A Deep Convolutional and Recurrent Approach for Large Vocabulary Arabic Word Recognition

Faten Ziadi[1,2][a], Imen Ben Cheikh[1][b] and Mohamed Jemni[1][c]
[1]*University of Sousse, ISITCom, 4011, Sousse, Tunisia*
[2]*Latice Laboratory, ENSIT, University of Tunis, Tunis, Tunisia*

Keywords: CNN, LSTM, Arabic Writing, Large Vocabulary, Linguistic Knowledge, APTI Dataset.

Abstract: In this paper, we propose a convolutional recurrent approach for Arabic word recognition. We handle a large vocabulary of Arabic decomposable words, which are factored according to their roots and schemes. Exploiting derivational morphology, we have conceived as the first step a convolutional neural network, which classifies Arabic roots extracted from a set of word samples int the APTI database. In order to further exploit linguistic knowledge, we have accomplished the word recognition process through a recurrent network, especially LSTM. Thanks to its recurrence and memory cabability, the LSTM model focuses not only prefixes, infixes and suffixes listed in chronological order, but also on the relation between them in order to recognize word patterns and some flexional details such as, gender, number, tense, etc.

1 INTRODUCTION

Recognition is an area that covers various fields such as facial recognition, fingerprint recognition, image and character recognition, number recognition, etc. For decades, considerable progress has been made in handwriting recognition.

Our work focuses primarily on this field of research, specifically Arabic writing recognition.

Arabic is a mother Semitic tongue (Ibrahim, Bilmas and Babiker, 2013) spoken by hundreds of millions of people in Middle Eastern countries.

The following figure (Figure 1) illustrates the 28 characters of the Arabic alphabet.

The morphological complexity of written Arabic and its cursivity remains a very broad subject area which has known in recent years a great progress in diverse fields.

Indeed, the complexity of the recognition process depends on the type of script (printed or handwritten), the approach (holistic, pseudo-global or analytic) (Touj,Ben Amara and Amiri, 2007) (Avila, 1996) and the vocabulary size (reduced, large).

خ Kh	ح Haa'	ج Jiim	ث Thaa'	ت Taa'	ب Baa'	أ 'Alif
ص Saad	ش Shiin	س Siin	ز Zaayn	ر Raa'	ذ (Th)aal	د Daal
ق Qaaf	ف Faa'	غ Ghayn	ع 'Ayn	ظ (Th)aa'	ط Taa'	ض Daad
ي Yaa'	و Waaw	ه Haa'	ن Nuun	م Miim	ل Laam	ك Kaaf

Figure 1: The 28 character of the Arabic alphabet.

Numerous approaches have been proposed, dealing with letters and / or pseudo-words, and several works have experienced statistical, neural, stochastic methods on different types of Arabic documents.

Among these approaches, we will focus on neural networks by presenting two variants: convolutional networks and recurrent networks with LSTM memory.

In this context, we have chosen to combine these two architectures in order to recognize a large vocabulary of Arabic words.

[a] https://orcid.org/0000-0002-1966-5630
[b] https://orcid.org/0000-0003-2119-1312
[c] https://orcid.org/0000-0001-8841-5224

Ziadi, F., Ben Cheikh, I. and Jemni, M.
A Deep Convolutional and Recurrent Approach for Large Vocabulary Arabic Word Recognition.
DOI: 10.5220/0010814800003122
In *Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2022)*, pages 213-220
ISBN: 978-989-758-549-4; ISSN: 2184-4313

The remainder of this paper is structured as follows: Section 2 displays the main characteristics of the Arabic script. Section 3 describes not only the performance of convolutional network and LSTM, but also various related works using this two models in the recognition field.

Section 4 reviews the related works suggested in the literature using respectively CNN model, LSTM model, and the hybridation of these two models for writing recognition. Section 5 introduces our suggested approach. Then, section 6 reports our preliminary results. Finally, section 7 summarizes the paper and highlights major directions for future research.

2 ARABIC LANGUAGE CHARACTERISTICS

Thanks to its stability, Arabic is a rich language in terms of its morphology, phonology and vocabulary (Cheriet and Beldjehem, 2006) (Ben Hamadou, 1993) (Kanoun,Alimi and Lecourtier, 2005) (Kammoun and Ennaji, 2004). An Arabic word can be decomposed into smaller units. If it derives from a root, it is said to be decomposable into morphemes (root, prefix, infix and suffix). Figure 2 illustrates our vision of Arabic words derived from several roots and conjugated according to different schemes.

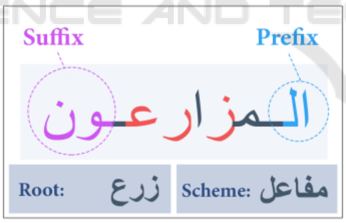

Figure 2: Our vision of an Arabic word.

Based on this principle, we were able to enlarge our vocabulary not only by combining words derived from the same root with several schemes, but also by applying conjugation rules (tense, number, gender, etc.).

Figure 3 presents several samples of words that are derived from the root صرف and conjugated according to different schemes. For instance, the root صرف derived according to the scheme انفعل gives the verb word انصرف. The latter conjugated in the past tense with the masculine plural pronoun “they” gives the verb انصرفوا.

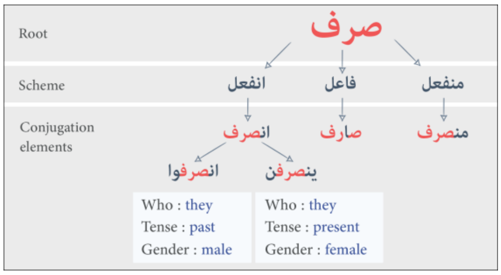

Figure 3: Various words derived from the root “صرف”.

In this context, we aim to classify Arabic word roots using a convolutional network. The output of this model will be used by a recurrent network (LSTM) to recognize whole word.

3 CNN AND LSTM OVERVIEW

3.1 Convolutional Neural Network

Convolutional Neural Network (CNN), proposed by LeCun (LeCun and Bengio, 1995), is a neural network model that combines three key architectural ideas, namely local receptive fields, shared weight and spacial subsampling. This network receives 2-D input (e.g. an image, a voice signal) and extracts high-level characteristics through a series of hidden layers. These layers consist of:

- Convolution layer includes a set of filters whose parameters need to be learned. These filters have the same shape as the input but with smaller dimensions. In the learning process, the filter of each convolutional layer is tiled across the input space (in the case of an image, for example, the filter is slid across the width and height of the input) and the input product is calculated. This calculation of the entire input leads to a filtering feature map.
- Pooling layer operates upon each feature map. It reduces the spatial size of the representation in order to reduce the number of parameters and computational time, and hence to also control over-fitting. Max-Pooling is a common approach that partitions the input space into non-overlapping regions and chooses the maximum value of each region.
- Fully connected layer takes the output of convolution/pooling and predicts the best label to describe the image.

3.2 LSTM Network

Recurrent neural networks (RNN) have recurrent connections. At time t, the RNN takes into account a number of past states, following the short term memory principle. We can deduce that RNNs are adapted for analyzing contextual applications, and more particularly for processing sequences. However, this "short-term memory" is not anymore sufficient in most applications involving long time differences (typically the classification of video sequences). In fact, "classic" RNNs are only able to memorize the so-called near past, and start to forget after multiple iterations.

These networks take as input: 1) the current input and 2) what they have observed previously over time. The decision reached by a recurrent neuron at time t-1 affects the learning process at instant t. However, RNNs suffer from the vanishing gradient problem.

As more layers using activation functions are added, the gradients of the loss function begin to approach to zero. This problem hampers learning of long data sequences and therefore makes the network hard to train. LSTM networks provide a solution to this problem.

In addition to its external connections, A LSTM neuron has a recurrent-self connection with a constant coefficient equal to 1. This allows us to save the successive states of the neuron, varying, from one iteration to another. Multiplication gates (input-gate, output-gate and forget-gate) protect the current state of memory (the cell state) and the memory of the whole network. The structure of the LSTM neuron is illustrated in Figure 4.

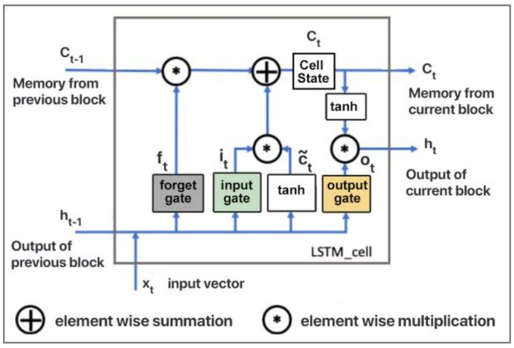

Figure 4: LSTM node architecture.

Equations (1) and (2) are used to calculate the current memory state (Ct) and current output (ht), respectively.

$$C_t = \sigma(f_t * C_{t-1} + i_t * \tilde{C}_t) \quad (1)$$

$$h_t = \tanh(C_t) * o_t \quad (2)$$

Where C_{t-1} is the previous memory cell, $\tilde{C}$ is the past value of the current memory. i_t, o_t and f_t stand for input-gate, output-gate and forget-gate, respectively. These outputs are estimated by the current neuron input and the previous neuron output.

4 RELATED WORKS

4.1 CNN for Writing Recognition

A special focus is put on Arabic word recognition.

Using CNN, various methods have been proposed and high recognition rates have been reported for handwriting recognition in English (Bai,Chen,Feng and Xu, 2014) (Yuan, Bai, Jiao and Liu, 2012), Chinese (Wu, Fan, He, Sun and Naoi, 2014) (Zhong, Jin and Feng, 2015) (Yang, Jin, Xie and Feng, 2015) (He, Zhang, Mao and Jin, 2015) (Zhong, Jin and Xie, 2015), Hangul (Kim and Xie, 2015), Malayalam (Anil, Manjusha, Kumar and Soman, 2015), Devanagari (Acharya, Pant, and Gyawali, 2015) (Singh,Verma and Chaudhari, 2016), Telugu (Soman, Nandigam and Chakravarthv, 2013).

The majority of these methods are based on Convolutional Neural Networks. The first CNNs date back to the 1980s with the work of K. Fukushima (Fukushima, 1980), but it was in the 1990s that these networks were popularized with the work of Y. Le Cun et al. about character recognition (LeCun, Boser, Denker, Henderson, Howard, Hubbardet and Jackel, 1990). In a study conducted by (Poisson and Gaudin, 2001), a particular neuron network TDNN (Time Delay Neural Network) was implemented for handwriting recognition. Their recognition results show the input of a hidden Markov model.

4.2 LSTM for Writing Recognition

Unidirectional and bidirectional LSTM networks have been tested in many handwriting applications. In (Graves, Fernandez, Liwicki, Bunke and Schmidhuber, 2008), a combined BLSTM-CTC network was put forward in an online handwriting dataset for unconstrained online handwriting recognition. In a study carried out by (Graves and Schmidhuber, 2009), a multidimensional LSTM was combined with CTC for offline handwriting recognition. (Su and Lu, 2015) set forth a segmentation-free approach in order to recognize text in scene images. A RNN was adapted to classify the word sequential characteristics obtained using Histograms of Oriented Gradients (HOG). (Graves, Fernandez, Liwicki, Bunke and Schmidhuber, 2007) offered a system that is able to manage online handwritten data. It is based on a RNN with an output layer designed for sequence labeling, associated with a probabilistic language model. (Mioulet, 2015) combined a bidirectional LSTM with a connectionist temporal classification (BLSTM-CTC) for handwriting recognition and keyword detection.

4.3 A Hybrid CNN-LSTM Network for Arabic Recognition

Several works have combined CNN and LSTM networks for Arabic language treatment.

(Hamdi et al., 2019) suggested an online Arabic character recognition system based on hybrid Beta-Elliptic model (BEM), CNN, bidirectional LSTM and SVM. Online handwriting trajectory features are extracted using BEM, while generic features are extracted using CNN. The classification is achieved by combining DBLSTM and SVM models.

To bridge the communication gap between deaf and hearing communities, (Agrawal and Urolagin, 2020) introduced a 2-way Arabic sign language translator using CNN-LSTM and NLP models for translating texts into signs and vice versa. CNN layers perform feature extraction and LSTM layers provide temporal sign prediction.

(Al Omari, Al-Hajj, Sabra and Hammami, 2019) recommended a combined CNN and LSTM model for Arabic sentiment classification. This model was tested on the Arabic Health Services (AHS) dataset. It achieved more interesting results on the Main-AHS and the Sub-AHS datasets compared with those of the previous methods (Alayba, Palade, England and Iqbal, 2018).

Since the combination of CNN and LSTM has proven its effectiveness for improving the quality of Arabic writing analysis, we use this hybridization of different linguistic knowledge in order to recognize Arabic decomposable words.

5 APPROACH PROPOSAL

With the recent emergence of deep learning technology, the aforementioned research studies described in previous sections inspired our work.

Thus, we have tried to combine two neural network variants, namely CNN and LSTM models.

The first step of our work consists in implementing a convolutional neural network that learns and recognizes roots from word images. In the second step, we feed a LSTM with word samples. This LSTM is specialized in recognizing words derived from the predicted root. It is conceived to recognize the schemes and the flexional details of the candidate word.

The CNN model comprises these different layers:

- A first convolutional layer with 64 filters, each with the size of (3*3), followed by a "relu" activation function.
- A max-pooling layer with a filter size (2*2).
- A second convolutional layer with 64 filters each of size (3*3) and a "relu" activation function.
- A max-pooling layer with a filter size (2*2).
- A flattening operation that flatten all data into a single vector.
- A Dense layers where each neuron is connected to all neurons in the previous layer. "relu" is the activation function of this layer.
- A Dense layer with 10 neurons and a "softmax" activation function to classify the 10 roots.

Note that the "Dropout" technique is used after each layer to prevent overfitting.

Figure 5 shows the implemented CNN architecture and decisions made at each layer.

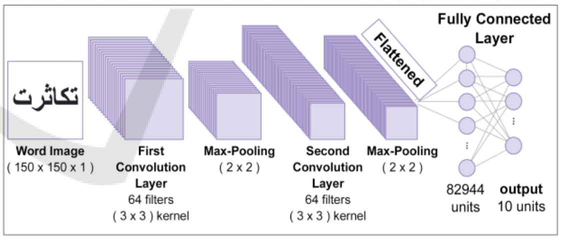

Figure 5: CNN architecture.

As mentioned before, after the classification phase, the recognized root and the whole word image will be passed to the LSTM model. We have started with a static segmentation where each word is split into vertical bands of a constant number of columns (25) as shown in Figure 6. Each band is converted into a 150 sized vector (image width), where each value corresponds to the sum of band row . At each time step, LSTM will be fed by six vectors.

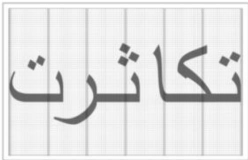

Figure 6: Splitting the word “ تكاثرت”.

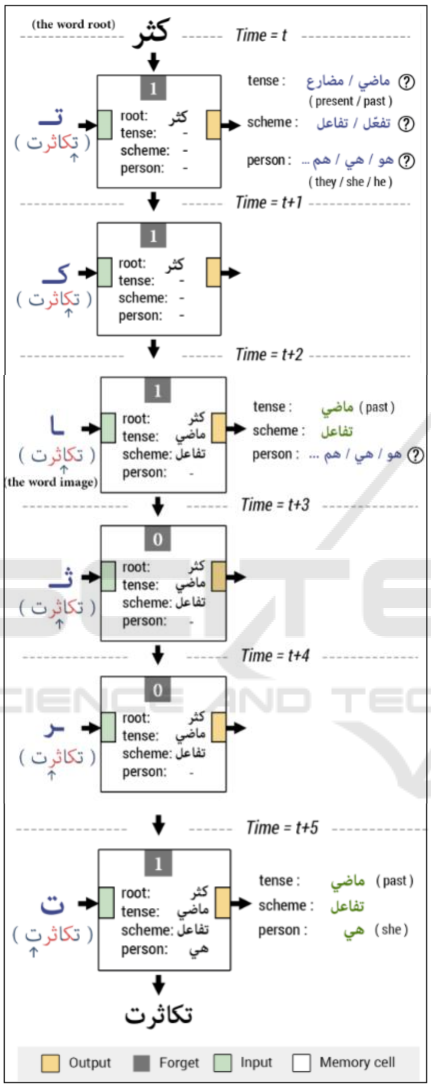

Figure 7: An example of the word “ تكاثرت ” recognition by the LSTM network.

Figure 7 displays multiple information propagation that allow the LSTM network to learn word derivations and flexions with their corresponding roots (CNN output).

In this example, at time t, the cell accepts two information: 1) the root كثر(It/he becomes abundant) and 2) the first letter “ت ” of the verb word to be recognized “ تكاثرت” (i.e. She has been reproduced).

To recall these two relevant data, the forget gate is activated (=1). Now, this neuron is going the try to predict the conjugation time, the scheme, etc. and its memory will save these information which will be exploited by the next cell.

At t+1, the second letter (“ك ”) is the input of the LSTM cell, the last memory will be used to go on, while the current memory will ignore the current input (it is actually the first root consonant) and decide to save conjugation information to propagate to the following cell at t+2.

At t+2, the following cell will determine the time (ماضي. i.e. past tense) and the scheme (تفاعل) thanks to the stored information and relevant input “ا”. Details concerning the second and third root letters at t+3 and t+4 are not memorized.

Finally, at time t+5, the last letter of the word (suffix) will determine the personal pronoun (هي. i.e. She). Once the root, time, scheme and different conjugation elements are known, the word is reconstituted and well-recognized.

6 EXPERIMENTAL RESULTS

6.1 Dataset

Our approach is tested using a database that contains Arabic printed word images, of various sizes and fonts, derived from three consonant roots (see Figure 8), displays samples extracted from the APTI database.

Figure 8: Samples of words extracted from the APTI database.

The corpus includes a set of Arabic words images. We manipulated 17600 samples of 1100 words derived from tri-consonant roots following different schemes. For data collection, we have gathered words from the APTI database and organized them into 10 sets. Each set corresponds to one root and contains many words that follow several schemes conjugation forms (e.g. different tenses and pronouns). Each word is given in 16 samples. The distribution of the words samples is as follows : 60% for the training, 25% for the validation and 15% for the test.

6.2 CNN Results

Both CNN and LSTM architectures have been created based on the Keras Framework. Before fitting the CNN model, we must compile it. Several parameters are necessary to carry out this phase like the optimizer, the loss function and optionally the evaluation metrics.

To compile the model,we choose the Root Mean Square Propagation (RMSPROP) optimizer, Categorical Cross-Entropy loss (also called Softmax Loss) and accuracy metric to evaluate our model.

In order to adjust the parameters of a model, a set of hyperparameters must be chosen appropriately. In our work we have used hyperparameters that we have adjusted manully. In fact, we choose them according to our experience. We then train the model, evaluate its accuracy and repeat the process. This loop is repeated until a satisfactory accuracy is noted.

After the training phase, estimated at about ten hours, the CNN model achieved a 97% classification rate after 30 epochs.

Some false predictions were detected during the classification phase, such as the images presented in the following figure (Figure 9) where the predicted label is the root "كثر", while the true root is "كبر" which is too similar to the predicted label.

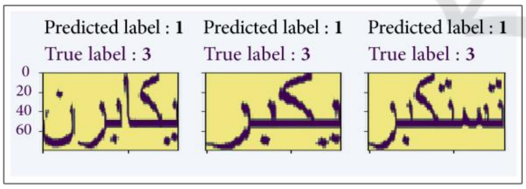

Figure 9: Examples of word predictions derived from the root كبر.

Figure 10 and Figure 11 represent curves illustrating the two metrics: accuracy and loss, respectively.

Based on the two figures, we notice that:

- The classification rate is 97%.
- The plot of training loss decreases to a point of stability.
- The plot of validation loss decreases to a point of stability and has a small gap with the training loss.

We can deduce that this is the case of a good fit.

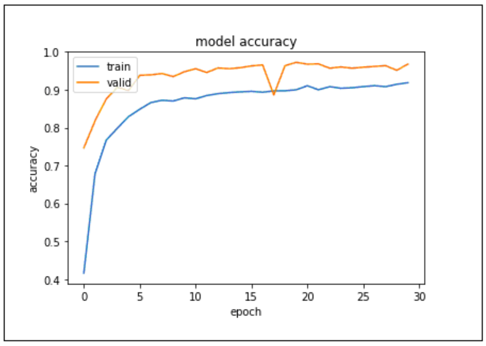

Figure 10: Performance curves on training and validation data.

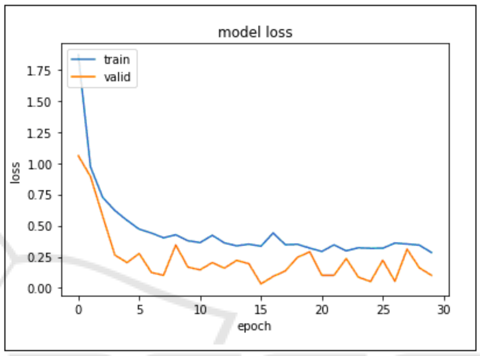

Figure 11: Representative curves of loss function on training and validation data.

6.3 LSTM Results

Before fitting it in the LSTM, each word is split into vertical bands with a constant number of columns (25). In the short term, we intend to operate the vertical projection of a word and send bands delimited by local minima.

As first, the LSTM model was evaluated using a corpus composed of 5000 samples. The first results are promising. In fact, the recognition rate of the LSTM algorithm is equal to 87.2% .

In fact, false results are detected mainly because of the segmentation phase. As mentioned, in this work we started with a static segmentation where each word sample is divided into 6 bands to detect each letter of the word. Since we have words composed of more than 6 letters as well as less than 6 letters, letters are not properly fed to the LSTM.

In addition, we will expand the lexicon by using more than 100 roots. Hence, we build not only vocabulary of about 4000 words, but also a corpus of around 60000 samples for training and testing a hybrid CNN-LSTM model.

We notice that: 1) the LSTM model browses the entire word from right to left and learns to focus on schemes and flections rather than on the consonants of the root. 2) We are confident that our model is reliable thanks to the ability of memory cells. The latter are able to store information and help neighbour neurons acquire relevant knowledge. Our model fits well with the Arabic lexicon philosophy by taking into account the significant coherence between prefixes and suffixes and their complicity in incarnating the time of an action verb and its corresponding pronoun.

6.4 Comparison with Related Works

Table 1 illustrates models that combined CNN and LSTM for Arabic recognition. It also indicates the rates obtained on specific datasets.

Table 1: Arabic rates recognition.

Model	Dataset	Result
BEM + CNN + BLSTM + SVM for an online Arabic character recognition (Hamdi et al., 2019)	LMCA	99.11 %
	Online-Khatt	93.98 %
CNN-LSTM for Arabic sign language translator (Agrawal and Urolagin, 2020)	Minisi MN. Arabic sign language dictionary	88.67 %
CNN-LSTM for textual reviews (Al Omari, Al-Hajj, Sabra and Hammami, 2019)	Main-AHS	88 %
	Sub-AHS	96 %
	Ar-Twitter	84 %
	ASTD	79 %
	OCLAR	90 %
CNN-LSTM for an Arabic sentiment analysis (Alayba, Palade, England and Iqbal, 2018)	Main-AHS	94 %
	Sub-AHS	95 %
Our CNN model for classification	APTI Dataset	97 %
Our CNN-LSTM for word recognition	APTI Dataset	87.2 %

7 CONCLUSIONS

In this work, we proposed a deep recurrent approach based on a hybrid CNN-LSTM model for tri-consonantal Arabic word recognition.

Preliminary experiments were carried out on a corpus including 17600 samples collected from the APTI Dataset. The CNN model achieved a 97% classification rate in root word recognition.

Preliminary experiments with the LSTM model gave a score equal to 87.2%. The obtained results are motivating. The combination of LSTM and CNN is expected to be more and more interesting, due to the significant coherence between the LSTM model, (with a special focus on its concepts of forgetting and memory), and the Arabic philosophy of constructing words around the root, playing with prefixes, infixes and suffixes and respecting Arabic derivations and conjugation patterns.

Our future works will put special focus on: 1) We will expand the vocabulary to handle with hundreds of roots for training, validating and testing the performance of the network on large datasets. 2). Apply cross-validation to demonstrate the reliability of our model. 3) Experiment the reliability of the LSTM model to achieve the recognition, notably, of the scheme, the tense and the pronoun, while changing the way of word segmentation.

REFERENCES

Ibrahim, M.N., Bilmas, S.N, Babiker, A. (2013). A framework of an online self-based learning for teaching arabic as second language (TASL).*Fifth International Conference on Computational Intelligence, Modelling and Simulation*, pp. 255-260.

Touj, S., Ben Amara, N., Amiri, H. (2007). A hybrid approach for off-line arabic handwriting recognition based on a planar hidden markov modeling. *ICDAR'07,* Brazil, 964- 968J.

Avila, J.M. (1996). Optimisation de modèles markoviens pour la reconnaissance de l'écrit. *PHD Thesis,* University of Rouen.

Cheriet, M., Beldjehem, M. (2006). Visual processing of arabic handwriting: challenges and new directions.*SACH'06,* India, pp 1-21.

Ben Hamadou, A. (1993). Vérification et correction automatiques par analyse affixale des textes écrits en langage naturel. *PHD Thesis*, Faculty of Sciences of Tunis.

Kanoun, S., Alimi, A., Lecourtier, Y. (2005). Affixal approach for arabic decomposable vocabulary recognition: a validation on printed word in only one font. *ICDAR'05,* Seoul.

Kammoun, W., Ennaji, A. (2004). Reconnaissance de textes arabes à vocabulaire ouvert. *CIFED,* France.

LeCun, Y., Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks, 3361(10), 1995.Bai, J., Chen, Z., Feng, B. ,Xu, B. (2014). Image character recognition using deep convolutional neural network learned from different languages.IEEE International Conference on Image Processing (ICIP),* pp. 2560-2564.

Yuan, A., Bai, G., Jiao, L., Liu, Y. (2012). Offline handwritten english character recognition based on convolutional neural network. *Document Analysis Systems (DAS), 10th IAPR International Workshop*, pp. 125-129.

Wu, C., Fan, W., He, Y., Sun, J., Naoi, S. (2014). Handwritten character recognition by alternately trained relaxation convolutional neural network. *Frontiers in Handwriting Recognition (ICFHR), 14th International Conference,* pp. 291-296.

Zhong, Z., Jin, L., Feng, Z. (2015). Multi-font printed chinese character recognition using multipooling convolutional neural network.*Document Analysis and Recognition (ICDAR), 13th International Conference,* pp. 96-100.

Yang, W., Jin, L., Xie, Z., Feng, Z. (2015). Improved deep convolutional neural network for online handwritten chinese character recognition using domain-specific knowledge. *Proc. Document Analysis and Recognition (ICDAR), 13th International Conference,* pp. 551-555.

He, M., Zhang, S., Mao, H., Jin, L. (2015). Recognition confidence analysis of handwritten chinese character with CNN.*Proc. Document Analysis and Recognition (ICDAR), 13th International Conference,* pp. 61-65.

Zhong, Z., Jin, L., Xie, Z. (2015). High performance offline handwritten Chinese character recognition using GoogLeNet and directional feature maps.*Proc. Document Analysis and R (ICDAR), 13th International Conference,* pp. 846-850.

Kim, I.-J., Xie, X. (2015). Handwritten hangul recognition using deep convolutional neural networks. *International Journal on Document Analysis and Recognition (IJDAR),* 18, pp. 1-3.

Anil, R., Manjusha, K., Kumar, S.S., Soman, K.P. (2015). Convolutional neural networks for the recognition of malayalam characters.*Proc. Document Analysis and Recognition (ICDAR), 13th International Conference,* pp. 1041-1045.

Acharya, S., Pant, A.K., Gyawali, P.K. (2015). Deep learning based large scale handwritten devenegari character recognition. *9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, pp. 1-6.

Singh, P., Verma, A., Chaudhari, N.S. (2016). Deep convolutional neural network classifier for handwritten devanagari character recognition. *Information Systems Design and Intelligent Applications: Proceedings of Third International Conference INDIA, 2,* pp. 551-561.

Soman, S.T., Nandigam, A., Chakravarthv, V.S. (2013). An efficient multiclassifier system based on convolutional neural network for offline handwritten Teugu character recognition.*Proc. Communications (NCC), 2013 National Conference,* pp. 1-5.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics,* 36(4):193–202.

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbardet, W., Jackel, L.D. (1990). Handwritten digit recognition with a backpropagation network. *Advances in Neural Information Processing Systems,* pages 396–404.

Poisson, E., Viard-Gaudin, C. (2001). Réseaux de neurones à convolution: reconnaissance de l'écriture manuscrite non contrainte.*Polytechnic School of the University of Nantes,* France.

Graves, A., Fernandez, S., Liwicki, M., Bunke, H., Schmidhuber, J. (2008). Unconstrained online handwriting recognition with recurrent neural networks.*Advances in Neural Information Processing Systems,* 20:1-8.

Graves, A., Schmidhuber, J. (2009). Offline handwriting recognition with multidimensional recurrent neural networks.*Advances in Neural Information Processing Systems,* 21:545-552.

Su, B., Lu, S. (2015). Accurate scene text recognition based on recurrent neural network. *In: Cremers D., Reid I., Saito H., Yang MH. (eds) Computer Vision -- ACCV 2014. ACCV 2014. Lecture Notes in Computer Science, vol 9003.* Springer, Cham.

Graves, A., Fernandez, S., Liwicki, M., Bunke, H., Schmidhuber, J. (2007). Unconstrained online handwriting recongnition with recurrent neural networks.*Advances in Neural Information Processing Systems.*

Mioulet, L. (2015). Reconnaissance de l'écriture manuscrite avec des réseaux récurrents, *PhD Thesis*, University of Rouen.

Hamdi, Y., et al. (2019). Hybrid DBLSTM-SVM based Beta-elliptic-CNN Models for Online Arabic Characters Recognition. *2019 InternationalConference on Document Analysis and Recognition (ICDAR).* IEEE.

Agrawal, T., Urolagin, S. (2020). 2-way Arabic Sign Language translator using CNNLSTM architecture and NLP.*Accepted in ACM, Ei Compendex and Scopus, ISI Web of Science, International Conference on Big Data Engineering and Technology BDET-20*.

Al Omari, M., Al-Hajj, M., Sabra, A., Hammami, N. (2019).Hybrid CNNs-LSTM Deep Analyzer for Arabic Opinion Mining.*2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS).*

Alayba, AM., Palade, V., England, M., Iqbal, R. (2018). A combined CNN and LSTM model for arabic sentiment analysis.I*n International Cross-Domain Conference for machine learning and knowledge Extraction,* pp.197-191.