the network to 1/18 of the actual state of the baseline model with only a 0.21% loss in accuracy. Li *et al.* [26] proposed CNN based architectures for Chinese character recognition and tests are carried on the ICDAR-2013 offline HCCR dataset, and the suggested approach takes only 6.9 milliseconds on average to classify a character image and achieves 97.1 % accuracy while requiring only 3.3 megabytes of storage.

Ahmad *et al.* [27] applied a Stacked Denoising Autoencoder (SDA) for offline Urdu CR. Pre-training has been performed using unsupervised way and fine tuning has been done with supervised then the final trained and tuned network is used to recognize the Urdu ligatures. The network SDA is trained on 178573 ligatures. The testing has been performed on Urdu Printed Text Image (UPTI) dataset and for comparison the SVM classifier has been trained and tested with the network. The authors concluded that the use of hand-crafted features is error prone and the application of resulted in better outcomes as compared to the traditional approaches. The results shows that the proposed Stacked Denoising Autoencoder (SDA) has higher accuracy rate than Support Vector Machine (SVM). For such a big dataset of ligatures, the experimental findings show accuracies in the range of 93% to 96%, which are better than previous Urdu OCR (Optical Character Recognition) systems [27]. However, the scope of study in [27] was limited to OCR of Urdu Nastaleeq font only. The authors in [28], presented an offline OCR (Optical Character Recognition) system for recognition of

eight handwritten Arabic characters and achieved an accuracy of 77.25%.

Elenwar *et al.* [29] proposed a framework for Arabic CR (Character Recognition) utilizing rule-based methods for segmentation and recognition of word portions. The authors also introduced a technique for separation of text line. The Top50 list members correctly choose the test strokes 92% of the time and 95% of the time for the test characters [29]. Khan *et al.* [30] proposed an approach based on wavelet transform and back propagation neural network classifier for handwritten CR of Urdu Nastalique font. However, the database used in [30] is of low generalization as it was prepared by only four writers. AlKhateeb [31] presented an Arabic characters database comprising of 28 thousand Arabic characters and the samples were collected by 100 different writers.

In [32], the authors presented an deep neural network based approach for recognition of Farsi handwritten characters. The proposed system for recognising Farsi handwritten phone numbers is achieved an accuracy of 94.6 %. The proposed method can search the phone book after recognising the digits of the phone number [32]. In another recent work, Khan *et al.* [33] proposed an (Optical Character Recognition) OCR system for recognition of Pashto characters and developed a database comprising of 4488 handwritten Pashto characters. The proposed experiments reveal that the support vector machine, artificial neural network, and proposed OCR system have accuracy of 56%, 78%, and 80.7%, respectively [33].

Azad *et al.* [34] introduced an auto-encoder with deep CNN (which they call DConvAENNet) for recognizing handwritten Bangla characters. The suggested DConvAENNet model scored 95.21% on BanglaLekha-Isolated for 84 classes, 92.40% on CMATERdb 3.1 for 238 classes, and 95.53% on Ekush for 122 classes using this supervised and unsupervised learning technique [34]. Husnain *et al.* [35] performed a survey on off-line Urdu hand written text recognition. According to the [35], Urdu is derived from Arabic language and it is bidirectional and cursive in its nature so it (Urdu) has almost same challenges as Arabic have for recognition purpose but Urdu has higher complexity. Authors presented survey of articles published from 2004-2019 with 04 levels which are: character, word, ligature, and sentence level [35]. All public available handwritten datasets are highlighted and some comparison has been performed with commonly used datasets and methods.

Naz *et al.* [36] surveyed cursive language like Urdu for OCR (optical character recognition). According to [36], many similar languages such as Sindi, Pashto, and Urdu are interest of researchers but with Nastaliq and Naskh scripts. The literature survey has been discussed on focusing the major steps: a) pre-processing, b) segmentation, c) feature extraction, d) classification and e) recognition for printed, handwritten and online character recognition. Authors performed a survey and reported accuracy and a critical analysis of methods [36].

Chen *et al.* [37] proposed a method multilingual text recognition framework for script identification and handwriting recognition. Spatial and temporal knowledge are used to perform feature extraction to encode the input into features [37]. The proposed network has the advantage that it can be beneficial for both two multilingual schemes via multi task learning. Testing has been performed on 5 different languages English, Kannada, Urdu, French, and Bangla [37]. The results demonstrated that the proposed framework has 99.9% script identification accuracy rate and system outperformed for handwriting recognition.

Naz *et al.* [38] proposed a hybrid approach based on convolutional-recursive deep learning which is combination of CNN (Convolutional Neural Network) and MDLTSM (Multi-dimensional Long Short-Term Memory). Neural Networks is used for Urdu Nastaliq recognition. Naz *et al.* extracted features by single layer of CNN and uses its 6 filters to filter with contoured image. Then extracted and contoured image are used with some weights as an input to MDLTSM. Features are recursively mapped to lower space dimension by every neuron then the resultant feature vector has been done for output layer [38]. Experiments are performed for evaluation purpose on UPTI (Urdu Printed Text Image) dataset and comparison of result has been performed with state-of-the-art. The reported recognition rate is 98.12% [38].

In most of the literature reviewed so far related to Urdu script recognition is based on small datasets which restricts its generalization capability. The existing research efforts for Urdu CR(Character Recognition) are for printed text (typically OCR based applications) [27], [28], and only a limited amount of work is found for handwritten Urdu CR. In [19], Ali *et al.* presented a new dataset comprising of handwritten digits and characters of Urdu, and the samples were collected from 900 different individuals. The dataset was passed through different pre-processing stages like conversion from RGB to grey-scale, segmentation and removal of noise. The authors used an unsupervised deep learning algorithm called auto-encoder and CNN for handwritten Urdu CR for the first time.

For the recognition of handwritten Urdu characters and numerals, Mushtaq *et al.* [39] presented a Convolutional Neural Network (CNN) architecture. CNN is an unique image recognition technique that does not require explicit feature engineering and extraction and offers more efficient results than traditional handmade feature extraction approaches, with a recognition rate of 98.82% [39]. Naeem *et al.* [40] proposed a CNN-RNN architecture for the recognition of handwritten urdu characters and reach the required character error rate of 5.28%. In [41], authors presented his study for optical character recognition of Nastaliq Urdu like script language and proposed deep extreme learning machine based machine.

The most of the literature work and research in this domain is done on machine learning approaches such as Ahmad *et al.* [27] applied a Stacked Denoising Autoencoder (SDA), Elenwar *et al.* [29] utilized rule-based methods for

segmentation and recognition, Khan *et al.* [30] proposed an approach based on wavelet transform, Azad *et al.* [34] introduced an auto-encoder, Chen *et al.* [37] proposed a method based on script spatial and temporal knowledge and many others with comparatively low performance as compared to recent trends of deep learning.

Deep learning has overtaken traditional machine learning as the method of choice for the majority of AI-related challenges over the past several years. Deep learning has repeatedly shown its better performance on a number of tasks, including speech, natural language, vision, and playing games. This is the obvious cause for this.

Compared to traditional Machine Learning (ML) methods, deep learning approaches can be applied to a variety of domains and applications far more simply. First, using pre-trained deep networks for various applications within the same domain is now efficient with transfer learning.

For instance, in computer vision, object recognition and segmentation networks frequently use feature extraction front-ends that were trained on pre-trained image classification networks. The full model's training is facilitated by using these pre-trained networks as front-ends, which frequently leads to better performance in a shorter amount of time.

Additionally, deep learning's fundamental principles and methods are frequently extremely portable across fields. For instance, since the fundamental concepts are relatively similar, understanding how to apply deep networks to the field of natural language processing isn't too difficult once the underlying deep learning theory for the domain of speech recognition is understood. This isn't the case with conventional ML (Machine Learning) at all, as feature engineering and domain- and application-specific ML techniques are needed to create high-performance ML models. Depending on the topic and application, the knowledge base of classical ML differs significantly and frequently necessitates in-depth specialist study in each field.

Other than this, deep learning based approaches are now in trend and are being used for numerous application such as Cao *et al.* [42] used deep clustering networks with feature learning enhancements and Lui *et al.* [43] used similar approach for pedestrian movement modelling.

Now-a-days handwritten character recognition is gaining importance and interest of the researchers [44], this interest is not only limited to the Urdu handwritten character recognition but also for other languages as well such as arabic, farsi etc. [45]. There exists a wide research gap for recognition of handwritten Urdu characters. This research is of significant importance because of its vast applications as urdu sign board recognition and detection [46], in product manufacturing [47] and automated recognition system [48]. In this work, we aim to use deep features for recognizing handwritten urdu characters and digits, and determine the performance of pre-trained deep AlexNet features with SVM classifier and transfer learning. The presented methods can serve as a baseline in research on character recognition and in diverse related domains.

III. AlexNet: DEEP LEARNING MODEL

CNN is a deep neural network and is being widely used for image analysis based tasks. It allows the extraction of a large number of features from the input image set [49]. Unlike other models, CNN takes input data, perform training, feature extraction and automatically classifies the data into the desired output. Mainly the CNN is comprised of input and output layers with many hidden layers between them. The major principals used in CNN are: convolution, activation and max-pooling. Now-a-days, deep learning models are widely used in image processing and computer vision as have many advantages over traditional and classical methods. In our research, deep learning based AlexNet method is used to recognize urdu handwritten characters and digits. The major reason to choose this model is that it has ability to train the model on huge hyper-parameters. Beside this, AlexNet can classify objects up to 1000 categories, leverages GPU for training, reduces the computational complexity and has better performance over traditional methods [50].

The model is trained on over a million images and objects [50]. It can easily categorize images into 1000 classes (such as mouse, plants tiger, cat, and many animals). The model has learned rich feature representations for a large number of images. It is comprised of eight layers where first five layers are convolutional (CN) layers followed by three max pooling (M-POOL) layers so the proposed network is deeper than the common CNN networks. To avoid over fitting the 0.5% dropout ratio is used and applied on fully connected layers. The architecture of the AlexNet is shown in Figure 2. The architecture comprised of the following elements:

- CONV1 with 11×11 kernel size
- Rectified Linear Unit Layer Activation (ReLU)
- Response Normalization Layer, Maximum Pooling (4×4 kernel)
- CONV2 with 5×5 kernel size
- Rectified Linear Unit Layer (ReLU)
- Response Normalization Layer
- Maximum Pooling (3×3)
- CONV3 with 3×3 kernel size
- Rectified Linear Unit Layer Activation (ReLU)
- CONV4 with 3×3 kernel size
- Rectified Linear Unit Layer Activation (ReLU)
- CONV5 with 3×3 kernel size
- Rectified Linear Unit Layer Activation (ReLU)
- Maximum Pooling (3×3)
- FC6 (Fully Connected) Layer(4096 nodes)
- Rectified Linear Unit Layer Activation (ReLU)
- FC7 (Fully Connected) Layer (4096 nodes)
- Rectified Linear Unit Layer (ReLU)
- FC8 (Fully Connected Soft-max out Layer).

In this work, the dataset images are preprocessed for input image layer to the AlexNet. In preprocessing, the images are resized from 28×28 to 227×227 and input images are converted from gray scale to color channel (RGB-conversion) to make them suitable for AlexNet because the network

process 3 channel images. The proposed framework comprised of 5 convolutional layers, followed by 3 Max Pooling (M-POOL) and RELU layers. Consider an input image I , of size $W \times H \times C$ is subjected to a CONV L_i with K square kernel size and M output maps. Then $N_i = WHM$, $P_i = K^2CM$ and $U_i = WHK^2CM$ are the number of output units, weights (parameters) and connections respectively in case of CONV layers. When they are subjected to FC layers, $N_i = WHM$, $P_i = K^2H^2CM$ and $U_i = W^2H^2CM$ are the number of output units, weights (parameters) and connections respectively in case of CONV layers. For CONV1 layer, there are 96 kernels (output channels) each of size $11 \times 11 \times 3$ with Stride 4 the input W and H shrink by a factor of 4. The convolutional is defined as represented in Equation 1 for the image I with (i,j) dimension. Where G is the feature map and F is the convolution filter [51].

$$G(i, j) = I \times F(i, j) = \sum_x \sum_y I(i-x, j-y)F(x, y) \quad (1)$$

Then activation function is performed by ReLU layers where the important purpose of rectified linear units is to shaped the deep model into linear structure. The negative values are bring to zero and the mathematical representation of ReLU function is shown in Equation.

$$ReLU(x) = \max(x, 0) \quad (2)$$

After ReLU pooling is applied to reduce the number of features, which actually reduces the sizes of the input image which is sent to the next convolutional layer. The activation process by ReLU and pooling is repeated after every convolutional layer. For CONV2 layers, there are 256 kernels (output channels) each of size 5×5 [52]. For next two CONV layers, there are 384 kernels (output channels) each of size 3×3 . The last CONV layers has 256 kernels (output channels) each of size 3×3 . The summary of the units, weights and connections of the proposed method is presented in the Table 1. The network is comprised of 660K units, 600M connections and 61M parameters. It can be seen that there are larger number of connections and hyper-parameters in convolutional layers but for weights FC layers are responsible. Feature extraction is performed in CONV layers which in result generate feature maps which are passed to FC. Further, they are subjected to Softmax layer for the classification purpose. The last layer (Softmax) can easily classify images or objects into 1000 and more object categories but here in our case we have modified them accordingly to our datasets. In case of characters, there are 40 classes. In digits there are 10 classes and hybrid dataset consist of total 50 classes.

IV. RESEARCH METHODOLOGY

The main objective of this study was to analyze the performance of pre-trained CNN AlexNet for handwritten urdu digits and characters recognition. We have followed two approaches for our system as shown in Figure 3.

TABLE 1. Summary of the units, weights, connections and detail of AlexNet.

Layers	Units	Weights	Connections
CONV1 $96 \times 11 \times 11 \times 3$ convolutions with stride and padding	290,400	34,848	105,415,200
CONV2 $256 \times 5 \times 5 \times 48$ convolutions with stride and padding	186,624	307,200	111,974,400
CONV3 $384 \times 3 \times 3 \times 256$ convolutions with stride and padding	64,896	884,736	149,520,384
CONV4 $384 \times 3 \times 3 \times 192$ convolutions with stride and padding	64,869	663,552	112,140,288
CONV5 $256 \times 3 \times 3 \times 192$ convolutions with stride and padding	43,264	442,368	74,760,192
FC6 4096 fully connected neurons	4096	37,748,736	37,748,736
FC7 4096 fully connected neurons	4096	16,777,216	16,777,216
FC8 1000 fully connected neurons	1000	4,096,000	4,096,000
Total	659,272	60,954,656	612,432,416

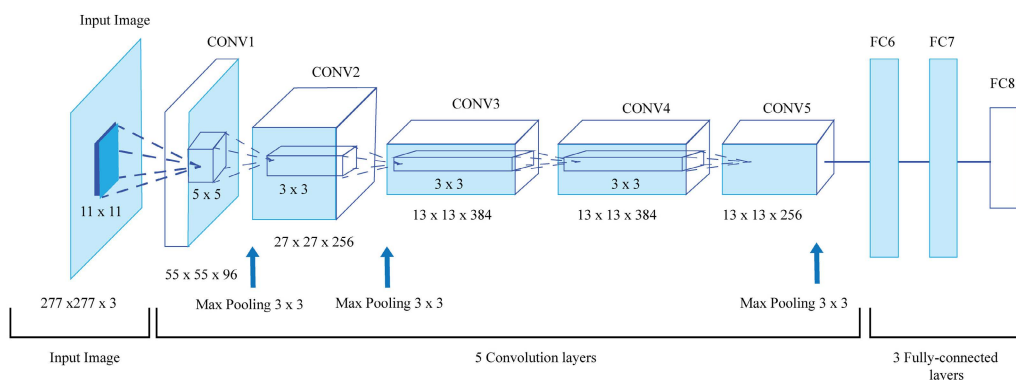


FIGURE 3. Overview of the research methodology.

Transfer Learning: Transfer Learning (TL) is a technique that is based on deep learning where we can use pre-trained network to learn a new task and use it as our starting point for a new domain or task. Tuning a network with TL is fast and easy than training a network from scratch with randomly initialized weights. With a lower amount of training pictures/images, you can rapidly transfer learned characteristics to a fresh or new domain or task. The examples of this approach are object detection, speech and image recognition and others. TL (Transfer Learning) is a popular machine learning approach because:

- By leveraging the models that have already been tested on big datasets, it allows to train model using comparatively little labeled information.
- Performing TL, there is no need to train for many epochs (full training cycle of entire training dataset). It reduces the training time.

The main purpose of this research is to present a technique for image classification in a way to save time and to overcome issue of isolated learning; since it is an issue with the traditional machine learning based approaches. Isolated learning does not take into account any other relevant information or previously acquired knowledge. The main issue with this type of isolated learning is that it lacks memory. It does not store previous knowledge and apply it to future learning. As a

result, in order to learn successfully, a huge number of training instances are required. To save time and overcome isolated learning problem we used an approach based on transfer learning with a deep neural network model named AlexNet with data augmentation. AlexNet is pre-trained model and has the ability to easily categorize objects into 1000 categories. The main reason to use data augmentation with AlexNet is to reduce over fitting and to manage the learning capacity of the Neural Network (size of the NN).

In transfer Learning approach, the basic objective is to learn the conditional probability distribution in new or target domain with the knowledge learned from old or source domain and from old or source task [55]. Formally it can be represented as, for a Task $T = L, p(*)$ which can consist of a label space L and related predictive function $p(*)$ which can be learned from existing or from source domain (training data) then the probability distribution of task T can be written as [56] and [55]:

$$T = L, P(L|F) \tag{3}$$

where L is a label space, $p(*)$ is the predictive function, F is the feature space.

The images of the dataset Urdu Characters, Digits and Hybrid are divided into 70:30 ratio for training and testing purpose. The conversion from gray scale to red, green

and blue channel is performed as a preprocessing. Then the images are subjected to AlexNet model for feature extraction. We have extracted features from CONV5 (fifth convolution layer), FC6, FC7 and FC8. The resultant feature map of extracted from each mentioned layer is subjected to SVM for training and classification. The feature extraction from CONV5, FC6, FC7 and FC8 is done for characters, digits and as well as hybrid dataset. The pictorial representation of AlexSVM is shown in Figure 4 and the settings used for feature extraction is represented in Figure 5. It can be clearly observed that the setting 1 means that the layers from first to setting 1 are transferred and remaining are replaced. Similarly, in the settings 2, 3 and 4 the layers are transferred from first to the setting point respectively and remaining layers are replaced accordingly.

B. AlexFT

AlexNet model is used for the recognition of urdu characters and digits. We have fine-tuned the AlexNet (AlexFT) model as shown in Figure 6. In our research work, five CONV, ReLU and response normalisation layers are used for maximum feature extraction from input images and to train the dataset with highest accuracy. The input dataset is divided into 70:30 ratio then images are resized and converted from gray scale to color channel. The pre processed images the fed into the network. The first five layers of the pre-trained model are used for maximum feature extraction and its last three layers are replaced. The network is fine-tuned to get highest accuracy. Fine-tuning of the model is done by setting the training hyper-parameters. Extensive experiments are performed with fine-tuned network by using Stochastic Gradient Descent with Momentum (SGDM) and different batch sizes.

The SGDM is used to minimize the loss function. As the algorithms SGDM updates the bias and weights the hyper-parameters of the network model by incremental moves in the direction of a negative loss gradient [57]. In SGDM, one step iteration increment towards the minimization of loss function is done by mini-batch. The size used of mini-batch which is used in each training iteration is called Mini-Batch Size and it is used for evaluation of updation of hyper-parameters used and gradient of loss function. The momentum added provide the advantage to lessen the oscillations. The Stochastic Gradient Descent with Momentum (SGDM) uses the same learning rate for whole hyper-parameters. The value of the learning rate can be constant and varies. If we choose very small value then the training process will be completed in a very long time and if we take its value high then it may diverge the results or sub-optimize. To limit the length of the training process, the maximum number of epoch are used. The extensive experiments have been performed to find out the optimal epoch size. The full pass of the training process on whole training dataset is known as Epoch. The training progress on Digits dataset with fine-tuned network is shown in Figure 7 (a) and Figure 7 (b) show the representations used in training graph.

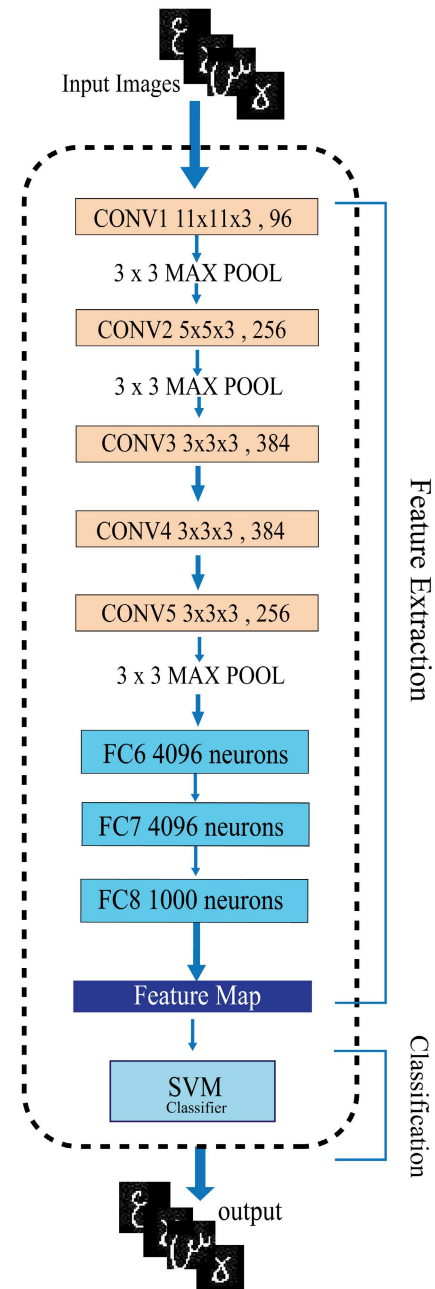


FIGURE 4. AlexSVM: AlexNet convolutional neural network with SVM.

We have changed the fully connected layers to the same size as to the number of classes in dataset for example we have set it 40 for urdu characters, 10 for digits and 50 for hybrid dataset. For fully connected layers, the values of “*WeightLearnRateFactor*” and “*BiasLearnRateFactor*” are increased to speed up the learning process in new layers than the transferred layers. The following hyper-parameters are tuned during the research as shown in Table 2.

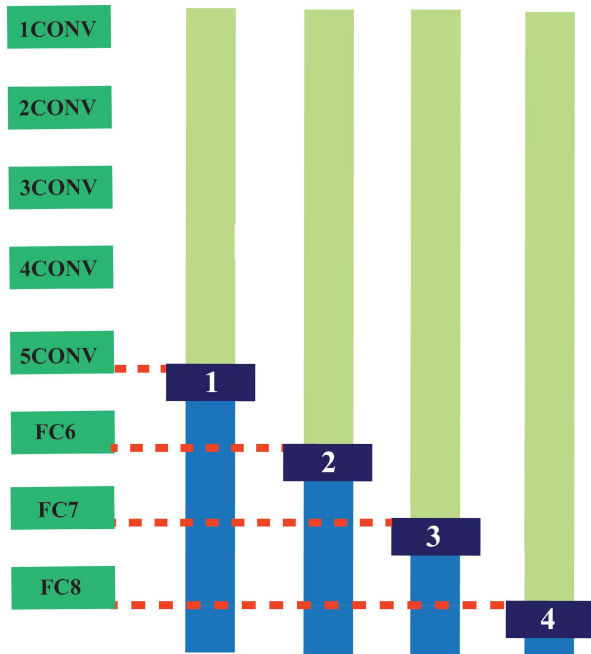


FIGURE 5. Settings used with SVM for feature extraction.

TABLE 2. Details about AlexNet hyper-parameters.

Parameter	Value
Epochs	20
Validation step	30
Optimizer	SGDM (Stochastic Gradient Descent Momentum)
Learning rate	1×10^{-3}
Decay	Default
Momentum	Default

Data Augmentation: In our approach, we used AlexNet model with data augmentation so that there will be lowest possible test error on real test data to make our model more generalized. Formally for a task say image classification, if we split our data into training data and test data then it can be written as Equation 4 and Equation 5 respectively for training and testing set.

$$(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)}) \quad (4)$$

$$(x_{testing}^{(1)}, y_{testing}^{(1)}), (x_{testing}^{(2)}, y_{testing}^{(2)}), \dots, (x_{testing}^{(m)}, y_{testing}^{(m)}) \quad (5)$$

With augmentation, our training set will become

$$BeforeAugmentation = \sum (x^{(i)}, y^{(i)}) \quad (6)$$

$$WithAugmentation = \alpha (x^{(i)}, y^{(i)}) \quad (7)$$

$$AfterAugmentation = \alpha \sum (x^{(i)}, y^{(i)}) \quad (8)$$

where $i = 1, 2, 3, \dots, m$

Where α is the constant number by which a training set samples is increased.

The suggested technique demonstrates how to fine-tune a pre trained CNN for classification task on a new set of

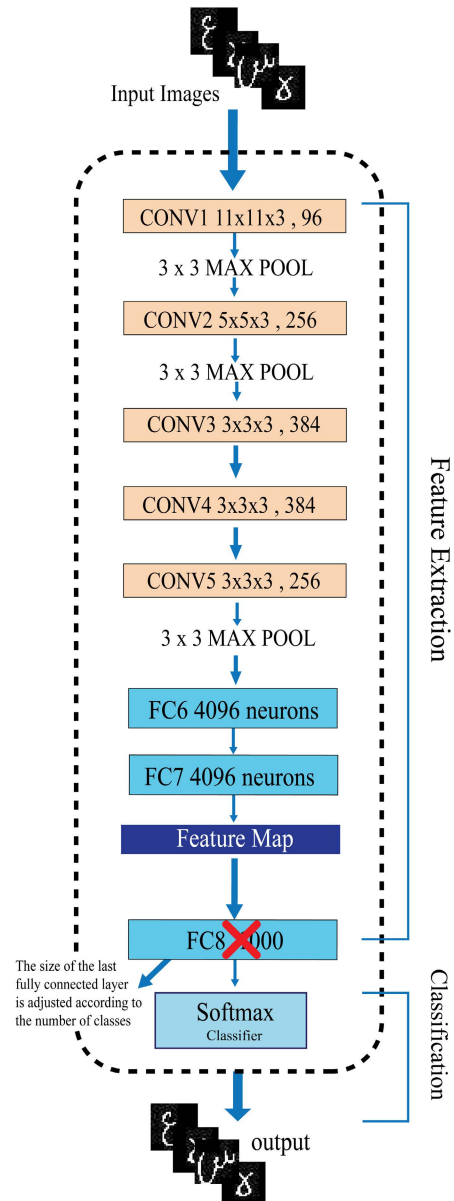
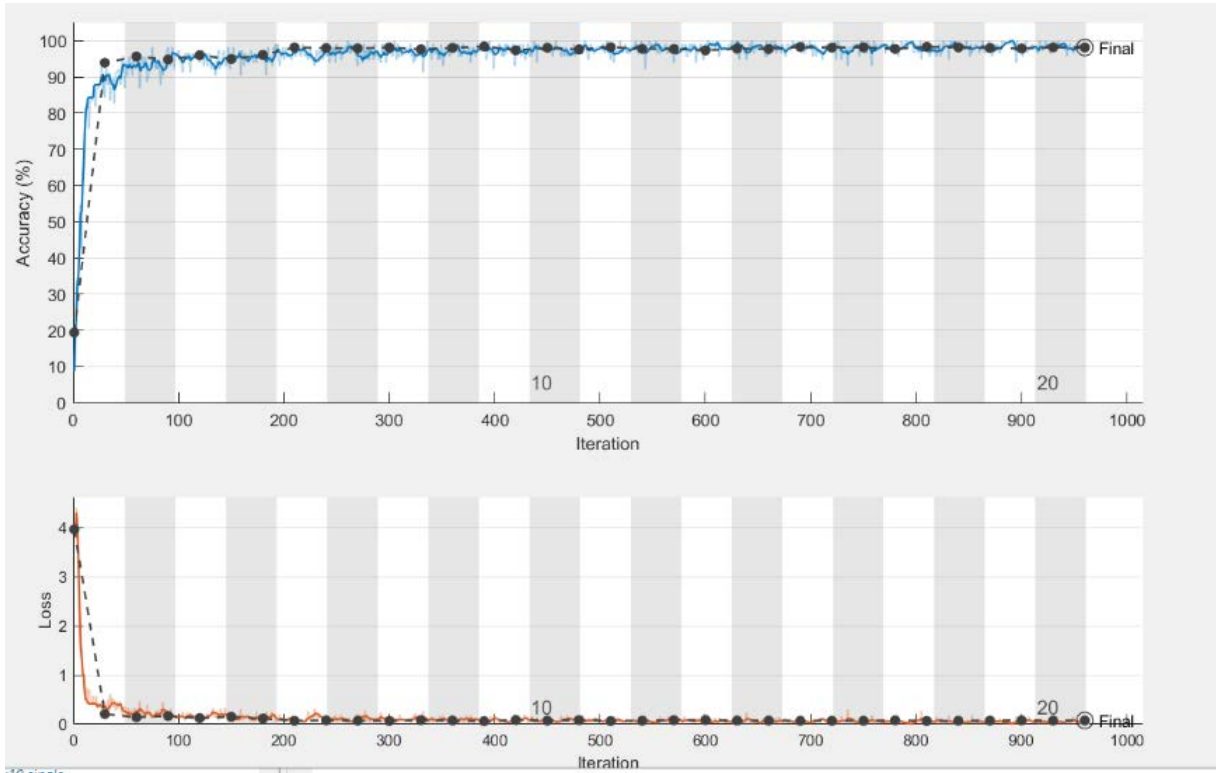


FIGURE 6. Diagrammatical representation of AlexFT.

images. Deep learning systems often employ transfer learning. You may use a pertained network as a launching point for learning a new mission. Transfer learning makes fine-tuned network much quicker and simpler than training a network from scratch with randomly initialised weights. With a smaller number of training images, you can easily pass learned features to a new task.

To prevent over fitting and to generalize the network well data augmentation is done. It helps prevent the network from memorizing the exact details of the training dataset [50]. AlexNet provide many options to perform augment operation on the training set: randomly translate upto specified pixel range default 30 vertically and horizontally, randomly flip



(a)

Accuracy



Loss



(b)

FIGURE 7. AlexFT (a) Training progress on Digits dataset with fine-tuned AlexNet (b) Representation used in training progress.

along vertical axis and other. We have augmented the dataset with the following augmentation operations: $RandRotation = [0, 20]$, $RandYShear = [0, 20]$, $RandXShear = [0, 20]$, $RandXTranslation = [-3, 3]$, and $RandYTranslation = [-3, 3]$. The results of these augmentation operations are shown in the following Figure 8 for the character dataset.

V. EXPERIMENTS AND RESULTS

This section provides the implementation details of the proposed research. The UHaT (Urdu Handwritten text Dataset) [19] is used for evaluation of the proposed research.

The UHaT dataset have handwritten characters and digits of Urdu language. There are 40 classes of characters and 10 classes of digits. The dataset is made by Hazrat Ali and is written by 900+ individuals [19]. The resolution of all images is 28×28 . The dataset is publicly available. Class representatives from digits and characters categories are shown in Figure 9.

As a first step, the image dataset is randomly split into a training and test subsets using a ratio of 70:30. The training images are used to train the classifier and test images are used for validation. The deep features are extracted using

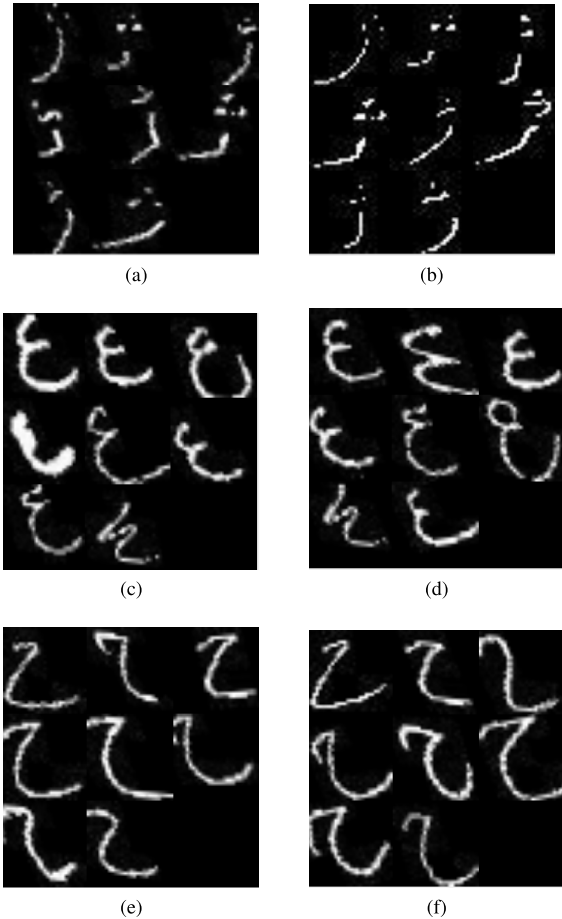


FIGURE 8. Urdu character images after augmentation operations ((a, b) For urdu character 'Zaa' (c, d) For urdu character 'Ain' (e, f) For urdu character 'Haa').

the pretrained AlexNet model. During preprocessing, all the images are resized to 227×227 according to the required input size of AlexNet network.

Hardware Specifications: MATLAB R2020a is used for implementation on Microsoft Windows 10. Experiments are conducted on Intel(R) Core(TM) i7-9750H CPU rate 2.60GHz with installed RAM 16.0 GB and NVIDIA GPU 8 GB RTX 2070 is used.

A. EVALUATION METRICS

The selection of the performance metrics depends on the nature of the application and user requirement or choice [6], [58]. The metrics used for evaluation of the proposed research are:

[I] Recognition Accuracy: The percentage of correct classification of test dataset is known as recognition accuracy. Ratio of correctly characters to the total number of characters is also known as recognition accuracy.

$$Accuracy = ACC = \frac{Tp + Tn}{Tp + Tn + Fn + Fp} \quad (9)$$

[II] Recall: It is defined as the ratio of correctly recognized images to the number of relevant images in dataset.

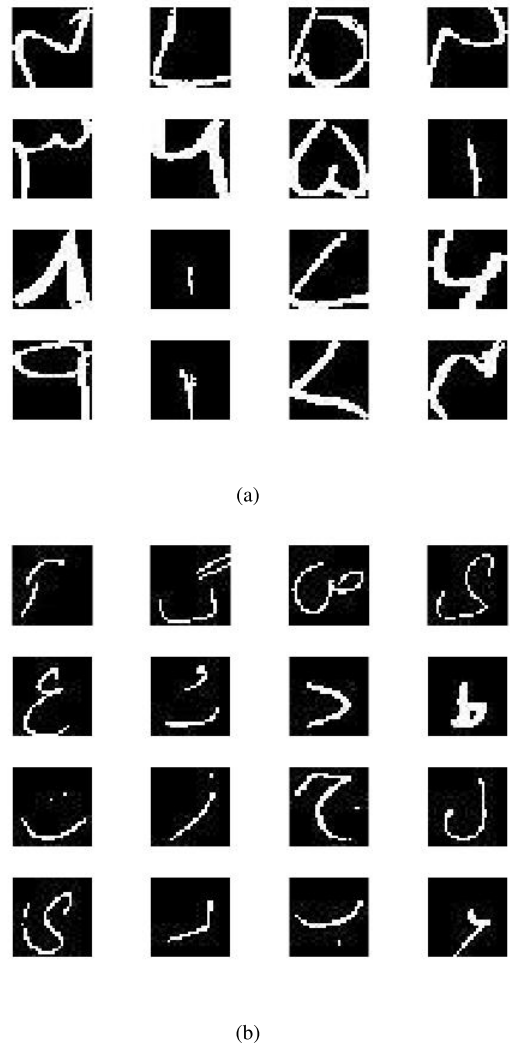


FIGURE 9. Example images from (a) Urdu Digits dataset (b) Urdu Character dataset.

Also referred as sensitivity and it is calculated as

$$Recall = Sensitivity = \frac{Tp}{Tp + Fn} \quad (10)$$

[III] Precision: It is also known as positive predictive value (PPV) and it is determined as

$$Precision = \frac{Tp}{Tp + Fp} \quad (11)$$

[IV] Error Rate (ER): It is determined as total number all incorrect predicted images to the total number of images in dataset.

$$ER = \frac{Fp + Fn}{Tp + Tn + Fn + Fp} \quad (12)$$

[V] F-measure: It is basically harmonic mean of recall and precision.

$$F - score = \frac{2(Precision \times Recall)}{Precision + Recall} \quad (13)$$

TABLE 3. Basic confusion matrix.

		Predicted Case	
		Negative	Positive
Actual Case	Negative	Tn = True Negative when model correctly predicts the negative case	Fp = False Positive when model incorrectly predicts the positive case
	Positive	Fn = False Negative when model incorrectly predicts the negative case	Tp = True Positive when model correctly predicts the positive case

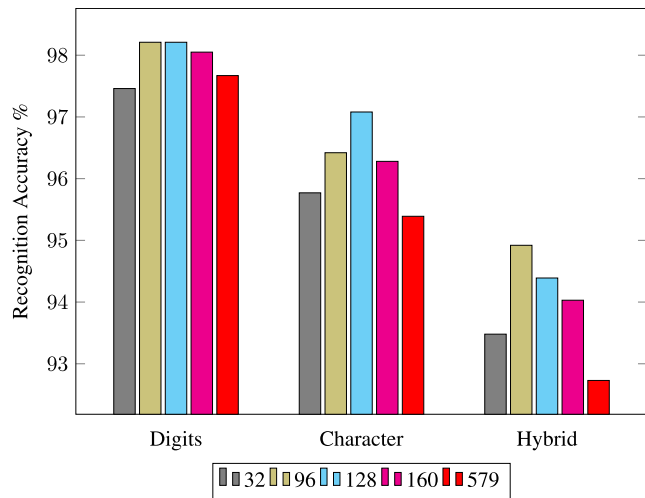


FIGURE 10. Impact of batch size on recognition accuracy of Digits, Character and Hybrid datasets using AlexFT.

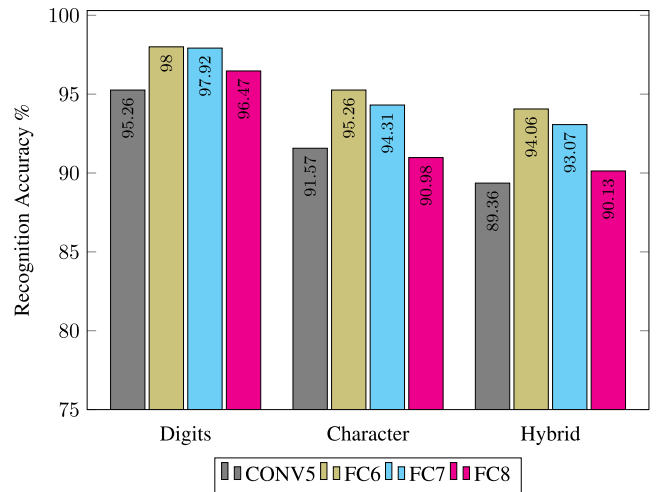


FIGURE 11. Recognition accuracy for AlexSVM on Digits, Character and Hybrid datasets with features from different layers.

where Tp, Fp, Tn and Fn are true positive, false positive, true negative and false negative respectively and their confusion matrix are shown in Table 3.

B. RESULTS FOR URDU DIGITS RECOGNITION

Experiments are conducted with different TL settings for AlexSVM on digits, character and hybrid datasets. It can be evidently seen from Figure 11, that the highest recognition accuracy is achieved by replacing FC6 among all settings. Table 4 shows the comparison of recognition accuracy among the proposed research using different settings and the state-of-the-art deep learning algorithms. The highest recognition accuracy is obtained for the AlexFT. Among the four different settings for AlexSVM, the features obtained at AlexSVM (FC6) achieve highest accuracy thereby outperforming AlexSVM (CONV5), AlexSVM (FC7) and AlexSVM (FC8) by 2.74%, 0.08% and 1.53% respectively.

Table 4 also demonstrates other important evaluation measures such as precision, recall, F-score and error rate for the proposed research. F-score is an important metric since in case of very low values of precision or recall, F-score helps to level the two measures. The higher values of the F-score are predictive of better results, with the worst possible 0 and 1 being the best. A good F-score is indicative

of a good precision and recall value. The average precision, recall, F-score and error rate for the digits dataset are obtained for AlexFT i.e. 97.08%, 97.04%, 97.01% and 2.92% respectively.

Table 5 shows the classification accuracy comparison of the proposed research with the traditional classification models as Logistic Regression, KNN classifier, Neural network and SVM for the digits category. The experimental results demonstrate that the proposed approach AlexFT outperforms the state-of-the-art methods by achieving 12.21%, 6.23%, 6.12% and 2.42% higher recognition accuracy as compared to Logistic regression, KNN classifier, Neural network and SVM respectively. The AlexFT achieves highest recognition accuracy as compared to the state-of-the-art algorithms CNN [19] and Autoencoder [19] by providing 1.51% and 0.91% higher classification performance. The average precision, recall, F-score and error rate for the digits dataset are obtained for AlexFT i.e. 98.23%, 98.21%, 98.22% and 1.79% respectively. The comparison of the proposed research approaches demonstrate that AlexFT achieves state-of-the-art performance as compared to the deep learning algorithms and the proposed AlexSVM.

The confusion matrices obtained using the four different settings with Alext-SVM, and AlexNet-FT with fine-tuned hyper-parameters and data augmentation for Urdu digits are shown in Figure 12 (a-e) respectively. The rows in the

TABLE 4. Comparison of the proposed methods for Urdu digits recognition.

Algorithms	Accuracy	Precision	Recall	F-Score	Error Rate
AlexSVM(CONV5)	95.26%	95.43%	95.34%	95.38	4.74%
AlexSVM(FC6)	98%	98.03%	97.99%	98%	2%
AlexSVM(FC7)	97.92%	97.93%	97.88%	97.9%	2.08%
AlexSVM(FC8)	96.47%	96.47%	96.47%	96.46%	3.53%
AlexFT	98.21 %	98.23%	98.21%	98.22	1.79

TABLE 5. Recognition accuracy comparison of the proposed research with traditional classification models for digits dataset.

Algorithms	Accuracy	Error Rate
Logistic regression [19]	86%	14%
KNN classifier [19]	92.09%	7.1%
Neural network [19]	91.98%	8.02%
SVM [19]	95.79%	4.21%
CNN [19]	96.7%	3.3%
Autoencoder [19]	97.3%	2.7%
Guassian NB [19]	69.3%	30.7%
Decision Tree [19]	82.00%	18%
Daubechies wavelet [59]	92.05%	7.95%
Fuzzy rule [60]	97.4%	2.6%
HMM (Hidden Markov Model) [60]	96.2%	3.8%
Hybrid approach (Fuzzy and HMM) [60]	97.8%	2.5%
Fuzzy rule base approach [61]	96.3%	3.7%
AlexSVM	98.00%	2%
AlexFT	98.21 %	1.79%

TABLE 6. Comparison of results for Urdu characters recognition with the state-of-the-art.

Algorithms	Accuracy	Precision	Recall	F-Score	Error Rate
Autoencoder [19]	81.2%	-	-	-	18.8%
CNN [19]	86.6%	-	-	-	13.4%
CNN [33]	80.7%	-	-	-	19.3%
SVM [33]	56%	-	-	-	44.0%
ANN [33]	78%	-	-	-	22.0%
LeNet [62]	90.34%	-	-	-	9.70%
SVM with polynomial kernel on 40x36 [63]	88.80 %	-	-	-	11.20%
SVM with Transfer Learning [64]	82.30%	-	-	-	17.7%
BLSM [65]	92-94%	-	-	-	8 - 6%
OCR-GoogleNet [64]	94.7%	-	-	-	5.3%
AlexSVM (CONV5)	91.57%	91.86%	91.47%	91.66%	8.43%
AlexSVM (FC6)	95.26%	95.29%	95.21%	95.23%	4.74%
AlexSVM (FC7)	94.31%	94.34%	94.27%	94.28%	5.69%
AlexSVM (FC8)	90.98%	90.98%	90.85%	90.85%	9.02%
AlexFT	97.08%	97.08%	97.04%	97.01%	2.92%

confusion matrix plot correspond to the predicted class (output class) and the columns represent the true class (target class). The values at the diagonal cells depict the correctly classified observations. The off-diagonal cells represent the observations that are incorrectly classified.

C. RESULTS FOR URDU CHARACTERS RECOGNITION

The experimental results for urdu character recognition are shown in Table 6. It can be seen that among different settings of AlexSVM i.e. AlexSVM (CONV5), AlexSVM (FC6), AlexSVM (FC7) and AlexSVM (FC8) the highest recognition accuracy is obtained by using features from AlexSVM (FC6). Of the two proposed research methods i.e. AlexSVM and AlexFT the later achieves the highest accuracy and outperforms the state-of-the-art-research. AlexFT obtained 15.88%, 10.48% and 1.82% higher accuracy as compared to Autoencoder, CNN and AlexSVM(FC6). The average precision, recall, F-score and error rate for the characters image

dataset are obtained for AlexFT i.e. 97.08%, 97.04%, 97.01% and 2.92% respectively.

D. RESULTS FOR HYBRID DATASET

The quantitative results for hybrid dataset are presented in Table 7. It can be evidently seen that AlexSVM (FC6) outperforms the AlexSVM (CL5), AlexSVM (FC7) and AlexSVM (FC8). However, the best classification performance is obtained by the proposed AlexFT. AlexFT outperforms the state-of-the-art methods i.e. Autoencoder and CNN by 12.92% and 12.12% respectively. It can be safely concluded that both the proposed methods i.e. AlexSVM (FC6) and AlexFT outperform the state-of-the-art results, with AlexFT being the best approach in terms of recognition accuracy achieved. The average precision, recall, F-score and error rate for the hybrid dataset are obtained for AlexFT i.e. 94.91%, 94.88%, 94.75% and 5.08% respectively.

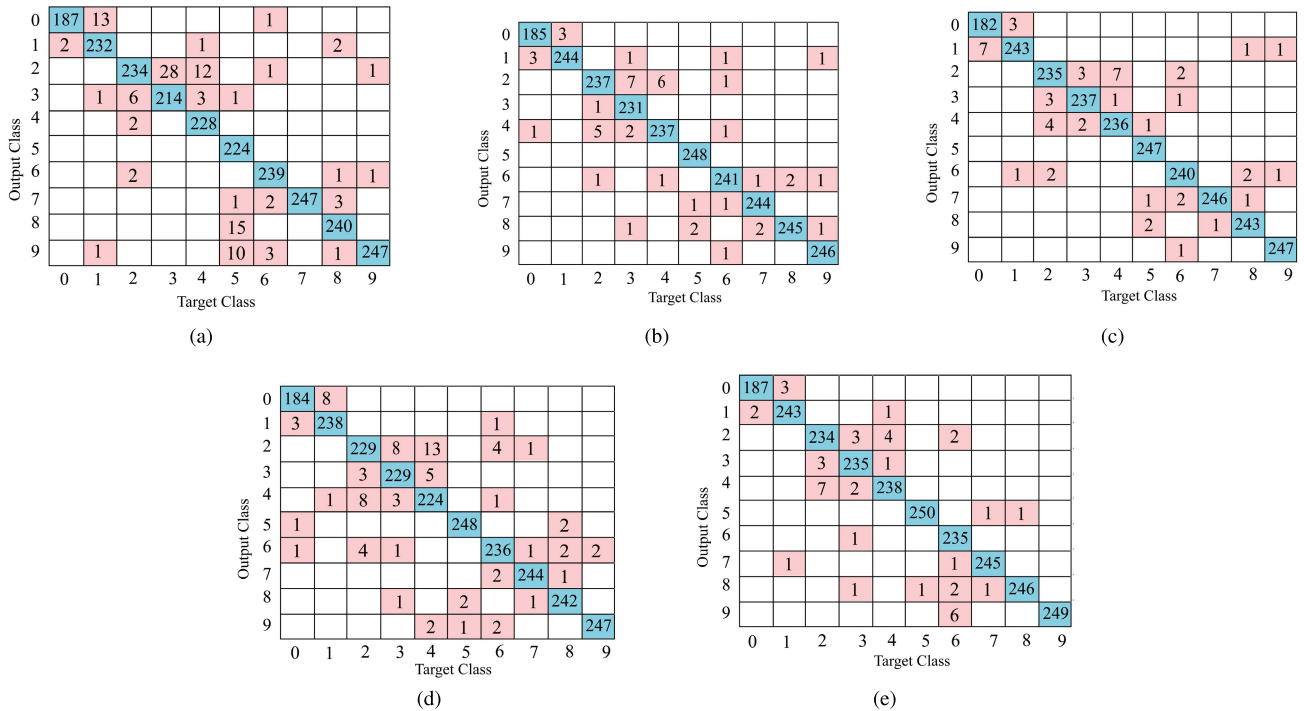


FIGURE 12. Confusion Matrix ((a) For AlexSVM with the setting CONV5 for Urdu digits (b) For AlexSVM with the setting FC6 for Urdu digits (c) For AlexSVM with the setting FC7 for Urdu digits (d) For AlexSVM with the setting FC8 for Urdu digits (e) For AlexFT with fine-tuned hyper-parameters for Urdu digits).

TABLE 7. Comparison of results for hybrid characters recognition with the state-of-the-art.

Algorithms	Accuracy	Precision	Recall	F-Score	Error Rate
Autoencoder [19]	82%	-	-	-	18%
CNN [19]	82.8%	-	-	-	17.2%
K-Nearest Neighbor (K-NN) [66]	38%	-	-	-	62%
Recurrent Neural Network (RNN) [66]	80%	-	-	-	20.0%
AlexSVM (CONV5)	89.36%	89.93%	89.20%	89.56%	10.64%
AlexSVM (FC6)	94.06%	94.06%	93.96%	93.99%	5.94%
AlexSVM (FC7)	93.07%	93.10%	93.06%	93.06%	6.93%
AlexSVM (FC8)	90.13%	90.16%	90.07%	90.09%	9.87%
AlexFT	94.92%	94.91%	94.88%	94.75%	5.08%

E. DEEP LEARNING APPROACHES VS TRADITIONAL APPROACHES

This section presents a theoretical comparison of deep learning methodology and the classical methodologies. The accuracy and performance of DL techniques appear to be high in most cases. DL has pushed the boundaries of image processing and method automation, producing extraordinary results. Because ML has a high sensitivity and specificity for recognition and detection, traditional techniques are thought to be less effective than the former. DL is mostly built on Artificial Neural Networks (ANNs), which are brain-like structures that function similarly to the brain. Because learned neural networks are employed rather than programmed, they provide more accuracy than traditional approaches. The goal of the first creation of neural networks was to replicate human brains. The evolution of shallow networks into deep architectures reduces resource requirements while maintaining

representation power. Deep learning methods have shown to be effective in a variety of domains of recognition tasks, computer vision, including textile image analysis, in recent years.

Deep neural networks are an innovative framework for analyzing vision and recognition. Deep nets' range of vision tasks is rapidly expanding, and they do represent a quantum leap forward in comparison to earlier computer vision systems. Deep networks can learn from examples to approximate functions and dynamics. Since the last decade of research, deep learning models such as convolutional neural networks, recurrent neural networks, and deep belief networks have dominated. These models have evolved in a variety of directions over the course of more than half a century. However, the limitations of deep learning models such as excessive training time, requirement of GPU for execution, parameter optimization and performance on small-scale datasets are still questionable.

VI. CONCLUSION

In this article deep neural network i.e. AlexNet is applied for automatic recognition of the hand-written characters and digits using different approaches. In the first approach, pre-trained AlexNet is applied for feature extraction and SVM is used for classification. In the second approach, transfer learning is used for extracting features from fine-tuned AlexNet model and Softmax layer is used for classification. Data augmentation is applied to avoid over fitting. The experiments are conducted using different hyper-parameter settings for the proposed approaches. Firstly, the classification performance for pre-trained CNN AlexNet is evaluated by extracting learned features and using a multi-class SVM classifier. The features are extracted by replacing different network layers to obtain the optimal performance. Secondly, the performance of transfer learning is evaluated from pre-trained CNN AlexNet by fine-tuning hyper-parameters and applying data augmentation for the classification task.

In the study, three different settings of data have been used for evaluation namely, digits only, characters only and hybrid; comprising of both digits and characters. The results for AlexNet with SVM confirmed that the features extracted from 'FC6' provided optimal performance. Experimental results have shown that the proposed approach using a fine-tuned AlexNet model with data augmentation resulted in better recognition performance and outperformed the state-of-the-art research. For digits and characters recognition, the highest accuracy obtained is up to 98.21% and 97.08%, respectively. For hybrid setting i.e. digits and characters, the accuracy is 94.92%. Experimental results demonstrate that the proposed fine-tuned AlexNet model with data augmentation and transfer learning outperforms the state-of-the-art classification methods. However, for deep learning model we have to find optimal values of hyper parameters against each dataset. To perform better than other strategies, deep learning models require a big volume of data. Because of the complicated data models training is exceedingly costly and requires the use of costly GPUs and hundreds of workstations. In future, we intend to further improve the recognition accuracy by applying some different feature extraction technique. Deep neural networks such as ResNet and GoogleNet; and other machine learning algorithms such as generative adversarial networks can be used in future research.

REFERENCES

- [1] Z. Chen, "Research on internet security situation awareness prediction technology based on improved RBF neural network algorithm," *J. Comput. Cogn. Eng.*, vol. 1, no. 3, pp. 103–108, Mar. 2022.
- [2] X. Zhang and G. Wang, "Stud pose detection based on photometric stereo and lightweight YOLOv4," *J. Artif. Intell. Technol.*, vol. 2, no. 1, pp. 32–37, 2022.
- [3] A. Shabbir, A. Rasheed, H. Shehraz, A. Saleem, B. Zafar, M. Sajid, N. Ali, S. H. Dar, and T. Shehryar, "Detection of glaucoma using retinal fundus images: A comprehensive review," *Math. Biosci. Eng.*, vol. 18, no. 3, pp. 2033–2076, 2021.
- [4] A. Shahzad, B. Zafar, N. Ali, U. Jamil, A. J. Alghadhban, M. Assam, N. A. Ghamry, and E. T. Eldin, "COVID-19 vaccines related user's response categorization using machine learning techniques," *Computation*, vol. 10, no. 8, p. 141, 2022. [Online]. Available: <https://www.mdpi.com/2079-3197/10/8/141>
- [5] S. Bai, S. Song, S. Liang, J. Wang, B. Li, and E. Neretin, "UAV maneuvering decision-making algorithm based on twin delayed deep deterministic policy gradient algorithm," *J. Artif. Intell. Technol.*, vol. 2, no. 1, pp. 16–22, 2022.
- [6] A. Rasheed, B. Zafar, A. Rasheed, N. Ali, M. Sajid, S. H. Dar, U. Habib, T. Shehryar, and M. T. Mahmood, "Fabric defect detection using computer vision techniques: A comprehensive review," *Math. Problems Eng.*, vol. 2020, pp. 1–24, Nov. 2020.
- [7] K. N. S. Nischal, G. N. Sai, C. Mathew, G. C. Gowda, and C. Bm, "A survey on recognition of handwritten zip codes in a postal sorting system," *Int. Res. J. Eng. Technol.*, vol. 7, Mar. 2020.
- [8] S. A. Husain, A. Sajjad, and F. Anwar, "Online Urdu character recognition system," in *Proc. MVA*, 2007, pp. 98–101.
- [9] K. Khan, R. Ullah, N. Ahmad Khan, and K. Naveed, "Urdu character recognition using principal component analysis," *Int. J. Comput. Appl.*, vol. 60, no. 11, pp. 1–4, Dec. 2012.
- [10] E. Ivarsson and R. M. Mueller, "Racing bib number recognition using deep learning," in *Proc. 25th Amer. Conf. Inf. Syst. (AMCIS)*, Cancun, Mexico, Aug. 2019.
- [11] Q. Luo, J. Wu, and M. Gombolay, "A generalized robotic handwriting learning system based on dynamic movement primitives (DMPs)," 2020, *arXiv:2012.03898*.
- [12] S. Fatima, N. A. Aslam, I. Tariq, and N. Ali, "Home security and automation based on Internet of Things: A comprehensive review," in *Proc. IOP Conf. Mater. Sci. Eng.*, 2020, vol. 899, no. 1, Art. no. 012011.
- [13] G. De Luca and Y. Chen, "Explainable artificial intelligence for workflow verification in visual IoT/robotics programming language environment," *J. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 21–27, 2021.
- [14] A. Shabbir, N. Ali, J. Ahmed, B. Zafar, A. Rasheed, M. Sajid, A. Ahmed, and S. H. Dar, "Satellite and scene image classification based on transfer learning and fine tuning of ResNet50," *Math. Problems Eng.*, vol. 2021, pp. 1–18, Jul. 2021.
- [15] M. Yang, "Research on vehicle automatic driving target perception technology based on improved msrpn algorithm," *J. Comput. Cognit. Eng.*, vol. 1, no. 3, pp. 147–151, 2022.
- [16] M. Mehmood, A. Shahzad, B. Zafar, A. Shabbir, and N. Ali, "Remote sensing image classification: A comprehensive review and applications," *Math. Problems Eng.*, vol. 2022, pp. 1–24, Aug. 2022.
- [17] F. Masood, H. Masood, H. Zahir, K. Driss, N. Mehmood, and H. Farooq, "Novel approach to evaluate classification algorithms and feature selection filter algorithms using medical data," *J. Comput. Cogn. Eng.*, May 2022.
- [18] K. M. Hosny, M. A. Kassem, and M. M. Fouad, "Classification of skin lesions using transfer learning and augmentation with alex-net," *PLoS ONE*, vol. 14, no. 5, May 2019, Art. no. e0217293.
- [19] H. Ali, A. Ullah, T. Iqbal, and S. Khattak, "Pioneer dataset and automatic recognition of Urdu handwritten characters using a deep autoencoder and convolutional neural network," *Social Netw. Appl. Sci.*, vol. 2, no. 2, p. 152, Feb. 2020.
- [20] M. A. Aslam, M. N. Salik, F. Chughtai, N. Ali, S. H. Dar, and T. Khalil, "Image classification based on mid-level feature fusion," in *Proc. 15th Int. Conf. Emerg. Technol. (ICET)*, Dec. 2019, pp. 1–6.
- [21] N. Ali, K. B. Bajwa, R. Sablatnig, S. A. Chatzichristofis, Z. Iqbal, M. Rashid, and H. A. Habib, "A novel image retrieval based on visual words integration of SIFT and SURF," *PLoS ONE*, vol. 11, no. 6, Jun. 2016, Art. no. e0157428.
- [22] U. Ravi Babu, A. Kumar Chintna, and Y. Venkateswarlu, "Handwritten digit recognition using structural, statistical features and K-nearest neighbor classifier," *Int. J. Inf. Eng. Electron. Bus.*, vol. 6, no. 1, pp. 62–68, Feb. 2014.
- [23] D. Gorgevik and D. Cakmakov, "Handwritten digit recognition by combining SVM classifiers," in *Proc. Int. Conf. 'Comput. Tool' (EUROCON)*, vol. 2, Nov. 2005, pp. 1393–1396.
- [24] R. Arnold and P. Miklos, "Character recognition using neural networks," in *Proc. 11th Int. Symp. Comput. Intell. Informat. (CINTI)*, Nov. 2010, pp. 311–314.
- [25] X. Xiao, L. Jin, Y. Yang, W. Yang, J. Sun, and T. Chang, "Building fast and compact convolutional neural networks for offline handwritten Chinese character recognition," *Pattern Recognit.*, vol. 72, pp. 72–81, Dec. 2017.
- [26] Z. Li, N. Teng, M. Jin, and H. Lu, "Building efficient CNN architecture for offline handwritten Chinese character recognition," *Int. J. Document Anal. Recognit.*, vol. 21, no. 4, pp. 233–240, 2018.
- [27] I. Ahmad, X. Wang, R. Li, and S. Rasheed, "Offline Urdu Nastaleeq optical character recognition based on stacked denoising autoencoder," *China Commun.*, vol. 14, no. 1, pp. 146–157, Jan. 2017.

- [28] R. S. Hussien, A. A. Elkhidir, and M. G. Elnourani, "Optical character recognition of Arabic handwritten characters using neural network," in *Proc. Int. Conf. Comput., Control, Netw., Electron. Embedded Syst. Eng. (ICCNTEE)*, Sep. 2015, pp. 456–461.
- [29] R. I. Elanwar, M. A. Rashwan, and S. A. Mashali, "Simultaneous segmentation and recognition of Arabic characters in an unconstrained on-line cursive handwritten document," *Int. J. Comput. Inf. Sci. Eng.*, vol. 4, pp. 203–206, Jan. 2007.
- [30] Q.-T.-A. Safdar and K. U. Khan, "Online Urdu handwritten character recognition: Initial half form single stroke characters," in *Proc. 12th Int. Conf. Frontiers Inf. Technol.*, Dec. 2014, pp. 292–297.
- [31] J. H. AlKhateeb, "A database for Arabic handwritten character recognition," *Proc. Comput. Sci.*, vol. 65, pp. 556–561, Jan. 2015.
- [32] M. Akhlaghi and V. Ghods, "Farsi handwritten phone number recognition using deep learning," *Social Netw. Appl. Sci.*, vol. 2, no. 3, pp. 1–10, Mar. 2020.
- [33] S. Khan, A. Hafeez, H. Ali, S. Nazir, and A. Hussain, "Pioneer dataset and recognition of handwritten pashoto characters using convolution neural networks," *Meas. Control*, vol. 53, nos. 9–10, pp. 2041–2054, Nov. 2020.
- [34] M. A. Azad, H. S. Singha, and M. M. H. Nahid, "Bangla handwritten character recognition using deep convolutional autoencoder neural network," in *Proc. 2nd Int. Conf. Adv. Inf. Commun. Technol. (ICAICT)*, Nov. 2020, pp. 295–300.
- [35] M. Husnain, M. M. Saad Missen, S. Mumtaz, M. Coustaty, M. Luqman, and J. Ogier, "Urdu handwritten text recognition: A survey," *IET Image Process.*, vol. 14, no. 11, pp. 2291–2300, Sep. 2020.
- [36] S. Naz, K. Hayat, M. I. Razzak, M. W. Anwar, S. A. Madani, and S. U. Khan, "The optical character recognition of Urdu-like cursive scripts," *Pattern Recognit.*, vol. 47, no. 3, pp. 1229–1248, Mar. 2014.
- [37] Z. Chen, F. Yin, X.-Y. Zhang, Q. Yang, and C.-L. Liu, "MuLTrE-Nets: Multilingual text recognition networks for simultaneous script identification and handwriting recognition," *Pattern Recognit.*, vol. 108, Dec. 2020, Art. no. 107555.
- [38] S. Naz, A. I. Umar, R. Ahmad, I. Siddiqi, S. B. Ahmed, M. I. Razzak, and F. Shafait, "Urdu Nastaliq recognition using convolutional–recursive deep learning," *Neurocomputing*, vol. 243, pp. 80–87, Jun. 2017.
- [39] F. Mushtaq, M. M. Misgar, M. Kumar, and S. S. Khurana, "UrduDeepNet: Offline handwritten Urdu character recognition using deep neural network," *Neural Comput. Appl.*, vol. 33, no. 22, pp. 15229–15252, Nov. 2021.
- [40] N. Sehr Zia, M. F. Naem, S. M. K. Raza, M. M. Khan, A. Ul-Hasan, and F. Shafait, "A convolutional recursive deep architecture for unconstrained Urdu handwriting recognition," *Neural Comput. Appl.*, vol. 34, no. 2, pp. 1635–1648, Jan. 2022.
- [41] S. S. R. Rizvi, M. A. Khan, S. Abbas, M. Asadullah, N. Anwer, and A. Fatima, "Deep extreme learning machine-based optical character recognition system for nastalique urdu-like script languages," *Comput. J.*, vol. 65, no. 2, pp. 331–344, Feb. 2022.
- [42] W. Cao, Z. Zhang, C. Liu, R. Li, Q. Jiao, Z. Yu, and H.-S. Wong, "Unsupervised discriminative feature learning via finding a clustering-friendly embedding space," *Pattern Recognit.*, vol. 129, Sep. 2022, Art. no. 108768.
- [43] A. K.-F. Lui, Y.-H. Chan, and M.-F. Leung, "Modelling of pedestrian movements near an amenity in walkways of public buildings," in *Proc. 8th Int. Conf. Control, Autom. Robot. (ICCAR)*, Apr. 2022, pp. 394–400.
- [44] Y. F. Tan, T. Connie, M. K. O. Goh, and A. B. J. Teoh, "A pipeline approach to context-aware handwritten text recognition," *Appl. Sci.*, vol. 12, no. 4, p. 1870, Feb. 2022.
- [45] M. Elkhayati and Y. Elkettani, "UnCNN: A new directed CNN model for isolated Arabic handwritten characters recognition," *Arabian J. Sci. Eng.*, vol. 47, pp. 10667–10688, Mar. 2022.
- [46] S. Y. Arafat, N. Ashraf, M. J. Iqbal, I. Ahmad, S. Khan, and J. J. P. C. Rodrigues, "Urdu signboard detection and recognition using deep learning," *Multimedia Tools Appl.*, vol. 81, no. 9, pp. 11965–11987, Apr. 2022.
- [47] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed, and M. Fayyaz, "Exploring deep learning approaches for Urdu text classification in product manufacturing," *Enterprise Inf. Syst.*, vol. 16, no. 2, pp. 223–248, Feb. 2022.
- [48] H. Zargar, R. Almahasneh, and L. T. Kóczy, "Automatic recognition of handwritten Urdu characters," in *Computational Intelligence and Mathematics for Tackling Complex Problems 3* (Studies in Computational Intelligence), vol. 959, I. A. Hartati, L. T. Koczy, J. Medina, and E. Ramirez-Poussa, Eds. Cham, Switzerland: Springer, 2022, doi: 10.1007/978-3-030-74970-5_19.
- [49] J. Ren, M. Green, and X. Huang, "From traditional to deep learning: Fault diagnosis for autonomous vehicles," in *Learning Control*. Amsterdam, The Netherlands: Elsevier, 2021, pp. 205–219.
- [50] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1097–1105.
- [51] M. B. Er and I. B. Aydilek, "Music emotion recognition by using chroma spectrogram and deep visual features," *Int. J. Comput. Intell. Syst.*, vol. 12, no. 2, pp. 1622–1634, 2019.
- [52] R. A. Minhas, A. Javed, A. Irtaza, M. T. Mahmood, and Y. B. Joo, "Shot classification of field sports videos using AlexNet convolutional neural network," *Appl. Sci.*, vol. 9, no. 3, p. 483, Jan. 2019.
- [53] N. Ali, K. B. Bajwa, R. Sablatneg, and Z. Mehmood, "Image retrieval by addition of spatial information based on histograms of triangular regions," *Comput. Electr. Eng.*, vol. 54, pp. 539–550, Aug. 2016.
- [54] N. Ali, B. Zafar, M. K. Iqbal, M. Sajid, M. Y. Younis, S. H. Dar, M. T. Mahmood, and I. H. Lee, "Modeling global geometric spatial information for rotation invariant classification of satellite images," *PLoS ONE*, vol. 14, no. 7, Jul. 2019, Art. no. e0219833.
- [55] S. Ruder. (2019). *Transfer Learning-Machine Learning's Next Frontier*. Accessed: Apr. 2017. [Online]. Available: <https://ruder.io/transfer-learning>
- [56] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2009.
- [57] P. Szymak, "Selection of training options for deep learning neural network using genetic algorithm," in *Proc. 24th Int. Conf. Methods Models Autom. Robot. (MMAR)*, Aug. 2019, pp. 24–29.
- [58] A. Latif, A. Rasheed, U. Sajid, J. Ahmed, N. Ali, N. I. Ratyal, B. Zafar, S. H. Dar, M. Sajid, and T. Khalil, "Content-based image retrieval and feature extraction: A comprehensive review," *Math. Problems Eng.*, vol. 2019, Aug. 2019, Art. no. 9658350.
- [59] R. Borse and I. Ansari, *Offline Handwritten and Printed Urdu Digits Recognition Using Daubechies Wavelet*. New Delhi, India: ER Publication, 2015.
- [60] M. I. Razzak, S. Hussain, A. Belaid, and M. Sher, "Multi-font numerals recognition for Urdu script based languages," *Int. J. Recent Trends Eng.*, Dec. 2009. [Online]. Available: <https://hal.inria.fr/inria-00437121>
- [61] M. I. Razzak, S. A. Hussain, and M. Sher, "Numeral recognition for Urdu script in unconstrained environment," in *Proc. Int. Conf. Emerg. Technol.*, Oct. 2009, pp. 44–47.
- [62] W. Jiang, "MNIST-MIX: A multi-language handwritten digit recognition dataset," *IOP SciNotes*, vol. 1, no. 2, Sep. 2020, Art. no. 025002.
- [63] H. Ali, K. Iqbal, G. Mujtaba, A. Fayyaz, M. F. Bulbul, F. W. Karam, and A. Zahir, "Urdu text in natural scene images: A new dataset and preliminary text detection," *PeerJ Comput. Sci.*, vol. 7, p. e717, Sep. 2021.
- [64] A. K. O. Mohammed and S. Poruran, "OCR-nets: Variants of pre-trained CNN for Urdu handwritten character recognition via transfer learning," *Proc. Comput. Sci.*, vol. 171, pp. 2294–2301, Jan. 2020.
- [65] S. B. Ahmed, S. Naz, S. Swati, and M. I. Razzak, "Handwritten Urdu character recognition using one-dimensional BLSTM classifier," *Neural Comput. Appl.*, vol. 31, no. 4, pp. 1143–1151, 2017.
- [66] M. A. Chhajro, "Handwritten Urdu character recognition via images using different machine learning and deep learning techniques," *Indian J. Sci. Technol.*, vol. 13, no. 17, pp. 1746–1754, May 2020.



AQSA RASHEED received the Bachelor of Engineering (B.E.) and M.S. degrees in software engineering from the Mirpur University of Science and Technology (MUST), in 2019 and 2021, respectively. She is currently serving as a Lecturer at the Department of Computer Science, University of South Asia, Raiwind, Lahore, Pakistan. She has published six articles in different SCIE-indexed journals. Her research interests include machine learning, deep learning, digital image processing, and computer vision.



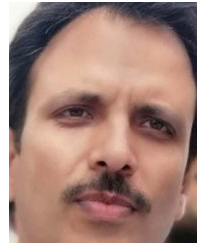
NOUMAN ALI received the Ph.D. degree in computer engineering from the University of Engineering and Technology (UET), Taxila, Pakistan. He is currently working as an Associate Professor and the Chairperson at the Department of Software Engineering, Mirpur University of Science and Technology (MUST), Mirpur, Azad-Kashmir, Pakistan. He has published more than 35 SCIE-indexed journal articles. His research interests include computer vision, digital image processing, and intelligent systems. He is serving as an Academic Editor for *Computational Intelligence and Neuroscience and Mathematical Problems in Engineering*. He is also serving as an Associate Editor for *International Journal of Computer Information Systems and Industrial Management Applications* and the Editorial Board Member of *Modern Intelligent Times*. He is also a reviewer of more than 40 SCIE-indexed journals.



BUSHRA ZAFAR received the Ph.D. degree in computer science from the National Textile University (NTU), Faisalabad, Pakistan. She is currently working as an Assistant Professor at the Department of Computer Science, Government College University Faisalabad (GCUF), Pakistan. Her research interests include computer vision, digital image processing, and artificial intelligence.



AMSA SHABBIR received the B.S. and M.S. degrees from the Department of Software Engineering, Mirpur University of Science and Technology (MUST), Mirpur, Pakistan, in 2018 and 2021, respectively. She has served as a Laboratory Engineer at the Department of Software Engineering, MUST, from March 2019 to September 2022. She has joined the Department of Software Engineering, Capital University of Science and Technology (CUST), Islamabad, Pakistan, as an Associate Lecturer. She has published four research articles in the field of computer vision and digital image processing. Her research interests include machine learning, artificial intelligence, computer vision, and digital image processing.



MUHAMMAD SAJID received the B.Sc., M.Sc., and Ph.D. degrees, all in electrical engineering, in 2002, 2008, and 2016, respectively. He is currently working as a Professor at the Department of Electrical Engineering, Mirpur University of Science and Technology. He has authored or coauthored 41 publications. His current research interests include artificial intelligence, machine learning, and digital image processing. He was awarded the Gold Medal Award by UAJK, in 2010, and the Pride of Pakistan Award, in 2022.



MUHAMMAD TARIQ MAHMOOD (Senior Member, IEEE) received the M.C.S. degree in computer science from the AJK University of Muzaffarabad, Pakistan, in 2004, the M.S. degree in intelligent software systems from the Blekinge Institute of Technology, Sweden, in 2006, and the Ph.D. degree in information and mechatronics from the Gwangju Institute of Science and Technology, South Korea, in 2011. He is currently working as an Associate Professor at the School of Computer Science and Engineering, Korea University of Technology and Education, South Korea. His research interests include image processing, 3D shape recovery, computer vision, and machine learning.

...