A Deep Learning based Arabic Script Recognition System: Benchmark on KHAT

Riaz Ahmad[1], Saeeda Naz[2], Muhammad Afzal[3], Sheikh Rashid[4], Marcus Liwicki[5], and Andreas Dengel[6]
[1]Shaheed Banazir Bhutto University, Sheringal, Pakistan
[2]Computer Science Department, GGPGC No.1 Abbottabad, Pakistan
[3]Mindgarage, University of Kaiserslautern, Germany
[4]Al Khwarizmi Institute of Computer Science, UET Lahore, Pakistan
[5]Department of Computer Science, Luleå University of Technology, Luleå
[6]German Research Center for Artificial Intelligence (DFKI) in Kaiserslautern, Germany

Abstract: *This paper presents a deep learning benchmark on a complex dataset known as KFUPM Handwritten Arabic TexT (KHATT). The KHATT data-set consists of complex patterns of handwritten Arabic text-lines. This paper contributes mainly in three aspects i.e., (1) pre-processing, (2) deep learning based approach, and (3) data-augmentation. The pre-processing step includes pruning of white extra spaces plus de-skewing the skewed text-lines. We deploy a deep learning approach based on Multi-Dimensional Long Short-Term Memory (MDLSTM) networks and Connectionist Temporal Classification (CTC). The MDLSTM has the advantage of scanning the Arabic text-lines in all directions (horizontal and vertical) to cover dots, diacritics, strokes and fine inflammation. The data-augmentation with a deep learning approach proves to achieve better and promising improvement in results by gaining 80.02% Character Recognition (CR) over 75.08% as baseline.*

Keywords: *Handwritten Arabic text recognition, deep learning, data augmentation.*

Received July 11, 2017; accepted April 25, 2018
https://doi.org/10.34028/iajit/17/3/3

1. Introduction

Handwriting recognition is one of the active and hot research problems in the field of Optical Character Recognition (OCR). The handwriting recognition has numerous applications in many fields; like automatic cheque processing in banks, signature verification, personality identification, postal/zip code recognition, form processing, writer identification, gender recognition etc., [25]. OCR research for the Latin script has reached to maturity level by obtaining 99% accuracy for printed text, however, in the case of handwriting recognition it still lag far behind. The case becomes more challenging while considering the handwritten Arabic script as a test case.

The Arabic script is a cursive script and is known for its complex and challenging behaviors [5, 17] as compared to handwritten Latin script [7]. The hurdles in Arabic script based recognition are exist due to the cursive writing, which are mainly include context dependency, complexity in segmentation, large number of words, variations in font style, inter and intra word overlapping, space anomalies and kerning. A very limited work exists regarding unconstrained handwritten Arabic text recognition [15, 24], which motivates us to accomplish with a more sophisticated deep learning based approach that can achieve better performances.

Deep learning is one of the Machine Learning (ML) approaches that have a sound reputation in solving a huge number of classification problems that include character recognition, speech recognition, machine translation, writer identification [26], named entity recognition, protein sequences classification etc. Some of the well-known deep learning models are AlexNet [12], GoogLeNet [28], Visioual Geometry Group (VGG), Bi-directional Long Short-Term Memory (BLTM) [14], MDLSTM [9], Hierarchical LSTM [23, 27], and Extreme Learning Model (ELM) [10].

In this work, we consider the most complex dataset known as KFUPM Handwritten Arabic TexT (KHATT) [16]. Another, we use 5-fold cross validation to cross check the entire KHATT dataset for its complexities. Further, we have used an extra pre-processing step where we have prune the extra white regions that are present in Arabic text-lines. Similarly, we have properly de-skewed some of the Arabic text-lines that are skewed. We have adapted a deep learning benchmark based on MDLST networks. As accurate segmentation is one of the nightmares in Arabic cursive scripts, therefore, we use the implicit segmentation technique by using Connectionist Temporal Classification (CTC) method. The CTC has already achieved best performance regarding character recognition [4, 5, 18, 21].

The rest of the paper is organized as: In section 2, we present similar work. That mainly describes recognition systems on Arabic like script based on deep learning. Section 3 and section 5 describe detail about the proposed methodology and explain the experimental set-up. Finally, we conclude our study with future direction in section 6.

2. Similar Works

This section covers and reviews previous work on variants of Recursive Neural Networks (RNN) and KHATT data-set in the domain of Arabic script based OCR systems. In addition to that, we also some of the successful deep learning based methodologies for features extraction and classification of characters or words, are given below:

Graves and Schmidhuber [7] presented one of the pioneering works using Bidirectional Longhort Term Memory (BLSTM) for recognition of online and off-line handwritten text of English. Different types of features extracted and then pass to BLSTM and reported character recognition rate of 88.5% and 81.8%, respectively. Graves *et al*. [8] applied MDLSTM on Institut fur nachritentechnik/Ecole Nationale d'Ingénieurs de Tunis (IFN/ENIT) for the recognition of Arabic handwritten words and achieved 91.4% recognition rate. Rashid *et al*. [25] also used MDLSTM for extraction of features from raw pixels considering low-resolution images and then fed the output of MDLSTM to the CTC layer (120 units) for labeling of data. They achieved 99% recognition on Araic Printed Text Image base (APTI) data-set. Chherawala *et al*. [6] investigated a number of handcrafted features and raw pixels using data-set of IFN/ENIT. The authors employed MDLSTM and reported an accuracy of 88.8% for combined features. Ahmad *et al*. [1] conducted experiment on APTI dataset by sliding an adaptive window for feature extraction on characters and words. They reported Error Rate of 0.57 on character level and 2.12 on words. Furthermore, they also experimented on KHATT database. They retrieved highest Character Error Rate (CER) of 1.04 for character recognition. Ahmad *et al*. [3] deployed LSTM and its variants BLSTM and MDLSTM on KPTI[1] data set. They evaluated experiments on both normalized and non-normalized data. They reported better CER of 9.22% using MDLSTM.

Naz *et al*. [22] deployed 2DLSTM to compute the zoning features. They performed experiment on Urdu Printed Textline Images (UPTI) data-set and achieved character recognition rate of 93.39%. In another work, Naz *et al*. [20] evaluated an experiment on Nastaliq printed text using MDLSTM and achieved 98% recognition accuracy. Another attempt was [21] to combine Convolutional Neural Network (CNN) and MDLSTM hierarchically. Using the hybrid of CNN and MDLSTM, they achieved 98.12% recognition rate on 44 classes considering UPTI data-set. Naz *et al*. [18] used statistical features using sliding window. They performed experiments on printed Nastaliq text lines of UPTI data-set. Using RNN they achieved better performance of 94.97%. Another experiment using UPTI data-set was conducted in [19]. They employed MDLSTM-RNN and reported 96.40% recognition rate.

Here, we review specifically the published work on KHATT data-set in terms of character or word recognition. Mahmoud *et al*. [15] presented a recognition system on KHATT dataset using Hidden Markov Model (HMM). They used pixel density of text-line images, derivatives of horizontal & vertical edges, and statistical features and gradient features, separately. They reached to 51.2% accuracy for the character recognition using gradient features.

We can summarize from these works, that deep learning based RNN and its variants systems provided good results. We can also conclude from the above literature that the state of the art character recognition accuracy on complete hand-written KHATT data-set is 75.8% [4]. This work extends the work of [4] and specifically investigates the data-augmentation technique with 5 Fold-cross validations. The data-augmentation technique not only increases the input data for a single instance but also extends the level of abstraction in patterns due to synthetic variations. Another, large number of samples of a single instance or text-line is a key pre-requisite for deep learning models which significantly affects the recognition rate.

3. Proposed System

The handwritten Arabic character recognition system is presented in Figure 3. First, the main data-set is explained. Secondly, the number of data-augmentation techniques for increasing the input data images is implemented in pre-processing step. Finally, experimental setup and training process are presented.

3.1. Khatt Dataset

In this study, we use the unique text-lines of the data-set named as KHATT data-set (KFUPM Handwritten Arabic TexT). It is proposed and developed in [16]. The KHATT dataset contains forms. Each form is collected from 1000 writers and from 46 different sources. Each form consists of 4 pages. The first page contains writer information. The second and third pages contain a fixed text paragraph and a free text paragraph respectively. Finally, on the fourth page open and free text is written on a ruled baseline to avoid skew. Mahmood *et al*. [15] divided the unique text-lines images into 4, 825 (train set images) 937 (validation set images) and 966 (test set images). In this work, we exactly use the same split reported in

[1]https://github.com/rahmad77/KPTI

[15]. However, after data-augmentation the each text-line of the data-set is re-generated using data-augmentation techniques as explain in following section and the size of data-set becomes 5 times multiple of 6728 unique text lines images. The statistics of augmented data-set is illustrated in Table 1.

Table 1. The statistics of KHATT data-set after data augmentation.

Sets	Unique Text lines	5 variation by Data Augmentation
Train Set	4825	24125
Test Set	966	4685
Total Size	6728	33640

The unique characters in all unique text-line images of KHATT data-set are 116. It means that unique basic characters are 115 and one blank character. So we need a total of 116 class label for preparing the ground truth or transcript file for our proposed model. An example of Text-line and its ground truth is shown in Figure1.

a) Original Text-line image.

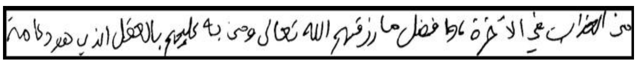
يقدرSعلىSا تمامSعملهSوSا كماله

b) Ground Truth /Transcription.

Figure 1. A sample of Text line and its ground truth.

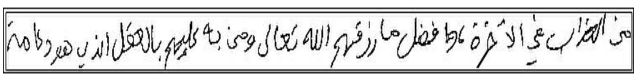
3.2. Data Augmentation

In data augmentation, we have introduced 5 extra samples for each original KHATT instance. These instances are shown in Figure 2. The first instance is the blur version of the original instance. The second instance emphasizes on the contours of the text-line. The third and fourth instances represent enhanced edges and bleed through effect. Finally, the fifth instance represent distorted image having 25% pixels as off.

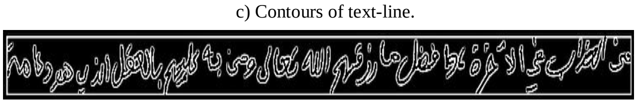


a) Original sample.

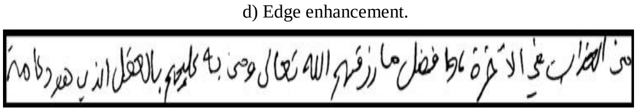
b) Blur sample.



c) Contours of text-line.



d) Edge enhancement.



e) Bleed through effect.

Figure 2. The resulted samples text-lines after Data-augmentation.

3.3. Pre-Processing

The Arabic text-lines are regenerated using bleed through, high frequency, blurriness, edge enhancement and edge more algorithms. Figure 2. [a-e] shows original image and the resultant images after data-augmentation. We also applied the pre-processing techniques for skew correction, removal of extra information and normalization from the work reported in [2]. We fixed 48 pixels for the height of text-lines using constant aspect ratio.

3.4. Training and Recognition

The main objective of this study is to evaluate and investigate the effectiveness of multiple representations of input images for training of MDLSTM model. For the first time, data-augmentation techniques deployed to evaluate the performance of the MDLSTM considering the case of Arabic characters. The architecture of our proposed network is inspired by the network in [3].

The images after applying data-augmentation methods are given to the proposed deep learning model. For more clarity, we have explained the proposed MDLSTM model in the following section.

3.4.1. MDLSTM Architecture

The proposed recognition system is a hierarchal model having three layers, mainly the input layer, hidden layer, and output layer. However, the main distinction of our proposed model is the utilization of LSTM units, such that, to scan the input image in all 4 directions. These three layers are depicted in Figure 3, with names like MDLSTM-4x4, MDLSTM-4x20, and MDLSTM-4x100. For example, in the notation MDLSTM-4x4, the first 4 represents the number of directions to be scan, while the 20 represents the number of LSTM units. The input image is first divided into patches of size (input_ block) 1x4, such that, these 1x4 patches collapse to single values. In this particular case, the height of the image is reduced by 4 times, while the length of the image is fixed. All the values of these patches are then passed to the first MDLSTM-layer; the output of the first layer is then again reduced by subsample size (here it is 2x4), and then the entire collapsed and reduced image is passed to tanh layer of size 16 units. The tanh layer is used to reduce the connection and weights. The same process is repeated until final layer reached. The activations of the final layer are then feed to the CTC layer. The CTC layer is used as final layer for prediction and alignment of predicted labels with the given label of ground truth file.

We set $1e^4$ and 0:9 for learning rate and momentum, respectively. There are total of 97 class labels therefore the output layer has 97 output units representing 96 basic characters and one for blank space.

The total numbers of epochs are set to 70 for getting best network but the proposed network on augmented KHATT data-set starts to converge after 5th epoch. One epoch takes 3 hours in training, however the time required for training is once, while test takes 0.14 Second per text-line recognition.

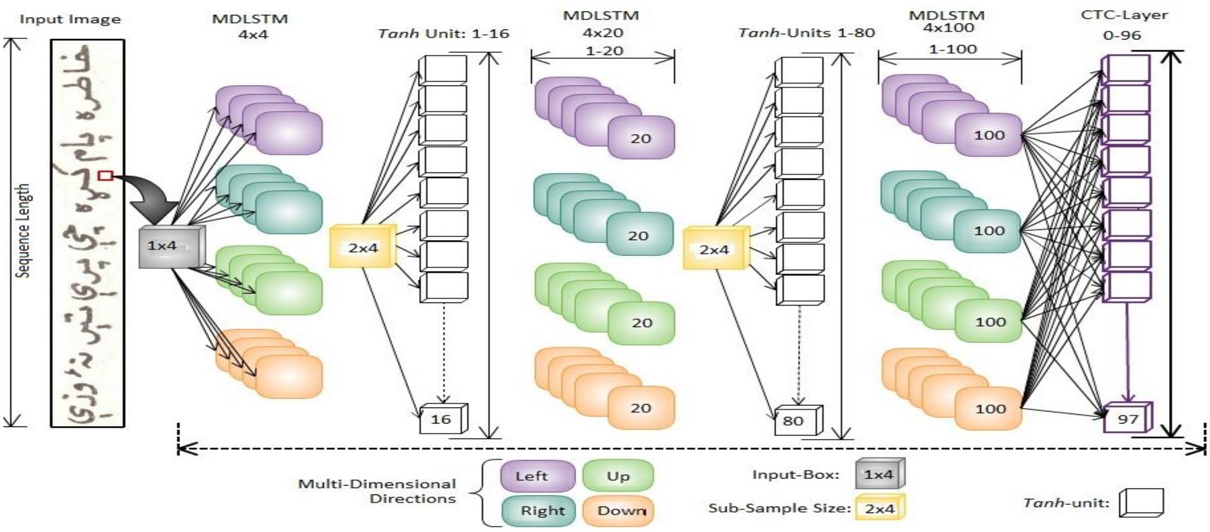


Figure 3. Architecture of Proposed System based on MDLSTM and CTC for handwritten Arabic character recognition [19].

4. Results, Discussion and Comparison

4.1. Results

The numbers of experiments are conducted on five variants after data-augmentation and original images of unique-text-lines of KHATT dataset. Mahmood *et al*. [15] proposed the split of data-set. So, we used the same split in our study but applied five techniques of data augmentation as explained in section 3.2.

In this study, the Levenshtein's [13] distance is used to report the error rate of Arabic character. Levenshtein distance is used as measurement metric between the predicted Arabic character and the annotations. Table 2 reports characters error rate and learning rate on augmented KHATT data-set.

Table 2. Performance of proposed system in term of character error and learning rates.

EXPERIMENT	Label Error Rate (%)
TRAIN SET	19
TEST SET	19.98

The results are shown in Table 3 that our proposed system outperforms in the literature on KHATT data-set. We achieved 4.22% character error rates as the state of the art results.

Table 3. Compares the results of proposed approach with the state of the art results using implicit segmentation.

Systems	Classifier	Features Set	Size	Recognition Rate (%)
Mahmoud *et al*. [15]	HMM htk tool	Adaptive gradients	Original dataset	46.13
Ahmad *et al*. [4]	MDLSTM	Raw Pixels	Original dataset	75.80
Proposed	MDLSTM	Raw Pixels	Augmented Dataset	80.02

4.2. Cross Validation

In proposed work, we use the cross validation [11] that is statistical process for evaluation and generalization of results of the proposed system on a given data-set. For this purpose, the augmented KHATT data-set is divided into equal 5 subsets. Then the proposed learning model is trained K-1 sets and test on remaining one set. Then training continued according to pictorial representation in Figure 4.

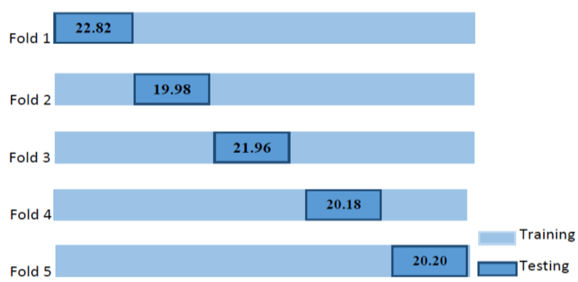


Figure 4. The Cross validation for generalization of results.

4.3. Comparison

We have not used language modeling or any post processing for improvement of results. It is all due to data augmentation technique and ability of context learning and scanning in all direction to cover the information of MDLSTM architecture with CTC layer. Therefore, we got this performance with distinct 4.22%. The experiment shows that our proposed system achieves the state-of-the-art performance due to data augmentation and deep learning architecture as compare to HMM htk tool based system in [15] and Simple text-lines based MDLSTM system [4].

5. Conclusions

This paper investigated the impact of data-augmentation to an MDLSTM on KHATT data-set in terms of Arabic character recognition. For alignment of most probable value to the predicted label, the CTC layer is used. The KHATT data-set is the challenging one data-set due to handwritten Arabic text. The literature primarily focused on the given data-set (limited text-lines). We exploited and introduced data-augmentation based techniques for providing more patterns to enrich the input feature space for deep learning model. In addition to raw handwritten Arabic text-line images, we used images of bleed through, blurred, enhanced, more enhanced and contour in order to enhance the performance of the MDLSTM. Furthermore, we also performed cross validation for generalization of result on KHATT data-set and reported results with considerable high margin of 4.22%. The average Arabic character recognition rate is up to 80.02% on KHATT dataset.

In future, the KHATT dataset may need special treatment to achieve the best performance. The major issue is the limited size of the dataset. It can be extended by adding more variations using data augmentation technique. The more we have the data the more deep-LSTM based network could be used to treat the Arabic text-line recognition.

References

[1] Ahmad I., Mahmoud S., and Fink G., “Open-Vocabulary Recognition of Machine-Printed Arabic Text Using Hidden Markov Models,” *Pattern Recognition*, vol. 51, no. c, pp. 97-111, 2016.

[2] Ahmad R., Rashid S., Afzal M., Liwicki M., Dengel A., and Breuel T., “A Novel Skew Detection and Correction Approach For Scanned Documents,” *in Proceedings of 12th International IAPR Workshop on Document Analysis Systems*, At Santorini, 2016.

[3] Ahmad R., Afzal M., Rashid S., Liwicki M., Breuel T., and Dengel A., “KPTI: Katib’spashto Text Image Base and Deep Learning Benchmark,” *in Proceedings of 15th International Conference on Frontiers in Handwriting Recognition*, Shenzhen, pp. 453-458, 2016.

[4] Ahmad R., Naz S., Afzal M., Rashid S., Liwicki M., and Dengel A., “KHATT: a Deep Learning Benchmark on Arabic Script,” *in Proceedings of 14th IAPR International Conference on Document Analysis and Recognition*, Kyoto, pp. 10-14, 2017.

[5] Ahmed S., Naz S., Swati S., Razzak I., Umar A., and Khan A., “UCOM Offline Dataset- An Urdu Handwritten Dataset Generation,” *The International Arab Journal of Information Technology*, vol. 14, no. 2, pp. 239-245, 2017.

[6] Chherawala Y., Roy P., and Cheriet M., “Feature Design for Offline Arabic Handwriting Recognition: Handcrafted vs Automated?,” *in Proceedings of 12th International Document Analysis and Recognition*, Washington, pp. 290-294, 2013.

[7] Graves A. and Schmidhuber J., “Offline Handwriting Recognition with Multidimensional Recurrent Neural Net-Works,” *in Proceedings of the 21st International Conference on Neural Information Processing Systems*, Red Hook, pp. 545-552, 2009.

[8] Graves A., Liwicki M., Fernandez S., Bertolami R., Bunke H., and Schmidhuber J., “A Novel Connectionist System for Unconstrained Handwriting Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 855-868, 2009.

[9] Graves A., *in Guide to OCR for Arabic Scripts*, Springer, 2012.

[10] Huang G., Song S., Gupta J., and Wu C., “Semi-Supervised and Unsupervised Extreme Learning Machines,” *Transactions on Cybernetics*, no. 44, no. 12, pp. 2405-2417, 2014.

[11] Kohavi R., “A Study of Cross-Validation and Boot-Strap For Accuracy Estimation and Model Selection,” *in Proceedings of the International Joint Conference on Artificial Intelligence*, Montreal, pp. 1137-1145, 1995.

[12] Krizhevsky A., Sutskever I., and Hinton G., “Imagenet Classification with Deep Convolutional Neural Networks,” *in Proceedings of Neural Information Processing Systems*, Lake Tahoe, pp. 1097-1105, 2012.

[13] Levenshtein V., “Binary Codes Capable of Correcting Deletions, Insertions and Reversals,” *in Soviet Physics Doklady*, vol. 10, pp. 707-710, 1966.

[14] Liwicki M., Graves A., Bunke H., and Schmidhuber J., “A Novel Approach To on-Line Handwriting Recognition Based on Bidirectional Long Short-Term Memory Networks,” *in Proceedings of the 9th International Conference on Document Analysis and Recognition*, United States, pp. 367-371, 2007.

[15] Mahmoud S., Ahmad I., Al-Khatib W., Al-Shayeb M., Parvez M., Margner V., and Fink G., “Khatt: An Open Arabic Offline Handwritten Text Database,” *Pattern Recognition*, vol. 47, no. 3, pp. 1096-1112, 2014.

[16] Mahmoud S., Ahmad I., Alshayeb M., Al-Khatib W., Parvez M., Fink G., Margner V.,

and El Abed H., "Khatt: Arabic Offline Handwritten Text Database," *in Proceedings of* International Conference on *Frontiers in Handwriting Recognition*, Bari, pp. 449-454, 2012.

[17] Naz S., Hayat K., Razzak M., Anwar M., Madani S., and Khan S., "The Optical Character Recognition of Urdu-Like Cursive Scripts," *Pattern Recognition*, vol. 47, no. 3, pp. 1229-1248, 2014.

[18] Naz S., Umar A., Ahmad R., Ahmed S., Shirazi S., and Razzak M., "Urdu Nastaliq text Recognition System Based on Multi-Dimensional Recurrent Neural Network and Statistical Features," *Neural Computing Application*, vol. 26, no. 8, pp. 1-13, 2016.

[19] Naz S., Umar A., Ahmad R., Ahmed S., Shirazi S., Siddiqi I., and Razzak M., "Offline Cursive Nastaliq Script Recognition Using Multidimensional Recurrent Neural Networks with Statistical Features," *NeuroComputing*, vol. 177, pp. 228-241, 2016.

[20] Naz S., Umar A., Ahmad R., Razzak M., Rashid S., and Shafait F., "Urdu Nastaliq Text Recognition Using Implicit Segmentation Based on Multi-Dimensional Long Short Term Memory Neural Networks," *SpringerPlus*, vol. 5, no. 1, pp. 2010, 2016.

[21] Naz S., Umar A., Ahmad R., Siddiqi I., Ahmed S., Razzak M., and Shafait F., "Urdu Nastaliq Recognition Using Convolutional-Recursive Deep Learning," *NeuroComputing*, vol. 243, pp. 80-87, 2017.

[22] Naz S., Ahmed S., Ahmad R., and Razzak M., "Zon-ing Features and 2DLSTM for Urdu Text-Line Recognition," *Procedia Computer Science*, vol. 96, pp. 16-22, 2016.

[23] Pan P., Xu Z., Yang Y., Wu F., and Zhuang Y., "Hierarchi-Cal Recurrent Neural Encoder for Video Representation with Application to Captioning," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Hynes Convention Center in Boston, pp. 1029-1038, 2016.

[24] Parvez M and Mahmoud S, "Offline Arabic Handwritten Text Recognition: A Survey," *ACM Computing Surveys*, vol. 45, no. 2, pp. 23, 2013.

[25] Rashid S., Schambach M., Rottland J., and Null S., "Low Resolution Arabic Recognition With Multidimensional Recurrent Neural Networks," *in Proceedings of the 4th International Workshop on Multilingual OCR*, New York, pp. 6, 2013.

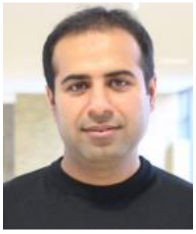
[26] Rehman A., Naz S., and Razzak M., "Writer Identification Using Machine Learning Approaches:A Comprehensive Review," *in Multimedia Tools and Applications*, vol. 78, no. 8, pp. 10889-10931, 2018.

[27] Yang Z., Yang D., Dyer C., He X., Smola A., and Hovy E., "Hierarchical Attention Networks for Document Classification," *in Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, pp. 1480-1489, 2016.

[28] Zhong Z., Jin L., and Xie Z., "High Performance Off-Line Handwritten Chinese Character Recognition Using Googlenet and Directional Feature Maps," *in Proceedings of the 13th International Conference on Document Analysis and Recognition*, Tunis, pp. 846-850, 2015.

Riaz Ahmad has received his PhD Degree from Technical University of Kaiserslautern, Germany, 2018 Currently; he is working as a faculty member at Shaheed Benazir Bhutto University, Sheringal, Pakistan. His areas of research include document image analysis, image processing and Optical Character Recognition. More specifically, his work examines the challenges pose by cursive script languages in the field of OCR systems.

Saeeda Naz received her PhD degree in Jan 2016 (Hazara University, Pakistan). Currently, she is Assistant Professor by designation and Head of Computer Science Department at GGPGC No.1, Abbottabad, Higher Education Department of Government of KPK, Pakistan, since 2008. Her areas of interest are Optical Character Recognition, Pattern Recognition, Machine Learning, Medical Imaging and Natural Language Processing and Multimedia.



Muhammad Afzal received his PhD degree in 2016 (German Research Center for Artificial Intelligence, University of Kaiserslautern, Germany). His research interests include generic segmentation framework for natural, document and, medical images, scene text detection and recognition.

Sheikh Rashid obtained Doctor of Engineering (Dr.-Ing.) from University of Kaiserslautern Germany. Research Center for Artificial Intelligence (DFKI) Kaiserslautern. Currently he is working as a director at Artificial Intelligence Research Lab, Al Khwarizmi Institute of Computer Science (KICS), UET Lahore, Pakistan.

Marcus Liwicki received his PhD degree from the University of Bern, Switzerland, in 2007. Currently he is an apl.-professor in the University of Kaiserslautern and a senior assistant in the University of Fribourg. His research interests include machine learning, pattern recognition, artificial intelligence, human computer interaction, digital humanities, knowledge management, document analysis, and graph matching.

Andreas Dengel is a member of the Management Board as well as Scientific Director at the German Research Center for Artificial Intelligence (DFKI) in Kaiserslautern where he is leading the Smart Data & Knowledge Services Research Department. In 1993 he became a Professor at the Computer Science Department of the University of Kaiserslautern. Since 2009 he also holds a Honorary Professorship at the Dept. of Computer Science and Intelligent Systems, Graduate School of Engineering of the Osaka Prefecture University.