International Journal of Electrical and Computer Engineering (IJECE)
Vol. 11, No. 3, June 2021, pp. 2327∼2334
ISSN: 2088-8708, DOI: 10.11591/ijece.v11i3.pp2327-2334

ATAR: Attention-based LSTM for Arabizi transliteration

Bashar Talafha[1], Analle Abuammar[2], Mahmoud Al-Ayyoub[3]
[1,3]Jordan University of Science and Technology, Jordan
[2]University of Southampton, UK

Article Info

Article history:

Received Apr 4, 2020
Revised Sep 30, 2020
Accepted Oct 9, 2020

Keywords:

Arabizi transliteration
Attention
Benchmark dataset
LSTM
Seq2seq

ABSTRACT

A non-standard romanization of Arabic script, known as Arbizi, is widely used in Arabic online and SMS/chat communities. However, since state-of-the-art tools and applications for Arabic NLP expects Arabic to be written in Arabic script, handling contents written in Arabizi requires a special attention either by building customized tools or by transliterating them into Arabic script. The latter approach is the more common one and this work presents two significant contributions in this direction. The first one is to collect and publicly release the first large-scale "Arabizi to Arabic script" parallel corpus focusing on the Jordanian dialect and consisting of more than 25 k pairs carefully created and inspected by native speakers to ensure highest quality. Second, we present ATAR, an ATtention-based LSTM model for ARabizi transliteration. Training and testing this model on our dataset yields impressive accuracy (79%) and BLEU score (88.49).

Corresponding Author:

Mahmoud Al-Ayyoub
Department of Computer Science
Jordan University of Science and Technology
Irbid, Jordan
Email: maalshbool@just.edu.jo

1. INTRODUCTION

As stated by many researchers [1–3], social media users express themselves in ways different from standard format. Social media content exhibit frequent use of informal vocabulary, non-standard abbreviation, typos, and many idiosyncrasies such as repeating letters for emphasis and writing out non-linguistic content like emojis and sound reactions [4–6]. For several reasons, these issues are more complicated for Arabic content. Examples of these reasons include the prevalent use of dialectal Arabic (DA) and its grave deviations from modern standard Arabic (MSA) [7]. Another reason is the common use of a non-standard romanized way of writing Arabic words known as *Arabizi*. There are many reasons for the widespread of Arabizi such as the lack of support for Arabic script on some devices/platforms, the existence of some difficulties in using Arabic script, the relative ease of code-switching between Arabizi and English or French compared with Arabic script. Even though Arabizi is not known to all social media users, it is common enough to warrant studies focusing solely on it [8–17].

For most state-of-the-art tools and applications for natural language processing (NLP) and information retrieval (IR) of Arabic text, the expected input is Arabic words written in Arabic script [18–20]. Therefore, there is an obvious need for a system to automatically transliterate content written in Arabizi into Arabic orthography [2]. Previous studies [2, 9, 21–30] presented tools and resources for this problem. However, to the best of our knowledge, very few of them [27–30] followed deep learning approaches such as Recurrent neural networks (RNN) and its extensions such as long short-term memory (LSTM) [31]. The others mostly follow character-level rule-based approaches.

In this work, we are addressing the problem of Arabizi transliteration by presenting ATAR, an ATtention-based encoder-decoder model for ARabizi transliteration. This novel neural network-based approach follows the celebrated attention-based encoder-decoder model of [32]. To evaluate ATAR, we present a "first of its kind" dataset consisting of 21.5 K words from the Jordanian dialect.

The rest of this paper is organized as: The following section gives a high-level view of the related work while section 3 presents our ATAR model and discusses its details. Section 4 discusses the dataset we create and section 5 presents our evaluation of the proposed model on the collected dataset. Finally, the paper is concluded in section 6.

2. RELATED WORK

Due to the importance of Arabizi-Arabic script transliteration problem, several companies, such as Google and Microsoft, have invested money and effort into developing tools for this problem. Examples of such tools include: Google Ta3reeb, in http://www.google.com/ta3reeb; Microsoft Maren, in https://www.microsoft.com/en-us/download/details.aspx?id=20530; Facebook's automatic translation services, in https://engineering.fb.com/ml-applications/expanding-automatic-machine-translation-to-more-languages/; Rosette Chat Translator, in https://www.basistech.com/text-analytics/rosette/chat-translator/; Yamli, in https://www.yamli.com/;. However, these tools are mostly closed-source, and very little is known about the approaches they follow or the resources they employ. On the other hand, the effort within the Arabic NLP research community to address the Arabizi-Arabic script transliteration problem has been rather shy. The existing resources are limited and are not publicly available and the proposed approaches do not follow the new and exciting approaches in the field of sequence learning [33].

Existing work on Arabizi transliteration, such as [2, 22–24, 34, 35], followed basic approaches that used character-to-character mappings in order to generate lattices of multiple alternative words. The approach proposed by [36] combines a rule-based model and a discriminative model based on conditional random fields (CRF) for transliterating Tunisian dialect Arabizi texts to standard Arabic. A further selection from these words is done using language models. As for the dataset they used, only that of [23] is reported to be publicly available [2], however, it is very small with only 2.2 K word pairs. It was used in the development of [24]'s system in addition to 6,300 Arabic-English proper name pairs from [37]. The reported accuracy of [24]'s system is 69.4% and it was later used by [2].

Another interesting effort in creating useful resources for the Arabizi transliteration problem is the work of Bies *et al.* [2]. The authors discussed how the linguistic data consortium (LDC) collected and annotated a huge parallel corpus of Arabizi content and its Arabic script counterpart as part of the DARPA broad operational language translation (BOLT) program (Phase 2). The corpus consisted of more than 408 K words and it mainly focused on the Egyptian dialect.

Few papers [27–30] discussed the use of deep learning for the problem of Arabic transliteration. In [27, 28], the authors claimed to use a standard RNN encoder-decoder model for transliterating sentences written in Algerian dialect, but they did not provide any details of the model. Moreover, the dataset they considered is rather small (1.3 k sentences). In a mode detailed work, Younes *et al.* [29] used a standard RNN encoder-decoder model for transliterating words in Tunisian dialect. Their dataset was relatively big with 45.6 k word pairs. In a follow-up work [30], the expanded their work and discussed how to adapt three well-known models in machine translation for the problem of transliterating Tunisian dialect. The first one was a CRF, while the second one was a Bidirectional RNN with Long short-term memory cells (BLSTM). As for the third one, it was a BLSTM with CRF decoder. The results show the superiority of the latter approach over the former two approaches.

Transliteration systems have been proposed for many languages other than Arabic. However, such systems are usually designed to transliterate between two closely related languages. Examples include the work of Musleh *et al.* [38] on transliterating Urdu to Hindi, the work of Nakov *et al.* [39] on transliterating Portuguese and Italian to look like Spanish and the work of Nakov *et al.* [40] on transliterating Macedonian to Bulgarian.

3. ATAR: ATTENTION-BASED LSTM FOR ARABIZI TRANSLITERATION

Over the past decade, deep learning approaches have made a ground-breaksaing impact on many fields such as NLP, image processing, computer vision [41–44]. A particularly interesting and challenging set of problems, known as sequence learning problems, has been heavily studied by deep learning researchers. A special kind of neural networks, known as recurrent neural networks (RNN), has been shown to perform very well for many sequence learning problems in natural language understanding (NLU) and natural language generation (NLG). However, RNN suffers from some issues like the vanishing gradient problem. To address this problem, Hochreiter and Schmidhuber [31] proposed to equip RNN with memory cells creating what they called LSTM networks.

For sequence-to-sequence problems (like the one we have at our hands), a general approach known as the encoder-decoder approach was found to be very successful. The approach is based on the idea of learning efficient representations of the input using an RNN (or LSTM) as an “encoder network” and using another RNN (or LSTM) as a “decoder network” to take this feature representation as input, process it to make its decision, and produce an output in https://www.quora.com/What-is-an-Encoder-Decoder-in-Deep-Learning. In the rest of this section, we present the details of our attention-based LSTM model for Arabizi transliteration, which we call ATAR.

3.1. Model architecture

our transliteration model is inspired by the attentional sequence-to-sequence (seq2seq) model proposed by [32], which is based on the encoder-decoder architecture as shown in Figure 1.

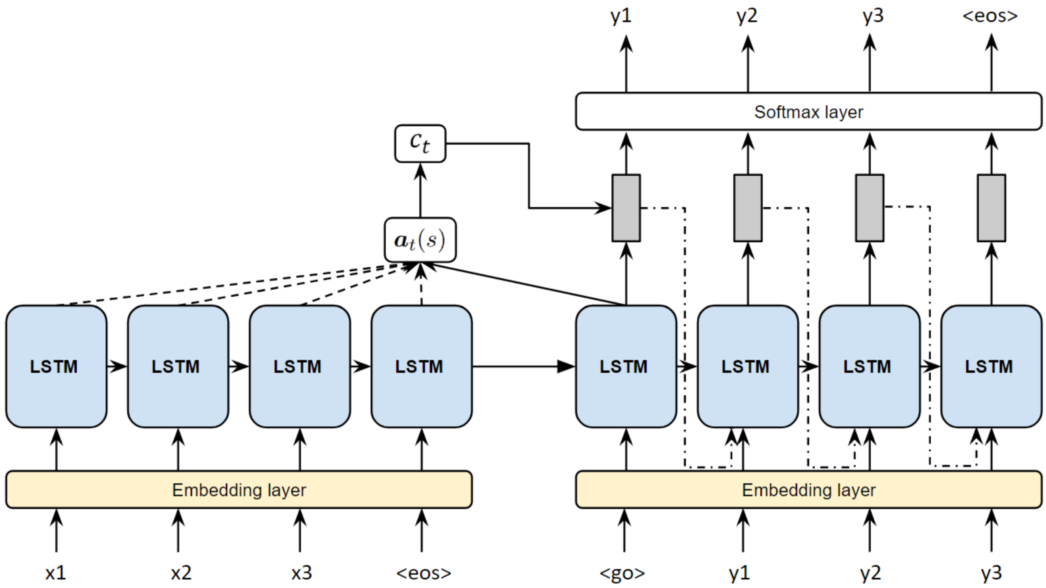

Figure 1. Illustration of the sequence-to-sequence architecture based on LSTM with the attention mechanism

The seq2seq architecture consists of an RNN encoder to learn representations of input sequence $X = \{x_1, x_2, \ldots, x_n\}$ of varying length and an RNN decoder, which reads the hidden representation produced by the encoder and generates output sequence $Y = \{y_1, y_2, \ldots, y_m\}$ of varying length. The model takes input from the embedding layer that maps a one-hot encoding vector of vocab size, which in our case is the number of letters consisting of different 47 letters in Arabizi and 36 in Arabic, as an input and generates a fixed size dense vector that represents the semantic features of the input letter. It is worth mentioning that there is no $<$UNK$>$ token in our case because each word (i.e., sequence) is a combination of limited predefined letters. In our architecture, each unit in the encoder and decoder is an LSTM cell which solves the problem of vanishing gradients with its memory cells [31].

Instead of relying on one thought vector from the encoder, many researchers [32, 45] proposed the encoder-decoder architecture with attention. The idea behind the attention mechanism is to link each time step of the decoder with the most "convenient" time step(s) of the encoder input sequence. This is done by utilizing the idea of global attentional model which takes all the hidden states of the encoder $\overline{h}_s$ and the current target state h_t into consideration to calculate the attention score. In this paper, the dot product function is used in order to perform the attention score calculation.

$$score(h_t^T, \overline{h}_s)$$

Following the previous step, the alignment vector a_{ts} is computed for each state by applying a softmax function to normalize all scores; therefore, a probability distribution based on the target state will be produced.

$$a_{ts} = \frac{\exp(score\left(h_t^T, \overline{h}_s\right))}{\sum_{\acute{s}} \exp(score\left(h_t^T, \overline{h}_{\acute{s}}\right))}$$

The decoder then computes a global context vector c_t as a weighted average, based on the alignment vector a_t over all the source states.

$$c_t = \sum_s a_{ts} \overline{h}_s$$

Therefore, the decoder will take the context vector as an additional input vector at the next time step s_t.

4. DATASET

We use Arabizi-Arabic script parallel words in order to perform our Arabizi transliteration experiments. Due to the lack of such available parallel data, we have crawled only Arabizi data written in the Jordanian dialect from different resources, such as Twitter, Facebook and ASK. These crawled words are regularly used on daily life basis. We were able to collect 21.5 K unique Arabizi words, which were then translated to the Jordanian dialect using only Arabic letters. A group of native speakers validated the parallel data by correcting any spelling mistakes, removing redundant letters and omitting any unneeded special characters.

One of the contributions of this work is to make this "first of its kind" dataset publicly available. In https://github.com/bashartalafha/Arabizi-Transliteration Table 1 shows samples of our parallel data. The average length of the collected words is about 5 letters per word, maximum word length of 12 letters and the minimum is 2 letters.

Table 1. Examples of our parallel corpus

Arabizi	Arabic Script
bl9odfa	بالصدفه
2b3tele	ابعتيلي
wr$eh	ورشه

It is worth mentioning that the same word in Arabizi could have different representations in the Jordanian dialect since not all people would write it in the same way but still they are all correct. Table 2 shows few such examples. This issue was faced by earlier work on Arabizi transliteration such as [2, 9] and it is discussed in details therein. As stated by these researchers, such things could penalize the model and give it lower score considering some transliterations are right but the reference is different.

Table 2. Examples with different representations

Arabizi (different representations)	Arabic script
keef, kaif	كيف
khaled, khalid, 5aled, 5alid	خالد
3arabi, 3araby, 3rabe	عربي
3alaykom, 3alaikom, 3laykoTyoum	عليكم
yemken, yomken, ymken	يمكن
almostashfa, elmostashfa, almosta$fa, almostashfah	المستشفى

5. EXPERIMENTS AND EVALUATION

To evaluate the performance of our proposed model, we implement it using TensorFlow (We select TensorFlow for its efficiency and ease of use. For a comparison of different deep learning frameworks, the interested readers are directed to [46]), and perform several experiments using our dataset. After shuffling and lowercasing the data, we use the first 80% of the dataset as the training set, the next 10% as the validation set and the remaining 10% as the testing. As for the evaluation metric, we use the two most common measures for the Arabizi transliteration task: accuracy and bilingual evaluation understudy (BLEU) [47]. Finally, to aid the reproducibility of our results, both the dataset and the model are made publicly available, in https://github.com/bashartalafha/Arabizi-Transliteration.

Using an attentional encoder-decoder sequence-to-sequence translation model, we have to worry about the many hyperparameters that can affect its performance. This issue is so important that complete studies have been dedicated for it such as [48], which reported the use of more than 250 K of GPU hours for experimentation. For our work, we use the work of Britz *et al.* [48] as well as Ruder's blog in http://ruder.io/deep-learning-nlp-best-practices/ and Brownlee's blog in https://machinelearningmastery.com/configure-encoder-decoder-model-neural-machine-translation/ to guide us in our experiments to search for the best values for the hyperparameters. The ones that give the best performance are listed in Table 3. For this configuration, the accuracy is 79% and the BLEU score is 88.49.

Table 3. The values of our model's hyperparameters that gives the best performance

Hyperparameter	Value
LEARNING_RATE	0.001
BATCH_SIZE	64
HIDDEN_NODES	256
NUMBER_OF_LAYERS	1
EMB_SIZE	50
EPOCHS	30
LOSS	"categorical_crossentropy"
OPT	"adam"
BI	"No"
DROPOUT	0.2

ATAR does achieve good results. However, it does have its limitation such as the lack of support for the various Arabic dialects. To address this, one might benefit from existing multi-dialect parallel datasets [49–54] or build new ones (perhaps, by benefiting from unsupervised approaches for dialect translation [7]). Another issue that can be addressed before adopting ATAR in real-life scenarios is trying to increase the model's accuracy. This can be done by either considering other sequence-to-sequence models, such as Facebook's convolutional sequence-to-sequence model [55] and Google's attention-only Transformer model [56] or by combining it with a neural diacrization model [57, 58].

6. CONCLUSION

In this paper, we addressed the Arabizi transliteration problem. This work has two significant contributions to this problem. The first one is to collect and publicly distribute the first large-scale Arabizi-Arabic script parallel corpus focusing on the Jordanian dialect and consisting of more than 25 k pairs carefully created and inspected by native speakers to ensure the highest quality. In the second contribution, we presented one of the first detailed and reproducible efforts to employ the celebrated attention-based seq2seq model for Arabizi transliteration. The presented model, which we called ATAR, performed very well in the experiments we conducted. It reached an impressive level with an accuracy of 79% and a BLEU score of 88.49. Future directions include experimenting with other sequence-to-sequence models, such as Facebook's convolutional sequence-to-sequence model and Google's attention-only Transformer model. We are also thinking of ways to expand our work to other Arabic dialects. Finally, we will explore the generation of more accurate MSA text from the transliteration by looking into combining our model with a neural diacrization model.

ACKNOWLEDGEMENT

The authors would like to thank the Deanship of Research at the Jordan University of Science and Technology for supporting this work (through Grant #20190180). The authors would also like to thank Nesreen Al-Qasem and Areen Bany Salim for their efforts in creating the dataset.

REFERENCES

[1] N. Y. Habash, "Introduction to arabic natural language processing," *Synthesis Lectures on Human Language Technologies*, vol. 3, no. 1, pp. 1–187, 2010.

[2] A. Bies, Z. Song, M. Maamouri, S. Grimes, H. Lee, J. Wright, S. Strassel, N. Habash, R. Eskander, and O. Rambow, "Transliteration of arabizi into arabic orthography: Developing a parallel annotated arabizi-arabic script sms/chat corpus," *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)*, 2014, pp. 93–103.

[3] W. A. Hussien, Y. M. Tashtoush, M. Al-Ayyoub, and M. N. Al-Kabi, "Are emoticons good enough to train emotion classifiers of arabic tweets?" *7th International Conference on Computer Science and Information Technology (CSIT)*, 2016, pp. 1–6.

[4] A. I. Alharbi and M. Lee, "Combining character and word embeddings for affect in arabic informal social media microblogs," *International Conference on Applications of Natural Language to Information Systems. Springer*, 2020, pp. 213–224.

[5] W. Hussien, M. Al-Ayyoub, Y. Tashtoush, and M. Al-Kabi, "On the use of emojis to train emotion classifiers," *arXiv preprint arXiv:1902.08906*, 2019.

[6] K. A. Kwaik, S. Chatzikyriakidis, S. Dobnik, M. Saad, and R. Johansson, "An arabic tweets sentiment analysis dataset (atsad) using distant supervision and self training," *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, 2020, pp. 1–8.

[7] W. Farhan, B. Talafha, A. Abuammar, R. Jaikat, M. Al-Ayyoub, A. B. Tarakji, and A. Toma, "Unsupervised dialectal neural machine translation," *Information Processing and Management*, vol. 57, no. 3, 2020.

[8] J. May, Y. Benjira, and A. Echihabi, "An arabizi-english social media statistical machine translation system," *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas*, 2014, pp. 329–341.

[9] M. van der Wees, A. Bisazza, and C. Monz, "A simple but effective approach to improve arabizi-to-english statistical machine translation," *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*, 2016, pp. 43–50.

[10] R. M. Duwairi, M. Alfaqeh, M. Wardat, and A. Alrabadi, "Sentiment analysis for arabizi text," *2016 7th International Conference on Information and Communication Systems (ICICS)*, 2016, pp. 127–132.

[11] A. M. Abd Al-Aziz, M. Gheith, and A. S. E. Ahmed, "Toward building arabizi sentiment lexicon based on orthographic variants identification," *The 2nd International Conference on Arabic Computational Linguistics (ACLing)*, 2016.

[12] I. Guellil, A. Adeel, F. Azouaou, F. Benali, A.-e. Hachani, and A. Hussain, "Arabizi sentiment analysis based on transliteration and automatic corpus annotation," *Proceedings of the 9th workshop on computational approaches to subjectivity, sentiment and social media Analysis*, 2018, pp. 335–341.

[13] I. Guellil, F. Azouaou, F. Benali, A. E. Hachani, and M. Mendoza, "The role of transliteration in the process of arabizi translation/sentiment analysis," *Recent Advances in NLP: The Case of Arabic Language, Springer*, 2020, pp. 101–128.

[14] T. Tobaili, M. Fernandez, H. Alani, S. Sharafeddine, H. Hajj, and G. Glavas, "Senzi: A sentiment analysis lexicon for the latinised arabic (arabizi)," *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, 2019, pp. 1203–1211.

[15] F. Aqlan, X. Fan, A. Alqwbani, and A. Al-Mansoub, "Arabic–chinese neural machine translation: Romanized arabic as subword unit for arabic-sourced translation," *IEEE Access*, vol. 7, pp. 133 122–133 135, 2019.

[16] E. Gugliotta and M. Dinarelli, "Tarc: Incrementally and semi-automatically collecting a tunisian arabish corpus," *arXiv preprint arXiv:2003.09520*, 2020.

[17] M. Alkhatib and K. Shaalan, "Boosting arabic named entity recognition transliteration with deep learn-

ing," *The Thirty-Third International Flairs Conference*, 2020.
[18] I. El Bazi and N. Laachfoubi, "Arabic named entity recognition using deep learning approach," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 3, pp. 2025-2032, 2019.
[19] H. G. Hassan, H. M. A. Bakr, and B. E. Ziedan, "A framework for arabic concept-level sentiment analysis using senticnet," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 5, pp. 4015-4022, 2018.
[20] M. A. Ahmed, R. A. Hasan, A. H. Ali, and M. A. Mohammed, "The classification of the modern arabic poetry using machine learning," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 17, no. 5, pp. 2667–2674, 2019.
[21] K. Shaalan, H. Bakr, and I. Ziedan, "Transferring egyptian colloquial dialect into modern standard arabic," *International Conference on Recent Advances in Natural Language Processing (RANLP–2007)*, Borovets, Bulgaria, 2007, pp. 525–529.
[22] A. Chalabi and H. Gerges, "Romanized arabic transliteration," *Proceedings of the Second Workshop on Advances in Text Input Methods*, 2012, pp. 89–96.
[23] K. Darwish, "Arabizi detection and conversion to arabic," *arXiv preprint arXiv:1306.6755*, 2013.
[24] M. Al-Badrashiny, R. Eskander, N. Habash, and O. Rambow, "Automatic transliteration of romanized dialectal arabic," *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*, 2014, pp. 30–38.
[25] R. Eskander, M. Al-Badrashiny, N. Habash, and O. Rambow, "Foreign words and the automatic processing of arabic social media text written in roman script," *Proceedings of The First Workshop on Computational Approaches to Code Switching*, 2014, pp. 1–12.
[26] N. Altrabsheh, M. El-Masri, and H. Mansour, "Proposed novel algorithm for transliterating arabic terms into arabizi," *Research in Computer Science*, 2017.
[27] I. Guellil, F. Azouaou, M. Abbas, and S. Fatiha, "Arabizi transliteration of algerian arabic dialect into modern standard arabic," *Social MT 2017/First workshop on social media and user generated content machine translation*, 2017.
[28] I. Guellil, F. Azouaou, and M. Abbas, "Neural vs statistical translation of algerian arabic dialect writ- ten with arabizi and arabic letter," *The 31st Pacific Asia Conference on Language, Information and Computation PACLIC*, vol. 31, 2017, p. 2017.
[29] J. Younes, E. Souissi, H. Achour, and A. Ferchichi, "A sequence-to-sequence based approach for the double transliteration of tunisian dialect," *Procedia computer science*, vol. 142, pp. 238–245, 2018.
[30] J. Younes, H. Achour, E. Souissi, and A. Ferchichi, "Romanized tunisian dialect transliteration using sequence labelling techniques," *Journal of King Saud University-Computer and Information Sciences*, 2020.
[31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[32] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025*, 2015.
[33] M. Al-Ayyoub, A. Nuseir, K. Alsmearat, Y. Jararweh, and B. Gupta, "Deep learning for arabic nlp: A survey," *Journal of computational science*, vol. 26, pp. 522–531, 2018.
[34] G. Lancioni, E. Gugliotta, and V. Pettinari, "Lahajat: A rule-based converter of standard arabic lexical databases into spoken arabic forms," *2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)*, 2016, pp. 395–399.
[35] I. Guellil, F. Azouaou, F. Benali, A.-E. Hachani, and H. Saadane, "Hybrid approach for transliteration of algerian arabizi: a primary study," *arXiv preprint arXiv:1808.03437*, 2018.
[36] A. Masmoudi, M. E. Khmekhem, M. Khrouf, and L. H. Belguith, "Transliteration of arabizi into arabic script for tunisian dialect," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 19, no. 2, Nov. 2019, doi: 10.1145/3364319.
[37] T. Buckwalter, "Buckwalter arabic morphological analyzer version 2.0," *Web Download*, 2004.
[38] A. Musleh, N. Durrani, I. Temnikova, P. Nakov, S. Vogel, and O. Alsaad, "Enabling medical translation for low-resource languages," *International Conference on Intelligent Text Processing and Computational Linguistics. Springer*, 2016, pp. 3–16.
[39] P. Nakov and H. T. Ng, "Improving statistical machine translation for a resource-poor language using related resource-rich languages," *Journal of Artificial Intelligence Research*, vol. 44, pp. 179–222, 2012.

[40] P. Nakov and J. Tiedemann, "Combining word-level and character-level models for machine translation between closely-related languages," *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2. Association for Computational Linguistics*, 2012, pp. 301–305.

[41] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[42] I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning," *MIT press*, 2016.

[43] J. Patterson and A. Gibson, "Deep learning: A practitioner's approach," *" O'Reilly Media, Inc."*, 2017.

[44] G. Al-Bdour, R. Al-Qurran, M. Al-Ayyoub, and A. Shatnawi, "A detailed comparative study of open source deep learning frameworks," *arXiv preprint arXiv:1903.00102*, 2019.

[45] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[46] G. Al-Bdour, R. Al-Qurran, M. Al-Ayyoub, and A. Shatnawi, "Benchmarking open source deep learning frameworks," *Submitted to the International Journal of Electrical and Computer Engineering (IJECE)*, 2020.

[47] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," *Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics*, 2002, pp. 311–318.

[48] D. Britz, A. Goldie, M.-T. Luong, and Q. Le, "Massive exploration of neural machine translation architectures," *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 1442–1451.

[49] H. Bouamor, S. Hassan, and N. Habash, "The madar shared task on arabic fine-grained dialect identification," *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, 2019, pp. 199–207.

[50] B. Talafha, A. Fadel, M. Al-Ayyoub, Y. Jararweh, A.-S. Mohammad, and P. Juola, "Team just at the madar shared task on arabic fine-grained dialect identification," *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, 2019, pp. 285–289.

[51] B. Talafha, W. Farhan, A. Altakrouri, and H. Al-Natsheh, "Mawdoo3 ai at madar shared task: Arabic tweet dialect identification," *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, 2019, pp. 239–243.

[52] A. Ragab, H. Seelawi, M. Samir, A. Mattar, H. Al-Bataineh, M. Zaghloul, A. Mustafa, B. Talafha, A. A. Freihat, and H. Al-Natsheh, "Mawdoo3 ai at madar shared task: Arabic fine-grained dialect identification with ensemble learning," *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, 2019, pp. 244–248.

[53] C. Zhang, H. Bouamor, M. Abdul-Mageed, and N. Habash, "The shared task on nuanced rabic dialect identification (nadi)," *Proceedings of the Fifth Arabic Natural Language Processing Workshop*, 2020.

[54] B. Talafha, M. Ali, M. E. Za'ter, H. Seelawi, I. Tuffaha, M. Samir, W. Farhan, and H. T. Al-Natsheh, "Multi-dialect arabic bert for country-level dialect identification," *arXiv preprint arXiv:2007.05612*, 2020.

[55] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional sequence to sequence learning," *Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org*, 2017, pp. 1243–1252.

[56] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.

[57] A. Fadel, I. Tuffaha, M. Al-Ayyoub et al., "Arabic text diacritization using deep neural networks," *2019 2nd International Conference on Computer Applications and Information Security (ICCAIS)*, 2019, pp. 1–7.

[58] A. Fadel, I. Tuffaha, B. Al-Jawarneh, and M. Al-Ayyoub, "Neural arabic text diacritization: State of the art results and a novel approach for machine translation," *arXiv preprint arXiv:1911.03531*, 2019.