Open Access
Full Length Article

ORIGINAL ARTICLE

Urdu Handwritten Character Recognition using Deep Learning

Ali Jan[1], Junaid Baber[2], Anwar Ali Sanjrani[3], Maheen Bakhtyar[4]

[1,2,3,4] Department of Computer Science & IT, University of Balochistan, Quetta, Pakistan

ABSTRACT

Due to the emergence of ICT technologies, teaching and learning through the smart system is very trending. Kids start learning the basic alphabets of any language through the smart interaction of tablets and phones. However, very little research is reported on the Urdu language for such applications. Urdu is the national language of Pakistan and spoken across the subcontinent. In this paper, two famous deep learning-based models are evaluated for handwritten Urdu alphabets. Handwritten data is generated by the students and the teachers. Customized AlexNet is trained on these data, and accuracy is compared with the FERDNN model. The AlexNet is a famous model for various applications, and FERDNN is proposed for facial expressions. It has been observed during experiments that the AlexNet provides low accuracy using the default configurations in the first layer. Experiments show that AlexNet achieves competitive performance by changing the kernel size on the first layer, the customized AlexNet and FERDNN provides 0.95 and 0.83 accuracies on the training set, and 0.52 and 0.61 accuracies on the test set, respectively.

Keywords: Deep Learning, Urdu Language Processing, and Optical Character Recognition.

Author`s Contribution
[1,2,3,4]Manuscript writing, Planning of Research, Data Collection, Data analysis & interpretation, Conception, synthesis, planning of research & Discussion

Address of Correspondence
Junaid Baber
Email: junaidbaber@ieee.org

Article info.
Received: March 17, 2019
Accepted: Oct 29, 2019
Published: December 30,2019

Cite this article: *Jan A, Baber J, Sanjrani AA, Bakhtyar M. Urdu Handwritten Character Recognition Using Deep Learning. J. Inf. commun. technol. robot. appl.2019; 10(2):42-51*

Funding Source: *Nil*
Conflict of Interest: *Nil*

INTRODUCTION

The Urdu language is originated from Indo-European languages of the Indo Aryan family, and it is one of the famous languages of the subcontinent. Urdu is the national and official language of Pakistan which uses Persian and Arabic like scripts for writing. There are approximately 328.01 million native speakers in the subcontinent and 697.4 million across the world. The writing style of the Urdu language is from right to left. However, the numerals are written from left to right. The Urdu uses the Nastaliq style for writing scripts. The Nastaliq script has many challenges in developing and designing the OCR. The OCR is amongst the most successful and oldest machine learning and pattern recognition problems [1]. The peculiar challenges of Urdu Nastaliq include cursive writing style, context-sensitive style, diagonally, and multiple baselines. The Urdu language contains 39 characters, OCR development is complex because (i) it has a large number of character sets and (ii) similar ligature for different words. The cursive language, such as Urdu, Arabic, Sindhi, and Persian, have different shapes in ligature depending on the position of the character. The positions are initial,

middle, final, and standalone. Urdu has borrowed a large number of vocabularies from Persian and Arabic languages.

Much research has been reported on printed (a.k.a online) and handwritten (a.k.a offline), OCR on various famous languages. Many Asian languages have already matured in OCR research, such as Arabic, Persian, and Chinese over the recent decades. A large amount of the research work carried out for recognition of the isolated character with varying fonts, styles and sizes. So far there is no commercial character recognition system is available for Urdu scripts. As already mentioned, Urdu is a very old language and there exits abundant literature that should be digitized for larger public interest and knowledge sharing, and OCR plays a vital role in the automatic digitization of any literature. A typical OCR involves various steps such as preprocessing, feature extraction, and character classification [2]-[3]. The pre-processing is used to remove unnecessary pixel values and make the image more suitable for the segmentation and feature extraction. The widely used processes in preprocessing are binarization, thinning, thresholding, normalization, slant detection, edge detection, noise removal, smoothing and filtering, etc. Feature extraction refers to the characteristics of the input image represented in the numerical vector form. There are various categories of feature extraction and selection for achieving good performance on the OCR system [4].

Statistical and structural features are widely used in OCR. Statistical features include distance computation, the number of crossings, the number of pixel values, and moments [1]. The structural features involve horizontal and vertical lines, width and heights, curves, dots, and intersection points. Handwritten OCR is more challenging than printed OCR. Handwriting recognition is utilized in various applications such as personality identification, gender recognition, and writer identification [5]. The handwriting case becomes more challenging when the given input source has different sizes which are very common in handwriting.

Deep learning is among the family of neural networks that have influence in solving an enormous number of classification glitches like machine translation, speech recognition, character recognition, etc. Deep learning, so far, works as a black box where the input image is given and the output is obtained with very high accuracy. In deep learning, more than one layer is used for neural networks, and most of the recent models pipe convolutional layers before the neural networks. The convolutional layers are used to extract the low-level features from the given input. Among the many famous models, AlexNet is a widely used model for image classification

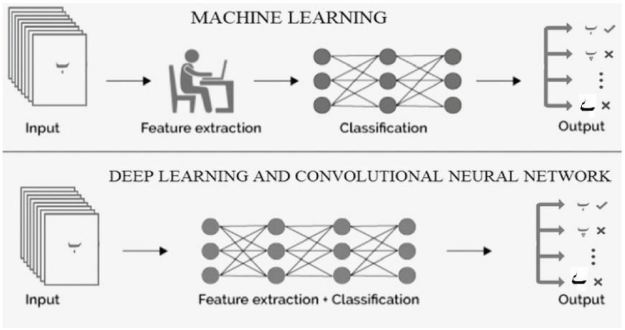


Figure 1. Architectural difference between classical machine learning and deep learning.

AlexNet is one of the famous models for image classification and recognition [6] -[7] in the family of deep learning models. Deep learning (DL) has been widely used for recognition in many applications such as image recognition, and object recognition. Recently, the idea of DL has been widely used for OCR as well. The architectural difference between classical machine learning and deep learning is shown in Figure 1. It can be seen that in classical machine learning, handcrafted features are used such as histograms, SIFT, and HoGs. Whereas, in deep learning, low-level features are learned by convolutional layers which are layer linked to fully connected layers.

In this research, two deep learning models, AlexNet [6] and FERDNN[8], are evaluated on Urdu handwritten characters. The AlexNet is a big and famous model and FERDNN is small and recently proposed for the facial expression recognition task. The AlexNet is modified for OCR applications as it gives a very poor performance on default configuration, as enlightened in the experimental section. The second contribution of the paper is an Urdu character dataset. Since there is no publicly available dataset for Urdu characters. Therefore, handwritten characters are obtained by different users of different ages. This dataset and the source code are kept online for everyone to use.

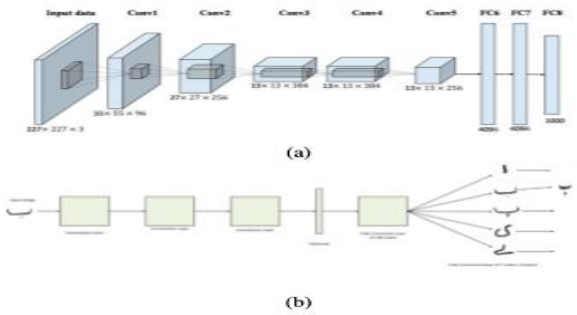


Figure 2. Methodologies for two DNN models used, (a) is of AlexNet, and (b) is of FERDNN. The configuration of FERDNN is explained in Section III-A1.

.

TABLE I: Urdu characters with their English pronunciation, Ci indicates the character index (ID), EP indicates English pronunciation, and UV indicates Urdu visual appearance.

Ci	EP	UV	Ci	EP	UV
1	Alef	ا	20	Soad	ص
2	Bey	ب	21	Zhoad	ض
3	Pay	پ	22	Tuain	ط
4	Tey	ت	23	Zuain	ظ
5	Thay	ٹ	24	Aeen	ع
6	Say	ث	25	Ghaeen	غ
7	Jeem	ج	26	Fay	ف
8	Chay	چ	27	Quaaf	ق
9	Hay	ح	28	Kaaf	ک
10	Khay	خ	29	Ghaf	گ
11	Daal	د	30	Laam	ل
12	Dahal	ڈ	31	Meen	م
13	Zaal	ذ	32	Noon	ن
14	Ray	ر	33	Wao	و
15	Rhay	ڑ	34	Hay	ہ
16	Zay	ز	35	Hay	ھ
17	Zhay	ژ	36	Hamza	ئ
18	Seen	س	37	Choti Yaee	ی
19	Sheen	ش	38	Barhi Yaee	ے

LITERATURE REVIEW

Many frameworks have been proposed for image classification and particularly for OCR. Most of the frameworks, either are statistical based recognition or it comes under the category of structural based recognition. In statistical pattern recognition, the feature vector is computed which is comprised of numerical values. The numerical values are either learned by some distances or probabilistic distribution of the given signal/image. Whereas, in structural based recognition, syntactic patterns are used as features to represent the objects. For example, in ECG recognition or facial expression recognition, the structure of the ECG or the face matters. Once, the features are learned, then these features are passed to some classifiers. Pattern recognition and classification algorithms are support vector machine (SVM), K-nearest neighbor (KNN), artificial neural networks (ANN), template matching algorithms, etc. The extensive study and development of handwritten characters are reported in the field of pattern recognition, computer vision, and image processing [9]-[10]. The pixel-level primitive features for the handwritten characters are color, projection histograms (horizontal and vertical histograms), moments, and texture. These are used in many applications of computer vision and image processing [11]-[12]. In deep learning study, the various deep neural networks have been applied for handwriting recognition. Moreover, there are some issues with deep learning networks in the fields of industries and academia [13]-[14]. The data gathering is the primary step for OCR applications. The limited data set is available for Urdu, Arabic, Pahto and Sindhi languages which are the famous languages of Asia. There is still room to improve the OCR system for these languages due to complex ligature and cursive nature. The most commonly and widely datasets as benchmarks for Latin character recognition are MNIST[15], IAMoffline [16] and UNLV-ISR [17]. The limited research has been carried out in Urdu isolated characters and ligatures. However, there is no benchmark dataset for Urdu characters and ligatures for carrying more research and study for developing improved and commercial level OCR systems. Later in this section, few famous frameworks are discussed on Urdu and related languages.

In [18], Urdu-Nastaliq handwritten dataset (UNHD) is trained with error rate up-to 6.04% to 7.93%. However, the accuracy reported is 89%. Sanjrani et al., proposed Sindhi handwritten numerals OCR using template matching and correlation coefficient techniques with an accuracy of 65% [19]. Sindhi numeral shapes are familiar with Urdu numerals. In Sindhi OCR, the simple correlation-based strategy is used to identify the false positives for each numeral class. It is claimed that finding the FP will help to tune the machine learning model more

accurately and a hierarchical classifier may be more effective in such cases. Hyat et al. [20], used ligatures for their cognition of units of text recognition in video frames and pre-trained deep convolutional networks for feature extraction. Whereas others evaluated the Back-Propagation (BP), neural network (NN) and (K-NN) classifiers [21]. They have achieved 93.61% accuracy on the Arabic language using a variation of NN. The variation of NN called Error Back-Propagation ANN (EBPANN). The EBPANN has received high accuracy due to high distinctiveness. Whereas, hybrid approaches, SVM along with NN are also used in the literature [22]. The reported accuracy of hybrid classifiers is significantly high compared to NN or single-stage classifiers. The combination of linear and nonlinear classifiers is very effective. Sagheer et al. [23] used a holistic approach and designed a handwritten-Urdu individual word recognition technique. For training and testing purposes, they used the CENPARMI Urdu word dataset that consists of 57 finance-related words. They have combined the validation with the training set with a new size of 14,407 samples. The testing set that contains 3,770-word images. In the experiments, the recognition results are compared with different sizes of the images for normalization, different operators for gradient feature extraction, the different features, and the different number of training samples. The data were normalized with 128 × 128 size and Sobel and Roberts filters are applied to extract gradient features. SVM is used for the classification purpose. In their method, the recognition rate has improved from 96.63% to 97.00% after increasing the size of the training set by approximately 3800 samples.

Data analyzing by scanning the documents is a formal practice in industrial applications. Though, a delay which is created by the error and this error produced by the OCR software that can increase the complexity of the task at hand [24]-[25]-[26]-[27]. OCR is a well-studied problem in the literature [28] with still open research challenges such as document layout analysis [29], word spotting [30], binarization [31], and digit recognition [32]. The OCR is often restricted to recognition and detection [33]-[34]. The handwritten OCR still has wide applications in industry. Many solutions have been proposed for handwritten OCR such as the chain-code quantization method [34]-[36] and ligature pattern matching method [37]. The ligature template matching method used NN with Nastalique font training. Tariq and Naru have proposed an Urdu OCR system in which research work is limited to isolated Urdu language characters [39], they claimed to achieve 97.43% accuracy.

All the above method uses handcrafted features which are then passed to the classifiers such as SVM and NN [40]. In the field of computer vision, low-level features are trained and passed to several layers of NN, which is the deep convolutional neural networks. The difference between deep learning and classical machine learning can be seen in Figure 1. Recently, all complex problems of NLP, signal processing, image processing, and computer vision are resolved with the help of deep learning. The main difference between these two paradigms is that classical machine learning work on two stages; one is feature extraction and the second is classification. Whereas, deep learning uses low-level features that are learned in the same network [41]. Deep learning technology is utilized in different objects recognition researches such as face recognition [42], herb leaf recognition [43], and character recognition [44]. Bag of features which is another famous technique in machine learning that is used in various applications of computer vision such as vehicle recognition [45], food recognition [46], and scene character recognition [41]-[47]

RESEARCH METHODOLOGY

As stated in the previous section, two deep learning models are evaluated on handwritten Urdu characters. The models include AlexNet [6] and FERDNN [8]. DL is the extension of neural networks. Instead of using a single layer and then the output layer, a large number of layers are added before the output layer, which makes the neural network deeper. DL has been state-of-the art for recognition. Both the above-mentioned models are discussed below.

3.1 The AlexNet Architecture

The AlexNet architecture which is a deep convolutional neural network (CNN) architecture, was first introduced by [6] in 2012 for the ImageNet. The ImageNet was a large-scale visual recognition challenge (ILSVRC2012). It is a simple yet effective architecture and is composed of many layers including convolution layers, pooling layers, rectified linear unit (ReLU) layers, and fully

connected layers. More specifically, the AlexNet layers are composed of five convolution layers followed by three fully connected layers before the final layer.

In this architecture, the extraction of convolution kernels is done using the back-propagation optimization using the cost stochastic gradient descent algorithm. The convolution layer is responsible for the generation of convolved feature maps using the input feature maps with sliding constitutional kernels, and the pooling layer is responsible to operate on the convolved feature maps. The pooling layer reduces the feature map by taking the average or max values of fix sliding kernel. The main reason for the success of the AlexNet is the ReLU nonlinearity layer and the dropout regularization technique. The ReLU is further explained in the next sub-section. The rectifier function is responsible to accelerate the training phase and prevent over-fitting. The dropout regularization technique regularize the process by stochastically setting several neurons. The network architecture of pre-trained AlexNet is shown in Figure 2 (a). The AlexNet is famous for its large kernel size on the first layer. Usually, smaller values are used as kernel size. However, larger values, 11 x 11 in the case of AlexNet surprisingly give good accuracy for several applications. In most of the image classification applications, the given image is mostly the clutter image that means there is a lot of edges and pixels. However, in the case of ORC, the input images are not clutter—a small set of pixels is used to describe the input shape. Mostly, the image has two colors only, one is used for the background and the second is for the foreground. For the foreground, black color is mostly used with white background. Sometimes in image processing, the color scheme is inverted: black for the background and white for the foreground. The schema is inverted, mostly when an image must be used as a binary image. It is easy to handle the pixel of the input shape in binary when the value is 1, and vice-versa to handle the background. Experiments show that AlexNet is not very effective for OCR applications on default configuration, particularly, the large kernel size on the first layer. Though, overall kernel size plays a vital role in other applications. In this paper, different kernel size options are evaluated and reported in the next section, as illustrated in Table-III.

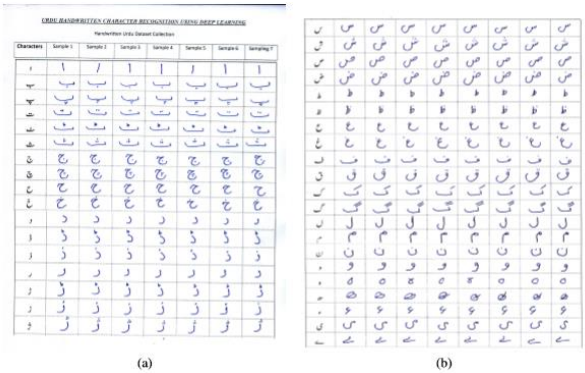


Figure 3. Handwritten Urdu characters collection from different users. A sample of the few classes from one user is shown.

3.2 The FERDNN Architecture:

The second architecture used in this paper is FERDNN that is shown in Figure 2 (b). This network consists of three convolution layers and two fully connected layers. The network architecture is initially proposed for facial expression recognition [8]. In CNN, the output of one layer is passed to the next layers. The output of one layer on CNN is obtained by traversing the sliding window, horizontally and vertically, of size n1 × n2, where n1 = n2. The stride, which is a unit of the sliding window is widely kept either one or two pixels. In the first three convolutional layers, 32, 64, and 32 filters are used in first, second, and third layers, respectively. Every convolutional layer is followed by a pooling function. The max-pooling function along with a kernel size of 3 × 3 is used in all layers. The filter kernel size is also kept 3 × 3 on each layer. Before the pooling operation, the features are passed to the activation function. The leaky Rectified Linear Unit (Leaky ReLU) which is a variation of simple ReLU, is used in FERDNN. It has been shown that Leaky ReLU gives better accuracy compared to simple ReLU. The equation of ReLU is given below which gives zero on negative values.

$$ReLU(x) = \begin{cases} x & if x \geq 0 \\ 0 & if x < 0 \end{cases}$$

The response against negative values is treated as zero. The zero response on negative values leads to the problems sometimes, as negative values received by the convolution layers have high importance.

TABLE II: AlexNet comparison with FERDNN

DNN	Train set	Test set
Default AlexNet [7]	0.39	-
Customized AlexNet	0.95	0.52

FERDNN [9]	0.83	0.61

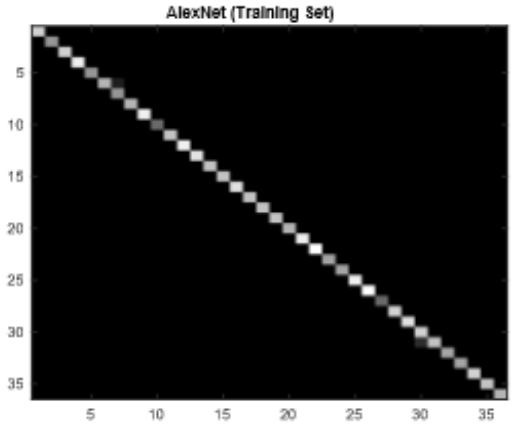
To preserve the importance of negative values in ReLU, following extension is used, which is known as Leaky ReLU

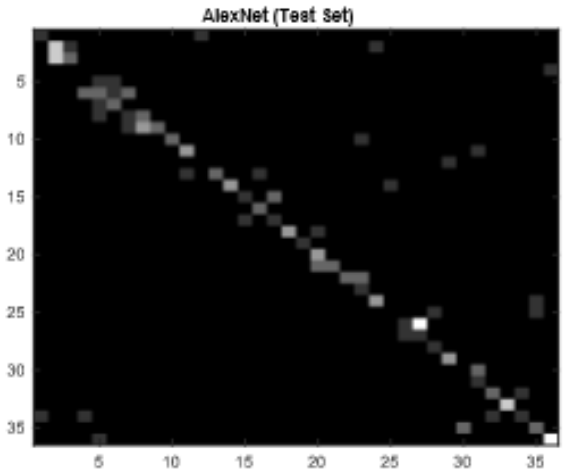
$$Leaky\ ReLU(x) = \begin{cases} x & if x \geq 0 \\ \frac{x}{\alpha} & if x < 0 \end{cases}$$

Where α is kept at 0.1 in proposed FERDNN, as suggested in the original paper. This model is small and simple. It was initially proposed to identify the FP of facial expressions. The FERDNN gives a high performance on the training set and low performance on the test set. It is a common problem of overfitting which is found in many deep learning models.

RESULTS AND DISCUSSION

Two A4 size pages of data entry formats have been designed for the collection of handwritten Urdu characters. The shape of every character is written on the left side of the page and writers are requested to write the same character at least 8 times on the same row. It enables us to capture the variation of a writer's style. Urdu characters are shown in Table I. It can be seen that many different characters have similar shapes, the only difference is the number and places of the dots. The handwritten data is scanned at 300 DPI. One of the scanned handwritten samples is shown in Figure 3. Later, all the characters are represented as an independent image of size 80×100 for training and testing.





(b)

Figure 4. Methodologies for two DNN models used, (a) is of AlexNet, and (b) is of FERDNN. The configuration of FERDNN is explained in Section III-A1.

The performance of both models are reported in Table II and Figures 4-5. AlexNet has achieved competitive performance on several image recognition and classification problems with default settings. The default AlexNet has 92 convolution filters with kernel size 11×11. However, in the case of OCR, it gives very low performance even on the training set. Most of the DNN models over-fit during the training, the AlexNet on default settings gives only 0.39 accuracy even several iterations and number of epochs. In customized AlexNet, in which, different kernel sizes of convolution filters starting from 3 × 3 to 13 × 13, are evaluated. The reported accuracy of customized AlexNet is on 5 × 5 kernel size on layer 1. The result of different configuration is shown in Figure 6. It can be seen that the performance starts deteriorating w.r.t the kernel size—only 0.39 accuracy of AlexNet default value. It is also recommended that the kernel size should be an odd number so that the pixel can be easily centered around the kernel.

The customized AlexNet over-fit the training set and gives a low performance on the test set, as shown in Figure 4. It can be seen that in a test set, AlexNet gives low accuracy which indicates that AlexNet over fits in training. In the case of FERDNN also marginally over-fit on the training set and gives a comparatively low performance on the test set. However, FERDNN has a better performance compared to AlexNet on the test set despite having very limited layers and filters, as shown in

Figure 5. To combat the overfitting in deep learning, the dropout layer is mostly used, whereas, dropout layer is not used in FERDNN.

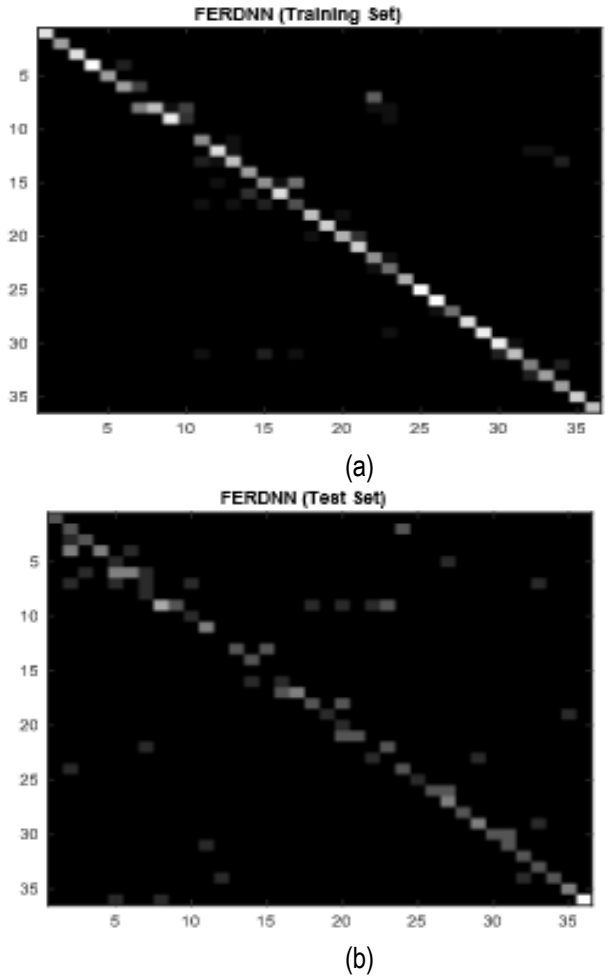


Figure 5. Methodologies for two DNN models used, (a) is of AlexNet, and (b) is of FERDNN. The configuration of FERDNN is explained in Section III-A1.

In FERDNN base paper, it gave a low performance on the test set for facial expression recognition. Whereas, it has given better performance on a test set as well.

CONCLUSION

Two DNN models, AlexNet and FERDNN, are evaluated on Urdu OCR. AlexNet is a widely used model for image classification and recognition, it has five convolution layers followed by three fully connected layers before the final layer. It is famous for her first layer in which a large kernel size is used, which is 11 x 11. However, AlexNet accuracy is very low on default settings on OCR applications due to large kernel size on first convolution layers, as illustrated in Table II. Large kernel size achieves better performance if they give an image/signal that is clutter. In many applications where AlexNet has achieved good performance, the image has a lot of information/content.

Whereas, in the case of OCR, the contents in images are very sparse: very few pixels are used to represent the characters with no background texture. Mostly, the character which has to be recognized is located at the center with a blank background, preferably white background. Therefore, the lower values of the kernel are used to train the model, which also gives better performance, as shown in Figure 6. In the case of FERDNN which is proposed for facial expression recognition, achieves better accuracy on the test set. However, AlexNet outperforms FERDNN on the training set, which is due to the over-fitting. Customized AlexNet and FERDNN achieve 0.95 and 0.83 accuracy on the training set, and 0.52 and 0.61 accuracies of the test set, respectively. All the experiments are done on handwritten characters of the Urdu language. Urdu characters are more challenging compared to English language characters. The visual appearance of Urdu characters can be seen in Table I and Figure 3. Table I shows the computer-generated fonts, and Figure 3 shows the handwritten acquisition of the Urdu characters. It is obvious that many characters have similar visual appearance with very minor changes. Results show that the famous deep learning models such as AlexNet cannot be generalized for all Urdu characters, as it has only 0.52 accuracy on unseen data (test set). Future works include extending the experiments to ligatures from the characters.

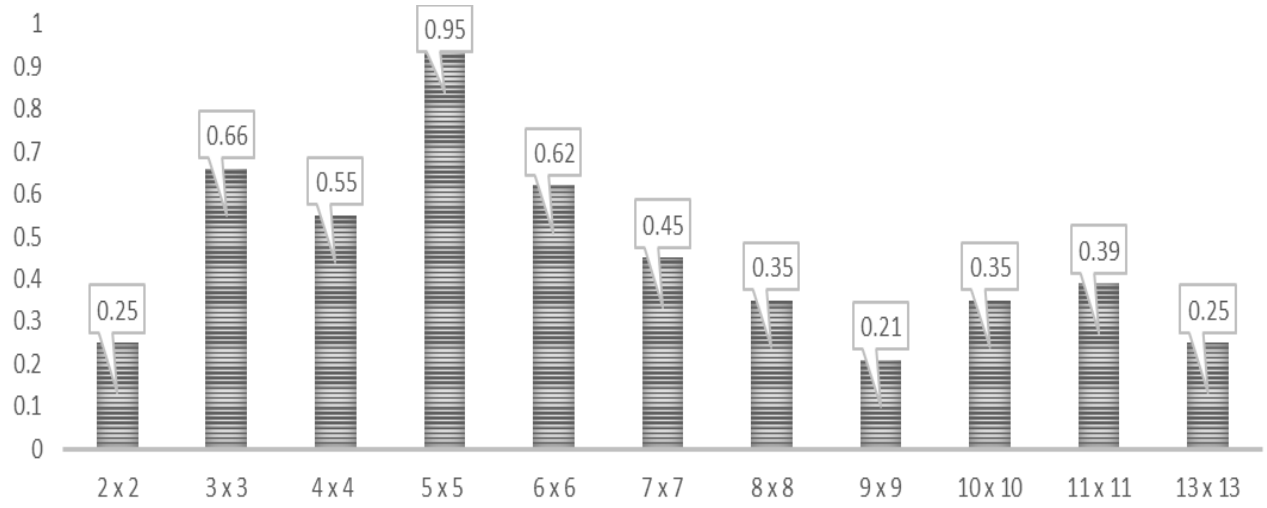


Figure. 6. Accuracy of AlexNet on different configuration of filter size

REFERENCES

1. D. Paul and B. B. Chaudhuri, "A BLSTM Network for Printed Bengali OCR System with High Accuracy," arXiv preprint arXiv:1908.08674, 2019.
2. X.-C. Yin, X. Yin, K. Huang, and H.-W. Hao, "Robust text detection in natural scene images," IEEE transactions on pattern analysis and machine intelligence, vol. 36, no. 5, pp. 970-983, 2013.
3. X.-C. Yin, X. Yin, K. Huang, and H.-W. Hao, "Accurate and robust text detection: A step-in for text retrieval in natural scene images," in Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, 2013: ACM, pp. 1091-1092.
4. Ø. D. Trier, A. K. Jain, and T. Taxt, "Feature extraction methods for character recognition-a survey," Pattern recognition, vol. 29, no. 4, pp. 641-662, 1996.
5. S. F. Rashid, M.-P. Schambach, J. Rottland, and S. von der Nüll, "Low resolution arabic recognition with multidimensional recurrent neural networks," in Proceedings of the 4th International Workshop on Multilingual OCR, 2013: ACM, p. 6.
6. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.
7. R. Ahmad, S. Naz, M. Afzal, S. Rashid, M. Liwicki, and A. Dengel, "A Deep Learning based Arabic Script Recognition System: Benchmark on KHAT."
8. J. Baber, M. Bakhtyar, K. U. Ahmed, W. Noor, V. Devi, and A. Sammad, "Facial Expression Recognition and Analysis of Interclass False Positives Using CNN," in Future of Information and Communication Conference, 2019: Springer, pp. 46-54.
9. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning. nature 521," 2015.
10. X.-Y. Zhang, Y. Bengio, and C.-L. Liu, "Online and offline handwritten chinese character recognition: A comprehensive study and new benchmark," Pattern Recognition, vol. 61, pp. 348-360, 2017.
11. Y. Mussarat, S. Muhamma, M. Sajjad, and I. Isma, "Content Based Image Retrieval Using Combined Features of Shape, Color and Relevance Feedback," KSII Transactions on Internet & Information Systems, vol. 7, no. 12, 2013.
12. A. Delaye and C.-L. Liu, "Contextual text/non-text stroke classification in online handwritten notes with conditional random fields," Pattern Recognition, vol. 47, no. 3, pp. 959-968, 2014.
13. Z. Rao et al., "Research on a handwritten character recognition algorithm based on an extended nonlinear kernel residual network," KSII Transactions on Internet & Information Systems, vol. 12, no. 1, 2018.
14. K. Khan, M. Siddique, M. Aamir, and R. Khan, "An efficient method for Urdu language text search in image based Urdu text," International Journal of Computer Science Issues (IJCSI), vol. 9, no. 2, p. 523, 2012.
15. L. Deng, "The MNIST database of handwritten digit images for machine learning research [best of the web]," IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 141-142, 2012.
16. U.-V. Marti and H. Bunke, "The IAM-database: an English sentence database for offline handwriting recognition," International Journal on Document Analysis and Recognition, vol. 5, no. 1, pp. 39-46, 2002.
17. K. Taghva, T. A. Nartker, J. Borsack, and A. Condit, "UNLV-ISRI document collection for research in OCR and information retrieval," in Document recognition and retrieval VII, 1999, vol. 3967: International Society for Optics and Photonics, pp. 157-164.

18. S. B. Ahmed, S. Naz, S. Swati, and M. I. Razzak, "Handwritten Urdu character recognition using one-dimensional BLSTM classifier," Neural Computing and Applications, vol. 31, no. 4, pp. 1143-1151, 2019.
19. A. A. Sanjrani, J. Baber, M. Bakhtyar, W. Noor, and M. Khalid, "Handwritten optical character recognition system for Sindhi numerals," in 2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), 2016: IEEE, pp. 262-267.
20. U. Hayat, M. Aatif, O. Zeeshan, and I. Siddiqi, "Ligature Recognition in Urdu Caption Text using Deep Convolutional Neural Networks," in 2018 14th International Conference on Emerging Technologies (ICET), 2018: IEEE, pp. 1-6.
21. M. Ali and H. Alasadi, "High Accuracy Arabic Handwritten Characters Recognition Using Error Back Propagation Artificial Neural Networks," IJACSA) International Journal of Advanced Computer Science and Applications, vol. 6, no. 2, 2015.
22. R. Kaur and G. S. Aujla, "Review on: Enhanced offline signature recognition using neural network and SVM," International Journal of Computer Science and Information Technologies", Volume5 (3), 2014.
23. M. W. Sagheer, C. L. He, N. Nobile, and C. Y. Suen, "Holistic urdu handwritten word recognition using support vector machine," in 2010 20th International Conference on Pattern Recognition, 2010: IEEE, pp. 1900-1903.
24. W. Bieniecki, S. Grabowski, and W. Rozenberg, "Image preprocessing for improving ocr accuracy," in 2007 International Conference on Perspective Technologies and Methods in MEMS Design, 2007: IEEE, pp. 75-80.
25. R. Smith, "An overview of the Tesseract OCR engine," in Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), 2007, vol. 2: IEEE, pp. 629-633.
26. A. Schmaltz, Y. Kim, A. M. Rush, and S. M. Shieber, "Adapting sequence models for sentence correction," arXiv preprint arXiv: 1707.09067, 2017.
27. K. Hakala, A. Vesanto, N. Miekka, T. Salakoski, and F. Ginter, "Leveraging Text Repetitions and Denoising Autoencoders in OCR Post-correction," arXiv preprint arXiv: 1906.10907, 2019.
28. G. Nagy, "Twenty years of document image analysis in PAMI," IEEE Transactions on Pattern Analysis & Machine Intelligence, no. 1, pp. 38-62, 2000.
29. F. Shafait, D. Keysers, and T. Breuel, "Performance evaluation and benchmarking of six-page segmentation algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 6, pp. 941-954, 2008.
30. V. Frinken, A. Fischer, R. Manmatha, and H. Bunke, "A novel word spotting method based on recurrent neural networks," IEEE transactions on pattern analysis and machine intelligence, vol. 34, no. 2, pp. 211-224, 2011.
31. N. R. Howe, "Document binarization with automatic parameter tuning," International Journal on Document Analysis and Recognition (IJDAR), vol. 16, no. 3, pp. 247-258, 2013.
32. R. Smith, "A simple and efficient skew detection algorithm via text row accumulation," in Proceedings of 3rd International Conference on Document Analysis and Recognition, 1995, vol. 2: IEEE, pp. 1145-1148.
33. A. Mishra, S. Shekhar, A. K. Singh, and A. Chakraborty, "OCR-VQA: Visual question answering by reading text in images," in Proceedings of the International Conference on Document Analysis and Recognition, 2019, vol. 1, no. 2, p. 5.
34. S. Hoque, K. Sirlantzis, and M. C. Fairhurst, "A New Chain-code Quantization Approach Enabling High Performance Handwriting Recognition based on Multi-Classifier Schemes," in ICDAR, 2003, pp. 834-838.
35. M. A. U. Rehman, "A new scale invariant optimized chain code for nastaliq character representation," in 2010 Second International Conference on Computer Modeling and Simulation, 2010, vol. 4: IEEE, pp. 400-403.
36. Z. A. Shah, "Ligature based optical character recognition of Urdu-Nastaleeq font," in International Multi Topic Conference, 2002. Abstracts. INMIC 2002., 2002: IEEE, pp. 25-25.
37. Z. Ahmad, J. K. Orakzai, I. Shamsher, and A. Adnan, "Urdu Nastaleeq optical character recognition," in Proceedings of world academy of science, engineering and technology, 2007, vol. 26: Citeseer, pp. 249-252.
38. H. Malik and M. A. Fahiem, "Segmentation of printed urdu scripts using structural features," in 2009 Second International Conference in Visualisation, 2009: IEEE, pp. 191-195.
39. J. Tariq, U. Nauman, and M. U. Naru, "Softconverter: A novel approach to construct OCR for printed Urdu isolated characters," in 2010 2nd International Conference on Computer Engineering and Technology, 2010, vol. 3: IEEE, pp. V3-495-V3-498.
40. S. Shalev-Shwartz and S. Ben-David, Understanding machine learning: From theory to algorithms. Cambridge university press, 2014.
41. S. A. Zulkeflie, F. A. Fammy, Z. Ibrahim, and N. Sabri, "Evaluation of Basic Convolutional Neural Network, AlexNet and Bag of Features for Indoor Object Recognition," International Journal of Machine Learning and Computing, vol. 9, no. 6, 2019.
42. N. A. B. M. Kasim, N. H. B. A. Rahman, Z. Ibrahim, and N. N. A. Mangshor, "Celebrity Face Recognition using Deep Learning," Indonesian Journal of Electrical Engineering and Computer Science, vol. 12, no. 2, pp. 476-481, 2018.
43. Z. Ibrahim, N. Sabri, and D. Isa, "Multi-maxpooling Convolutional Neural Network for Medicinal Herb Leaf Recognition," in Proceedings of the 6th IIAE International Conference on Intelligent Systems and Image Processing, Shimane, Japan, 2018, pp. 327-331.
44. M. M. Saufi, M. A. Zamanhuri, N. Mohammad, and Z. Ibrahim, "Deep Learning for Roman Handwritten Character Recognition," International Journal of Electrical Engineering and Computer Science, vol. 12, no. 2, pp. 455-460, 2018.

45. A. J. Siddiqui, A. Mammeri, and A. Boukerche, "Real-time vehicle make and model recognition based on a bag of SURF features," IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 11, pp. 3205-3219, 2016.
46. M. M. Anthimopoulos, L. Gianola, L. Scarnato, P. Diem, and S. G. Mougiakakou, "A food recognition system for diabetic patients based on an optimized bag-of-features model," IEEE journal of biomedical and health informatics, vol. 18, no. 4, pp. 1261-1271, 2014.
47. M. Tounsi, I. Moalla, and A. M. Alimi, "Supervised dictionary learning in BoF framework for Scene Character recognition," in 2016 23rd International Conference on Pattern Recognition (ICPR), 2016: IEEE, pp. 3987-3992.