A systematic review on sequence-to-sequence learning with neural network and its models

Hana Yousuf[1], Michael Lahzi[2], Said A. Salloum[3], Khaled Shaalan[4]
[1,2,3,4]Faculty of Engineering and IT, the British University in Dubai, United Arab Emirates
[3]Research Institute of Sciences and Engineering, University of Sharjah, United Arab Emirates

Article Info	ABSTRACT

We develop a precise writing survey on sequence-to-sequence learning with neural network and its models. The primary aim of this report is to enhance the knowledge of the sequence-to-sequence neural network and to locate the best way to deal with executing it. Three models are mostly used in sequence-to-sequence neural network applications, namely: recurrent neural networks (RNN), connectionist temporal classification (CTC), and attention model. The evidence we adopted in conducting this survey included utilizing the examination inquiries or research questions to determine keywords, which were used to search for bits of peer-reviewed papers, articles, or books at scholastic directories. Through introductory hunts, 790 papers, and scholarly works were found, and with the assistance of choice criteria and PRISMA methodology, the number of papers reviewed decreased to 16. Every one of the 16 articles was categorized by their contribution to each examination question, and they were broken down. At last, the examination papers experienced a quality appraisal where the subsequent range was from 83.3% to 100%. The proposed systematic review enabled us to collect, evaluate, analyze, and explore different approaches of implementing sequence-to-sequence neural network models and pointed out the most common use in machine learning. We followed a methodology that shows the potential of applying these models to real-world applications.

Corresponding Author:

Said A. Salloum
Research Institute of Sciences and Engineering
University of Sharjah
Academic city Road, P. O.Box 27272 Sharjah, United Arab Emirates
Tel: +971 6 5585000, Fax: +971 6 5585099
Email: ssalloum@sharjah.ac.ae

1. INTRODUCTION

Machine learning (ML) is a logical investigation of calculations and accurate models within computational frameworks that act without utilizing clear guidelines, depending on examples and surmising. It is viewed as a subset of computerized reasoning [1-4]. Performing ML includes making a model, which is prepared on some preparation information and afterward can process extra information to make forecasts [5-7]. Different kinds of models have been utilized and investigated for ML frameworks. These models include neural networks, decision trees, regression analysis and have a massive application that includes speech and object recognition [8-15]. The scope of this paper is focused on neural networks and their subsets, particularly neural networks and sequence-to-sequence learning. Neural systems or connectionist frameworks are registering frameworks dubiously motivated by the organic neural systems that establish creature cerebrums. Such frameworks "learn" to perform assignments by thinking about models, for the most part,

without being modified with task-explicit guidelines. Their applications are massive, and they have been developed to perform many difficult tasks, mainly including prediction and classification [16].

Sequence-to-Sequence learning is a part of ML and a method of neural networks that is mostly utilized in language processing models [17-23]. It can be implemented with recurrent neural networks (RNNs) using encoder-decoder based machine interpretation that maps an input sequence to a yield of a succession of output sequence with a tag and consideration esteem. The idea is to exploit two RNN that will cooperate with a unique token and attempt to anticipate the following state arrangement from the past succession. Sequence-to-Sequence models can also be implemented through what is known as connectionist temporal classification (CTC) and attention-based models. The sequence-to-sequence model was initially created by Google for machine translation and was introduced to train the model with a single command. The model was also made to be easily reproducible and extendable such that the code files were organized in a measured manner and that simple to expand upon and recreate.

The extent of this paper incorporates neural systems and their subsets, especially recurrent neural networks (RNN), connectionist temporal classification (CTC), and attention-based models to determine which is the best-suited approach to implement with sequence-to-sequence learning. This paper aims to explore the working of sequence-to-sequence learning and their different applications, which is reflected in the research questions of the paper.

This paper will conduct a systematic literature review on sequence-to-sequence neural network through exploring different academic research directories to look for peer-reviewed content. More on the method of conducting this systematic literature review is discussed in Section 3. The structure of this paper is as: Section 2 will talk about and dissect the accessible writing about RNNs, CTC, and attention-based models learning; it will explain the working of each network and discuss their strengths and weaknesses. Section 3 will study the examination methodology used to lead the systematic literature review, which incorporates the detailing and formulation of the research questions. Section 4 will utilize the academic writings to conduct a quality assessment, respond to the questions addressed, and lastly, section 5 will have a short conclusion to the paper.

2.   LITERATURE REVIEW

Neural networks is an established model of machine learning and has had extensive research done on it over the years; Sequence-to-sequence neural network is a new learning technique [9, 24, 25]. Despite this, there is still quite a substantial some of research done on both models and techniques, which will be expanded on in this section.

2.1. Background

Neural networks are inspired by biological neural networks systems that comprise animals' brains; they are designed to develop, progress, and solve complex problems that require a high level of comprehension to perform. There are many types of neural networks that are found to perform extremely well with such difficult tasks such as speech recognition and machine translation; one of such networks are the recurrent neural networks (RNN). The working principle behind RNNs is essentially constructed around neural network models that incorporate an encoder-decoder framework that can be used and trained end-to-end to map input sequences into output target sequences [26]. A wide-ranging definition of sequence-to-sequence models can be said to "refers to the broader class of models that include all models that map one sequence to another" [27]. Thus, by comparing it to the definition of [26], it is clear to see the relation between the two definitions.

Connectionist temporal classifications (CTC) is a kind of neural system yield related to scoring capacity, for preparing intermittent neural systems to handle grouping issues where the timing is variable. It may be utilized for assignments like online penmanship recognition or perceiving phonemes in discourse sound. CTC was presented in 2006 and alluded to the yields and scoring and is autonomous of the hidden neural system structure [28]. Historically, CTC has been used for the classification of unsegmented sequences with RNNs, such as cases of handwriting or speech recognition. RNNs on their own were not sufficient for the task as their standard neural system target capacities are characterized independently for each point in the preparation arrangement; basically, RNNs must be prepared to make a progression of autonomous mark orders.

In order to remove this dependency and enable RNNs to perform this task, the network had to decode the system outputs as a likelihood appropriation over all possible mark successions, adapted on a given input grouping. Given this dispersion, a target capacity can be determined that straightforwardly expands the probabilities of the right marking. Since the target work is differentiable, the system would then

be able to be prepared with standard backpropagation through time [29]. Thus, using this concept, therefore, use of RNNs in this way was known as CTC.

The attention mechanism is a type of neural network that allows the decoder aspect of the network to focus on certain parts of the sequence while the output is generated. Attention-based models help remove any dependencies on variable-length inputs without compressing them into fixed vectors by using variable-length memory where then, the model is free to use this memory in a truly adaptable way to create the output succession. In addition, various pieces of memory can be obtained at multiple yield time steps. These models are well-persuaded in light of the fact that data is lost by compacting long factor length groupings into a fixed-size vector, and choosing the pressure is an extra assignment to comprehend [30].

2.2. Sequence-to-sequence models

There are several approaches used to implement sequence-to-sequence algorithm models. The most common models are the connectionist temporal classification (CTC), RNNs, and attention-based model.

2.2.1. Connectionist temporal classification

The CTC algorithm proposed by [28]. This algorithm is a method of preparing start to finish models without a requirement of casing level arrangement of the objective names for a preparation articulation. CTC defines the probability of the output condition, estimated to use recurrent neural networks, simply known as encoders [31]. In addition, CTC uses for sequence-to-sequence method to help to address any issues related to the length of the output labels when it is shorter than the length of the input sequence since "CTC introduces a special blank label and allows for repetition of labels to force the output and input sequences to have the same length. CTC outputs are usually dominated by blank symbols" [32]. This gives CTC a major advantage when using sequence-to-sequence models in many applications such as translation.

2.2.2. Recurrent neural networks

The idea behind sequence-to-sequence models using the RNN approach utilizes two RNN that will cooperate with a unique token and attempt to anticipate the following state arrangement from the past succession. The RNN transducer differs on the encoder usage from the CTC alignment model by different repeat lease expectation arrangement over the output sequences. Instinctively, the encoder can be thought of as an acoustic model, while the expectation arranges practically equivalent to a language model. The expectation arrange gets as info and processes an output vector, subject to the whole sequence of labels [31].

2.2.3. Attention model

It is a consideration-based model contains an encoder organize, as in the RNN transducer model. In any case, in contrast to the RNN transducer, in which the encoder and the expectation arrange are displayed autonomously and consolidated in the joint system, a consideration based model uses a solitary decoder to deliver an appropriation over the marks molded on the full grouping of past forecasts and the acoustics [31]. The decoder network consists of several recurrent layers. The attention aspect of the model puts a higher weight on certain layers to produce an output using the end-to-end sequence method.

3. METHODOLOGY

The principle of research methodology is developed based on a systematic literature review. The paper follows the systematic review methodology illustrated by [33] to direct the deliberate literature audit. The reason for selecting the systematic review for sequence-to-sequence neural network is that no systematic review focuses on sequence-to-sequence neural network usage, limitations, and applications. Moreover, this methodology enabled us to collect, evaluate, analyze, and explore different approaches to implementing sequence-to-sequence neural network models and find the most common use in machine learning. The initial step to this method is to figure out the research hypothesis of our paper.

3.1. Research hypothesis

The research hypothesis developed for the paper were as:
1. What are the different applications of sequence-to-sequence neural network models?
2. How has this model been implemented and developed?
3. What are the advantages and limitations of implementing sequence-to-sequence models?
4. What is the best model to approach sequence-to-sequence implementation?
5. What are the countries that contributed to the development and implementation of sequence-to-sequence?
Table 1 shows the research hypothesis, including their motivation.

Table 1. Research motivation and hypothesis

RQ#	Research Hypothesis	Motivation
RQ1	What are the different applications of sequence-to-sequence models?	This question helps gain a broader understanding of all the applications of sequence-to-sequence models to relieve a lot of ambiguity surrounding the definition. This will later help identify the best possible approach to implement the model based on the best advantages it has on each application.
RQ2	How has this model been implemented and developed?	This question allows a deeper understanding of the development of the model. Moreover, It will highlight its shortcomings to identify the additional work needs to be done on the field and what is the best and least problematic approach to be used with common machine learning applications like speech recognition and translation.
RQ3	What are the advantages and limitations of implementing sequence-to-sequence models?	This question addresses the significance of the model and how the installation has helped in relieving some of the problems that were involved in the field of machine learning prior to its development. Moreover, this helps to highlight some challenges that are still not solved and new challenges that developed with the new model.
RQ4	What is the best model to approach sequence-to-sequence implementation?	This question will highlight the advantages and disadvantages of every approach. It will help choose the best approach for the purposes of suggesting a standard to use when implementing this model in common machine learning applications.
RQ5	What are the countries that contributed to the development and implementation of sequence-to-sequence?	This question will highlight the most countries contributed to the implementation and development Sequence-to-Sequence

3.2. Research strategy

The principle of research strategy in this paper is to conduct a careful checking on the subsequent database, peer-reviewed journals, and periodicals. Most of the papers utilized from arXiv, Google Scholar, Science Direct, IEEE Xplore, and Springer complete journals. When the most significant research registries were chosen, important keywords retrieved from the research hypothesis were utilized in search through pertinent titles and modified works of papers and articles. A Boolean technique adapted to string together the important search terms to locate the most significant peer-evaluated items for this paper. Specific perspectives were viewed when utilizing the search terms, such as equivalent words or plurals of terms. Table 2 shows the search terms used in various databases and the outcomes it accomplished. Overall, there were 871 papers appeared through the initial searching stage. From those, 86 papers appeared to be duplicated, and five papers were found through citations of found academic papers, which brought the total number of papers to 790. These papers were refined. The final pieces of literature used in this paper were screening according to the selection criteria and the preferred reporting items for systematic reviews and meta-analyses (PRISMA) Statement [34]. Figure 1 shows the PRISMA flowchart.

Table 2. Boolean technique search results

Database	Search String	Number of results
Google Scholar	("Connectionist temporal classifications") and ("sequence-to-sequence") and ("RNN") and ("CTC")	56
Science Direct	(("Attention model" or "attention mechanism") and ("sequence-to-Sequence"))	68
IEEE Xplore	"Sequence-to-sequence" and "neural network" and "challenges" or "limitation"	284
Springer Complete Journals	("Sequence-to-sequence") and ("recurrent neural networks" or "connectionist temporal classifications" or "attention model")	463

3.3. Inclusion and exclusion criteria

Inclusion and exclusion criteria were chosen criteria to measure the most formal writing for the extent of this paper. In view of these criteria, the papers which followed the research perspective criteria were included for the extent of research. Table 3 shows the inclusion and exclusion criteria.

Table 3. Inclusion and exclusion criteria

Inclusion criteria	Excision criteria
Must be peer-reviewed journal papers, articles, or books	Articles are written in another language
Should include the development and implementation of the sequence-to-sequence model and neural networks	Papers focus on specialized parts or working the sequence-to-sequence model and neural network
Must be written in English	Discuses as a subtopic in Sequence-to-Sequence, RNN, CTC, and attention models.

Academic papers that failed to meet the mentioned criteria were excluded from research. The initial screening was done on all the extracted 790 records in order to narrow the result. These records were

compared against pre-selection criteria that included refining the search fields to engineering and computer science, limiting the results to the no older than ten years. This resulted in the evaluated numbers to decrease to 413 papers, which were assessed by the selection criteria mentioned. Figure 1 represents the process for both the pre-selection and the selection criteria. This process resulted in 16 papers to be included in the systematic review in sequence-to-sequence neural network.
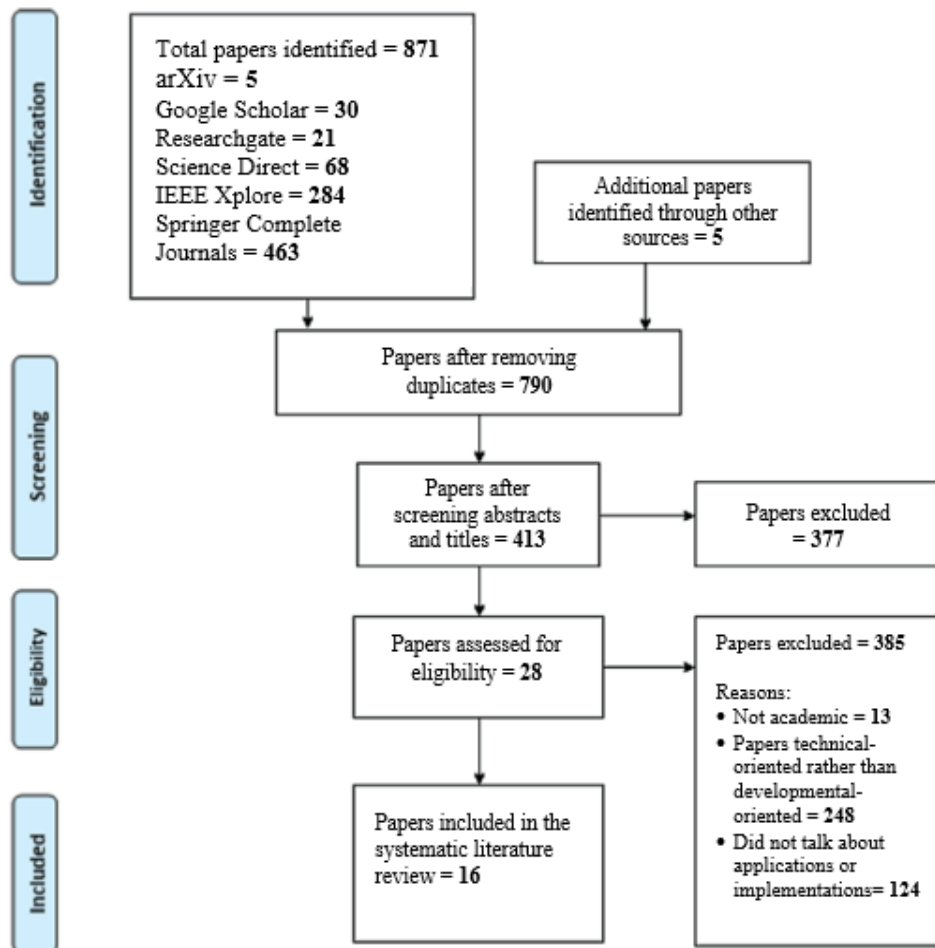
Figure 1. PRISMA flowchart

3.4. Quality assessment

Inclusion quality assessment is one of the most essential and critical parts of any systematic review [35-39]. The assessment quality assurance checklist for our systematic review consists of 6 questions for the 16 chosen papers, as shown in Table 4. The scoring of this process is done based on the work of [40] as: a 'Yes' to the question of the quality assessment is indicated by a 1, a 'No' was indicated by a 0, and a 'Partially' was indicated by a 0.5. As seen by the results in Table 5, all the chosen papers have passed the quality assessment.

Table 4. Quality assurance questions

Ques.	Quality Assurance Question
1	Are the research aims specified clearly?
2	Is the information presented clear and concise?
3	Does the study provide enough explanation of its methodology?
4	Do the findings of the study add to the understanding of sequence-to-sequence models?
5	Are the conclusions clearly identified?
6	Are the conclusions logical and concise with the flow of the paper?

Table 5. Quality assessment results

Study	Q1	Q2	Q3	Q4	Q5	Q6	Total	Percentage
S1	1	1	0.5	1	1	1	5.5	91.67%
S2	1	1	1	1	0.5	1	5.5	91.67%
S3	1	1	1	0	1	1	5	83.33%
S4	1	1	1	0.5	1	0.5	5	83.33%
S5	1	1	1	0.5	1	1	5.5	91.67%
S6	1	1	0.5	0.5	1	1	5	83.33%
S7	1	1	1	1	1	1	6	100%
S8	1	1	0.5	1	1	0.5	5	83.33%
S9	1	1	1	0	1	1	5	83.33%
S10	1	1	1	0	1	1	5	83.33%
S11	1	1	1	1	1	1	6	100%
S12	1	1	1	1	0.5	1	5.5	91.67%
S13	1	1	0.5	1	0.5	1	5.5	91.67%
S14	1	0.5	1	0.5	1	1	5.5	91.67%
S15	1	1	0	1	1	1	5	83.33%
S16	1	1	1	1	1	0.5	5.5	91.67%

4. RESULTS AND DISCUSSION

Utilizing the research procedure outlined in the previous section, the research questions that were tended to each paper are classified and analyzed based on their contribution to each question.

4.1. Classifications and analysis of studies

By studying all the 16 research papers included in the systematic review, a classification system was comprised based on their contribution to answering the research questions. Markings were made if the main focus of the paper was related to a particular category. Most papers, for example, talked in brief about the different applications of sequence-to-sequence (seq2seq) models. However, only studies that were mainly focused on a certain application or discussed several applications in depth were categorized as 'Applications of seq2seq'. The classification results can be seen in Table 6. Additionally, each study was analyzed in detail, and the results of this detailed study are outlined in Table 7 (see in Appendix). Figure 2 shows the publication distributions country wise. As the results show, there has been an increase in interest in this topic over the last two decades, as indicated by the increase in a number of publications since 1990. Most of this research is focused on the US, as mentioned.
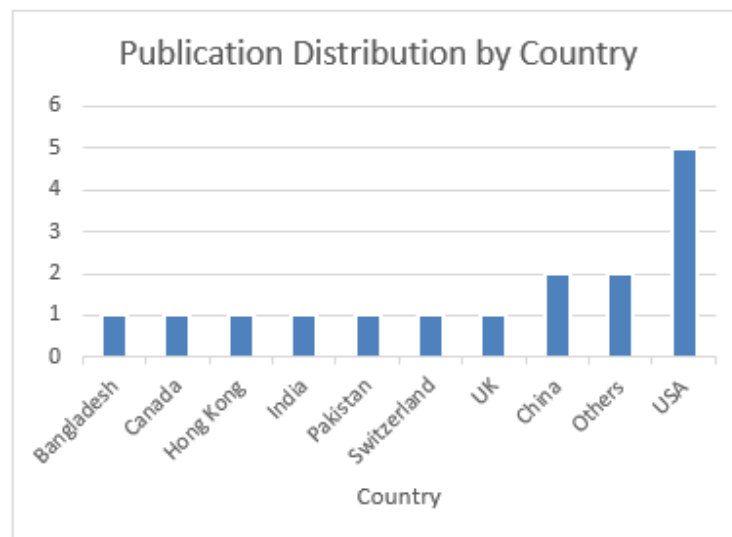

Figure 2. Publication distribution country wise

4.2. Quality assessment results

Using the questions outlined in Table 4, a score was given to each of the 16 studies used in this paper. The maximum score that a paper can have is 6. The results of each paper are presented in Table 6.

Table 6. Classification of literature review

Study	Source	Applications of seq2seq	Seq2seq via RNN	Seq2seq via CTC	Seq2seq via attention models	Challenges of seq2seq
S1	[41]	x	x			
S2	[32]			x	x	
S3	[28]		x	x		x
S4	[42]	x		x		
S5	[43]	x	x			
S6	[27]	x				x
S7	[16]	x	x			x
S8	[31]		x	x	x	
S9	[26]		x	x		
S10	[44]			x		
S11	[45]	x			x	
S12	[8]	x	x			
S13	[46]	x	x			
S14	[29]	x				x
S15	[47]	x	x			
S16	[48]	x	x			

4.3. Answers to research questions
a.   RQ1. What are the different applications of the sequence-to-sequence neural network model?

 As seen from **Error! Reference source not found.**, 11, or 68.75%, of the studies were relevant to the applications of sequence-to-sequence neural network models, indicating not only the relevance of this question but also its widespread interest in the field. The general consensus was that sequence-to-sequence models were best utilized for speech recognition and general linguistics, as suggested by [16, 31, 48]. In addition, sequence-to-sequence models can be used for video to text conversion [46] and handling large vocabularies, optimizing translation performance, and multi-lingual learning [27].

b.   RQ2. How has this model been implemented and developed?

 All 16 of the papers used in the systematic review mentioned different approaches to implementing and developing the sequence-to-sequence neural network models. For example: R. Prabhavalkar *et al.* [31] summarises three methods of implementation, which include: RNNs, Connectionist Temporal Classifications (CTC), and Attention models.

c.   RQ3. What are the advantages and limitations of implementing sequence-to-sequence models?

 I. Sutskever *et al.* [8] and Y. H. Chan *et al.* [41] talked mostly about implementations using RNNs and discussed many advantages and disadvantages of applying this model, while in [28, 42] discussed the implementation through CTC in details with the limitations in implementation. Similarly, in [32, 45] discussed the limitations of the attention model.

d.   RQ4. What is the best model to approach sequence-to-sequence implementation?

 The majority of the papers (62.5%) talk about RNNs and their implementation, limitations, and endorsement. Notably, the work of [31] compared the three different approaches and found the most promising approach to be "the RNN transducer, attention-based models, and a novel RNN transducer augmented with attention."

5.   CONCLUSION

 In conclusion, the paper aimed to conduct a systematic review on the topic of the sequence-to-sequence neural network and its models. The main aim of the review to gain insight into the sequence-to-sequence neural network models and to find the best approach to implement it. Three such approaches were found: through recurrent neural networks, connectionist temporal classifications (CTC), and attention models. The research question derived for the literature review were encompassing the applications of sequence-to-sequence models, their advantages and disadvantages, as well as the best implementation approach for them. The procedure done to conduct this systematic literature review included using the research questions to derive keywords that were then used to look at the subsequent database, peer-reviewed journals, and periodicals. Most of papers utilized from arXiv, Google Scholar, Science Direct, IEEE Xplore and Springer complete journals. Through initial searches, 790 papers and academic works were found, and with the help of selection criteria and PRISMA procedure, the number of papers reviewed in this paper was reduced to 16. Each of the 16 papers was categorized by their contribution to each research question, and they were analyzed. Finally, the research papers underwent a quality assessment where the resulting range was from 83.3% to 100%.

APPENDIX

Table 7. Analysis of the literature review

Study	Source	Purpose	Country	Database	Findings
S1	[41]	The authors suggested and experimented with an alternative approach to embedded emotional information at the encoder stage of sequence to sequence based emotional generation.	Hong Kong	IEEE Xplore	Different methods were tested on emotional encoding information for sequence to sequence generation response and evaluated the result, which was found to have a positive effect on a sentence level.
S2	[32]	The paper proposed a new method used in neural speech recognition by aligning the attention modeling within the CTC framework	Canada	IEEE Xplore	The proposed method boosted the end-to-end acoustic-to-word CTC model and achieved better WER than the traditional context-dependent phoneme CTC model decoded with a very large-sized language model
S3	[28]	The paper presented a novel method for training RNNs to label unsegmented sequences. "An experiment on the TIMIT speech corpus demonstrates its advantages over both a baseline HMM and a hybrid HMM-RNN".	Switzerland	ACM Digital Libraries	The method is derived from probabilistic principles and fits the framework of the neural network classifier. It allows the network to be trained directly for sequence labeling by removing the requirement of pre-segmented data.
S4	[42]	The authors developed a technique to base the RNN-T system towards a specific keyword.	USA	IEEE Xplore	The paper developed a streaming keyword by using (RNN-T) model to predict either phonemes or graphemes as sub-word, which allows detecting random expressions.
S5	[43]	The authors proposed an artificial Bangla text generator with LSTM and model it to validate the accuracy of the text generators.	Bangladesh	Science Direct	The paper worked with RNN structures and LTSM networks to be able to fulfill their purpose. They were able to "model for achieving multi-task sequence to sequence text generation and multi-way translation like Bengali articles, caption generation"
S6	[27]	The author introduced 'neural machine translation' or 'neural sequence-to-sequence models'	USA	arXiv	The paper covered the basics of neural machine translation and sequence to sequence models. It covered several application such as Handling large vocabularies, Optimizing translation performance, and Multi-lingual learning.
S7	[16]	This paper carries out a comprehensive review of articles that involve a comparative study of feed forward neural networks and statistical techniques used for prediction and classification problems in various areas of applications. Tabular presentations highlighting the important features of these articles are also provided	India	Science Direct	"Neural networks are being used in areas of prediction and classification, the areas where statistical methods have traditionally been used". "One of the important advantages of neural networks cited in the literature is that it can automatically approximate any nonlinear mathematical function".
S8	[31]	The authors conducted a detailed evaluation of sequence to sequence models which are used in the task of speech recognition such as connectionist temporal classification (CTC), the recurrent neural network (RNN) transducer, an attention-based model, and a model which augments the RNN transducer with an attention mechanism.	USA	Google Scholar	"[they] compared a number of sequence-to-sequence modeling approaches on an LVCSR task. In experimental evaluations, we find that the RNN transducer, attention-based models and a novel RNN transducer augmented with attention are comparable in performance to a strong state-of-the-art baseline on a dictation test set, even when evaluated without the use of an external pronunciation or language model"
S9	[26]	"We introduce an encoder-decoder recurrent neural network model called recurrent neural aligner (RNA) that can be used for the sequence to sequence mapping tasks. Like connectionist temporal classification (CTC) models, RNA defines a probability distribution over target label sequences, including blank labels corresponding to each time step in the input".	USA	Research Gate	"We presented the RNA model, which is a recurrent neural network model in the encoder-decoder framework. We applied it to end-to-end speech recognition and showed initial experimental results on YouTube transcription and mobile dictation tasks. We found that the RNA grapheme model can not perform as well as CTC CD phone models for mobile dictation task where the vocabulary size is larger, and the training data is relatively smaller."

Table 7. Analysis of the literature review (*Continue*)

Study	Source	Purpose	Country	Database	Findings
S10	[44]	"We propose SAN-CTC, a deep, fully self-attentional network for CTC, and show it is tractable and competitive for end-to-end speech recognition. We motivate the architecture for speech, evaluate the position and down-sampling approaches, and explore how to label alphabets (character, phoneme, subword) affect attention heads and performance".	UK	IEEE Xplore	"SAN-CTC trains quickly and outperforms existing CTC models and most encoder-decoder models, with character error rates (CERs) of 4.7% in 1 day on WSJ eval92 and 2.8% in 1 week on LibriSpeech test-clean, with a fixed architecture and one GPU. Similar improvements hold for WERs after LM decoding.
S11	[45]	The work is focused on using sequence to sequence the attention model by the Google brain team to generate the abstract of research papers. Moreover, temporal attention mechanism has been used in replacement to the global attention to cater to the problem of repetitive words.	Pakistan	IEEE Xplore	The result indicates that the temporal attention model is a useful method for generating summaries. Moreover, results indicate that with the increase in dataset size, the accuracy of the results also increases.
S12	[8]	The paper uses a multilayered long short-term memory (LSTM) to map the input sequence to a vector of fixed dimensionality. Then, another deep LSTM to decode the target sequence from the vector."	Canada	Google Scholar	The main result is that on an English to French translation task from the WMT-14 dataset, the translations produced by the LSTM achieve a BLEU score of 34.8 on the entire test set, where the LSTM's BLEU score was penalized on out-of-vocabulary words. Additionally, the LSTM did not have difficulty in long sentences.
S13	[46]	"We propose a novel end-to-end sequence-to-sequence model to generate captions for videos. For this, we exploit recurrent neural net- works, specifically LSTMs, which have demonstrated state-of-the-art performance in image caption generation."	USA	IEEE Xplore	[This paper] constructed descriptions using a sequence to sequence model, where frames are first to read sequentially, and then words are generated sequentially. This allows us to handle variable-length input and output while simultaneously modeling temporal structure. Our model achieves state-of-the-art performance on the MSVD dataset, and outperforms related work on two large and challenging movie-description datasets."
S14	[29]	"Basic backpropagation, which is a simple method now being widely used in areas like pattern recognition and fault diagnosis, is reviewed. The basic equations for backpropagation through time, and applications to areas like pattern recognition involving dynamic systems, systems identification, and control are discussed."	USA	IEEE Xplore	This paper presents the key equations of backpropagation, as applied to neural networks of varying degrees of complexity. It has also discussed other papers which elaborate on the extensions of this method to more general applications and some of the tradeoffs involved
S15	[47]	"This paper examines the most popular DNNs approaches: LSTM, Encoder-Decoder network and Memory network in sequence prediction field to handle the software sequence learning and prediction task."	China	IEEE Xplore	"Our results demonstrate that attention mechanism does not fit all seq2seq problems, especially in a weak mapping relationship. And additional information can benefit sequence prediction in neural networks."
S16	[48]	The paper presented a character level sequence to sequence the learning method. The author set in an RNN into an encoder-decoder framework and generate the character-level sequence in representation as an input	China	IEEE Xplore	Experimental results by the authors confirmed that the proposed approach achieved performance close to conventional word and phrase-based translation systems. The proposed methods allows to "reads quantized characters into the translation system, instead of using a predefined vocabulary with a limited number of words"

REFERENCES

[1]   S. A. Salloum, M. Alshurideh, A. Elnagar, and K. Shaalan, "Machine Learning and Deep Learning Techniques for Cybersecurity: A Review," *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, vol. 1153, 2020, pp. 50-57.

[2]   S. A. Salloum, M. Alshurideh, A. Elnagar, and K. Shaalan, "Mining in Educational Data: Review and Future Directions," *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, vol. 1153, 2020, pp. 92-102.

[3]   S. A. Salloum, R. Khan, and K. Shaalan, "A Survey of Semantic Analysis Approaches," *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, vol. 1153, 2020, pp. 61-70.

[4]   A. Y. Zainal, H. Yousuf, and S. A. Salloum, "Dimensions of Agility Capabilities Organizational Competitiveness in Sustaining," *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, vol. 1153, 2020, pp. 762-772.

[5]   H. AlGhanem, M. Shanaa, S. Salloum, and K. Shaalan, "The Role of KM in Enhancing AI Algorithms and Systems," *Advances is Science, Technology and Engineering System Journal*, vol. 5, no. 4, pp. 388-396, 2020.

[6]   H. Yousuf, A. Y. Zainal, M. Alshurideh, and S. A. Salloum, "Artificial Intelligence Models in Power System Analysis," in *Artificial Intelligence for Sustainable Development: Theory, Practice and Future Applications*, vol. 912, pp. 231-242, 2020.

[7]   M. AlShamsi, S. A. Salloum, M. Alshurideh, and S. Abdallah, "Artificial Intelligence and Blockchain for Transparency in Governance," in *Artificial Intelligence for Sustainable Development: Theory, Practice and Future Applications*, vol. 912, pp. 219-230, 2020.

[8]   I. Sutskever, O. Vinyals, and Q.V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, pp. 3104-3112, 2014.

[9]   T. -S. Nguyen, S. Stüker, J. Niehues, and A. Waibel, "Improving sequence-to-sequence speech recognition training with on-the-fly data augmentation," *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 7689-7693.

[10] I. Arel, J. B. Looks, A. Ziaei, and M. Lefkowitz, "Reconciliation between simulator and speech recognition output using sequence-to-sequence mapping," Google Patents, 2020.

[11] M. Blaauw and J. Bonada, "Sequence-to-Sequence Singing Synthesis Using the Feed-Forward Transformer," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 7229-7233.

[12] S. A. Salloum, M. Al-Emran, A. A. Monem, and K. Shaalan, "A survey of text mining in social media: facebook and twitter perspectives," *Advances is Science, Technology and Engineering System Journal*, vol. 2, no. 1, pp. 127-133, 2017.

[13] S. A. Salloum, A. Q. AlHamad, M. Al-Emran, and K. Shaalan, "A survey of Arabic text mining," *Intelligent Natural Language Processing: Trends and Applications*, vol. 740, pp. 417-431, 2017.

[14] S. A. Salloum, M. Al-Emran, A. A. Monem, and K. Shaalan, "Using text mining techniques for extracting information from research articles," *Intelligent Natural Language Processing: Trends and Applications*, vol. 740, pp. 373-397, 2017.

[15] S. A. Salloum, M. Al-Emran, S. Abdallah, and K. Shaalan, "Analyzing the Arab Gulf Newspapers Using Text Mining Techniques," *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics*, vol. 639, 2017, pp. 396-405.

[16] M. Paliwal and U. A. Kumar, "Neural networks and statistical techniques: A review of applications," *Expert Systems with Applications*, vol. 36, no. 1, pp. 2-17, 2009.

[17] P. Serai, A. Stiff, and E. Fosler-Lussier, "End to End Speech Recognition Error Prediction with Sequence to Sequence Learning," *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 6339-6343.

[18] K. Palasundram, N. M. Sharef, K. A. Kasmiran, and A. Azman, "Enhancements to the Sequence-to-Sequence-Based Natural Answer Generation Models," *IEEE Access*, vol. 8, pp. 45738-45752, 2020.

[19] I. Saito, K. Nishida, K. Nishida, and J. Tomita, "Abstractive Summarization with Combination of Pre-trained Sequence-to-Sequence and Saliency Models," *arXiv Prepr. arXiv2003.13028*, 2020.

[20] M. Surdeanu, "A Gentle Introduction to Deep Learning for Natural Language Processing," *google Scholar*, pp. 1-63, 2020.

[21] D. King, A. D. Sims, and M. Elsner, "Interpreting sequence-to-sequence models for Russian inflectional morphology," *Proceedings of the Society for Computation in Linguistics*, vol. 3, pp. 402-411, 2020.

[22] Y. Yan et al., "ProphetNet: Predicting Future N-gram for Sequence-to-Sequence Pre-training," *arXiv Prepr. arXiv2001.04063*, 2020.

[23] A. Sun, J. Wang, N. Cheng, H. Peng, Z. Zeng, and J. Xiao, "GraphTTS: graph-to-sequence modelling in neural text-to-speech," *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 6719-6723.

[24] R. Lokeshkumar, K. Jayakumar, V. Prem, and M. S. Nonghuloo, "Analyses and Modeling of Deep Learning Neural Networks for Sequence-to-Sequence Translation," *International Journal of Advanced Science and Technology*, vol. 29, no. 5, pp. 3152-3159, 2020.

[25] C. Chiu, N. Jaitly, I. Sutskever, and Y. Luo, "Recurrent neural networks for online sequence generation," US Patent, 2020.

[26] H. Sak, M. Shannon, K. Rao, and F. Beaufays, "Recurrent Neural Aligner: An Encoder-Decoder Neural Network Model for Sequence to Sequence Mapping," *Interspeech*, vol. 8, pp. 1298-1302, 2017.

[27] G. Neubig, "Neural machine translation and sequence-to-sequence models: A tutorial," *arXiv Prepr. arXiv1703.01619*, 2017.

[28] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 369-376.

[29] P. J. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550-1560, 1990.

[30] A. K. Tiwari, G. Ramakrishna, L. K. Sharma, and S. K. Kashyap, "Academic performance prediction algorithm based on fuzzy data mining," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 8, no. 1, pp. 26-32, 2019.

[31] R. Prabhavalkar, K. Rao, T. N. Sainath, B. Li, L. Johnson, and N. Jaitly, "A Comparison of Sequence-to-Sequence Models for Speech Recognition.," *Interspeech*, pp. 939-943, 2017.

[32] A. Das, J. Li, R. Zhao, and Y. Gong, "Advancing connectionist temporal classification with attention modeling," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB, 2018, pp. 4769-4773.

[33] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering-a systematic literature review," *Information and Software Technology*, vol. 51, no. 1, pp. 7-15, 2009.

[34] D. Moher, A. Liberati, J. Tetzlaff, and D.G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," *PLoS Medicine*, vol. 8, no. 7716, pp. 336-341, 2009.

[35] S. A. S. Salloum and K. Shaalan, "Investigating students' acceptance of E-learning system in Higher Educational Environments in the UAE: Applying the Extended Technology Acceptance Model (TAM)," *The British University in Dubai*, pp. 1-102, 2018.

[36] S. A. Salloum, A. Qasim Mohammad Alhamad, M. Al-Emran, A. Abdel Monem, and K. Shaalan, "Exploring students' acceptance of e-learning through the development of a comprehensive technology acceptance model," *IEEE Access*, vol. 7, pp. 128445-128462, 2019.

[37] S. F. S. Alhashmi, S. A. Salloum, and S. Abdallah, "Critical Success Factors for Implementing Artificial Intelligence (AI) Projects in Dubai Government United Arab Emirates (UAE) Health Sector: Applying the Extended Technology Acceptance Model (TAM)," *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2019,* vol. 1058, 2020, pp. 393-405.

[38] S. F. S. Alhashmi, S. A. Salloum, and C. Mhamdi, "Implementing Artificial Intelligence in the United Arab Emirates Healthcare Sector: An Extended Technology Acceptance Model," *International Journal of Information Technology and Language Studies*, vol. 3, no. 3, pp. 27-42, 2019.

[39] S. F. S. Alhashmi, M. Alshurideh, B. Al Kurdi, and S. A. Salloum, "A Systematic Review of the Factors Affecting the Artificial Intelligence Implementation in the Health Care Sector," *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, vol. 1153, 2020, pp. 37-49.

[40] M. Al-Emran, V. Mezhuyev, A. Kamaludin, and K. Shaalan, "The impact of knowledge management processes on information systems: A systematic review," *International Journal of Information Management*, vol. 43, pp. 173-187, 2018.

[41] Y. H. Chan and A. K. F. Lui, "Encoding emotional information for sequence-to-sequence response generation," *2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)*, Chengdu, 2018, pp. 113-116.

[42] Y. He, R. Prabhavalkar, K. Rao, W. Li, A. Bakhtin, and I. McGraw, "Streaming small-footprint keyword spotting using sequence-to-sequence models," *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, Okinawa, 2017, pp. 474-481.

[43] M. S. Islam, S. S. S. Mousumi, S. Abujar, and S. A. Hossain, "Sequence-to-sequence Bangla sentence generation with LSTM Recurrent Neural Networks," *Procedia Computer Science*, vol. 152, pp. 51-58, 2019.

[44] J. Salazar, K. Kirchhoff, and Z. Huang, "Self-attention networks for connectionist temporal classification in speech recognition," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, 2019, pp. 7115-7119.

[45] T. Siddiqui and J. A. Shamsi, "Generating Abstractive Summaries Using Sequence to Sequence Attention Model," *2018 International Conference on Frontiers of Information Technology (FIT)*, Islamabad, Pakistan, 2018, pp. 212-217.

[46] S. Venugopalan, M. Rohrbach, J. Donahue, R. Mooney, T. Darrell, and K. Saenko, "Sequence to sequence-video to text," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015, pp. 4534-4542.

[47] Q. Yang, Z. He, F. Ge, and Y. Zhang, "Sequence-to-sequence prediction of personal computer software by recurrent neural network," *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, 2017, pp. 934-940.

[48] H. Zhang, J. Li, Y. Ji, and H. Yue, "A character-level sequence-to-sequence method for subtitle learning," *2016 IEEE 14th International Conference on Industrial Informatics (INDIN)*, Poitiers, 2016, pp. 780-783.

BIOGRAPHIES OF AUTHORS

Hana Y. Zainal had a master's degree in Electrical Engineering from Rochester Institute of Technology .and a Bachelor Degree in Electrical and Computer Engineering from United Arab Emirates University. She is currently a Computer Science PhD candidate in British University in Dubai. The area of interest are Natural Language Processing, Sequence-to-Sequence, CryptDB, Word2Vec, Victorization, Deep Learning and Secure Encryption. Hana is working in Dubai and Water Authority (DEWA) in Planning Division as Deputy Senior Manager.

Michael Lahzi Gaid is a PhD student in CS in the British University in Dubai (BUiD). He received his Masters of CS from The Arab Academy for Science, Technology & Maritime Transport (AAST), Cairo Egypt in 2019. His particular interests in computer security and data confidentiality. He works as head developer in Setup Sys Company, developing ERP-System. Also he works as IT Manager in Ceramica Verdi, Aba Al Dahab Trading Company, Ceramica Glamour, Mica-Vit Factory, Egyptian Canadian Co. For Manufacturing Carton - Golden Packs, and Mountee Verde Co. For Manufacturing Oils. Michael has a successful business carrier, he has developed custom software in many properties all over the world e.g. Vatican, Vodafone Egypt, Top-Business, Moevenpick Hotels, Rotana Hotels, Ceramica Alfa).

Said A. Salloum had graduated from The British University in Dubai with a distinction with MSc in Informatics (Knowledge and Data Management). He got his Bachelor's degree in Computer Science from Yarmouk University. Currently, He is working at the University of Sharjah "Research Institute of Sciences and Engineering (RISE)" as a researcher on different research areas in Computer Science such as data analysis, machine learning, knowledge management, and Arabic Language Processing. Salloum is an Oracle expert since 2013, along with various recognized international certificates that are issued by Oracle. You can contact. Said Salloum via emails at: ssalloum@sharjah.ac.ae or salloum78@live.com

Khaled Shaalan is a full professor of Computer Science/Artificial Intelligence at the British University in Dubai (BUiD), UAE. He is an Honorary Fellow at the School of Informatics, University of Edinburgh (UoE), UK. Over the last two decades, Prof Khaled has been contributing to a wide range of research topics in AI, Arabic NLP, Knowledge management, health informatics, and educational technology. Prof Khaled has published 200+ referred publications. Prof Khaled's research work is cited extensively worldwide and the impact of his research using GoogleScholar's H-index metric is 35+. Prof Khaled has been actively and extensively supporting the local and international academic community. He acts as the chair of international Conferences, journals & books editor, keynote speaker, external member of promotions committees, among others.