Technical Sciences, 2024, 27, 113–125

UNIWERSYTET WARMIŃSKO-MAZURSKI W OLSZTYNIE

Yearbook peer-reviewed scientific journal
ISSN 1505-4675
e-ISSN 2083-4527
TECHNICAL SCIENCES
Homepage: https://czasopisma.uwm.edu.pl/index.php/ts/

DOI: https://doi.org/10.31648/ts.10228

NEURAL NETWORK ARCHITECTURES IN DATA SEQUENCE ANALYSIS: DIAGNOSIS OF REINKE'S OEDEMA AND LARYNGEAL POLYPS

Jarosław Szkoła

ORCID: 0000-0002-6043-3313
Institute of Computer Sciences
University of Rzeszow

Received 15 May 2024, accepted 05 June 2024, available online 06 June 2024.

K e y w o r d s: clinical decision support, recurrent neural networks, Reinke's oedema, laryngeal polyps, Bézier curves, speech signal analysis, temporal patterns, classification.

A b s t r a c t

In this review paper, we focus our attention on presenting selected neural network architectures dedicated to the analysis of sequential data, in particular to support the diagnosis of Reinke's oedema and laryngeal polyps. The research discussed here is located in the area of clinical computer decision support systems (CDS) based on the use of artificial recurrent neural networks (RNNs) for speech signal analysis. RNNs are able to predict time series due to their memory and local recurrent connections. In the experimental part, Elman-Jordan artificial neural networks are used, whose characteristics are speed and accuracy in pattern learning allowing real-time decision-making. In the review presented here, one important theme is the use of Bezier curves for preprocessing the speech signal, leading to data reduction and noise elimination. Elman-Jordan networks significantly speed up the learning process and show high classification accuracy in laryngopathy diagnosis.

Correspondence: Jarosław Szkoła, Instytut Informatyki, Zakład Metod Przybliżonych, Uniwersytet Rzeszowski, ul. Pigonia 1, 35-310 Rzeszów, e-mail: jszkola@ur.edu.pl.

Introduction

In this review, we will focus on the different artificial neural network architectures used to analyze data sequences in the context of the diagnosis of Reinke's oedema and laryngeal polyps. In our discussion, we will refer to several key works that use recurrent neural networks (RNNs) and their modifications in the analysis of speech signals of patients with laryngopathy. In studies (SZKOLA et al. 2010a, 2010b, 2011a), the authors showed how combining an Elman neural network (ELMAN 1993) with a Jordan neural network (JORDAN 1986) can speed up the learning process, which is crucial in real-time diagnoses. The combination of these networks retains the ability to distinguish between normal and pathological states, while improving learning efficiency. Work SZKOŁA (2011b) has shown that preprocessing the speech signal using Bézier curve approximation can significantly reduce the amount of data needed for learning and remove noise from the original signal, improving the accuracy of health state classification. Despite some limitations, this method indicates directions for further research. It is worth briefly mentioning the currently popular techniques based on convolutional networks in our review. In the literature on the application of deep learning to medical diagnosis, review papers such as BAKATOR and RADOSAV (2018) indicate that CNN architectures are widely used in the analysis of medical images and acoustic signals. In particular, the application of deep learning in the analysis of speech signals for the diagnosis of laryngopathy shows promise results (ZAIDI et al. 2011, SFAYYIH et al. 2023a, 2023b, TANVEER et al. 2023, MEHRISH 2023). The feature that distinguishes our research from convolutional networks is that the size of the input data does not need to be fixed and the data strings can be aggregated directly.

The next parts of our review work will focus on: Computer-assisted clinical decision support for laryngopathy using RNNs (SZKOLA et al. 2010a, 2010b) – see section *Computer aided clinical decision support for laryngopathy using recurrent neural networks* and *Recurrent neural networks in computerized clinical decision support for laryngopathy: An experimental study*. Improving the learning ability of recurrent neural networks in speech signal analysis (SZKOŁA et al. 2011a) – see section *Improving the learning capacity of recurrent neural networks – experiments on the speech signals of patients with laryngopathies*. Bézier curve approximation in the laryngopathy classification process (SZKOŁA et al. 2011b) – see section *A Bezier curve approximation of the speech signal in the classification process of laryngopathies*. Our results and conclusions will be useful in the further development of AI-based diagnostic tools that can support real-time diagnosis of laryngeal diseases – see section *Methodology – non-invasive detection of laryngeal disease* and *Conclusions*.

Methodology – non-invasive detection of laryngeal disease

In the work, we review our selected techniques for the non-invasive detection of laryngeal disease. Originally, statistical analysis was used for this purpose in medical centers and processing of the input data as speech signals into Fourier spectral form. We present our proprietary solutions, produced in the research group, which are focused, among other things, on the possibility of using them in real time.

Computer aided clinical decision support for laryngopathy using recurrent neural networks

In the first approaches, a statistical method was used to deviate the laryngeal frequency based on FFT, the SDA parameter was determined. The problem was that this worked differently. The main issue was the use of Fourier analysis, which loses information about time in the context of spectral distribution. And from direct observation of the samples, it is clear that disturbances occur at random points along the length of the recorded sample. Therefore, the idea of using a different technique that can respond to specific perturbations in selected parts of the waveform emerged in subsequent studies see (SZKOLA et al. 2010a). This paper presents a new approach to the analysis of speech samples, in terms of anomaly detection, relative to control group samples, in the form of an Elman neural network. The most important conclusions reached after applying the neural network are as follows: Spectral analysis does not fulfil its role as an effective tool for detecting anomalies; In speech samples that may be indicative of a disease entity. This is due to several reasons: Patients' speech samples are strongly individual, meaning that no standard frequency distribution can be identified for classifying speech samples into healthy or pathological categories; Depending on the severity of the disease, the age of the patient, the gender, the disturbance of the speech samples can show significant differences, it is usually the case that some parts of the sample look healthy, comparable to the control group, and some places show significant deviations from the pattern of the whole signal. In the case of frequency analysis, it is not possible to tell if it is just one small disturbance or if it repeats cyclically, the whole spectral waveform is a kind of averaging of the frequency components of the whole waveform. This is a big problem, because a patient who will show a single instance of a disturbance, and a patient with more such instances, the sample analysis of both patients will show great similarity, which interferes with the diagnostic process. The use of Elman's recurrent neural network allows the analysis of information that was not available in statistical methods with spectral analysis. The network learns each patient's phoneme articulation and is not sensitive to individual characteristics. The aim of Elman's recurrent neural network is to identify anomalies in a sequence of repeated samples containing phonemes.

Due to the nature of the data, which contains many repeated value sequences, the learning of the network is efficient and the model produced for a given patient is able to indicate where the anomaly occurs, as well as its relevance to the sample as a whole. The neural network used has a slightly different architecture to typical RNN, or LSTM, GRU recurrent neural networks, whose sole purpose is to correctly indicate the next value, based on a finite (usually short) preceding sequence. Typical recurrent neural networks have no internal memory, in which to store the necessary data, their prediction is based on taking into account previously calculated weights and the current input value. LSTM, GRU neural networks use very simple tags that can be propagated over long distances, but this too is not sufficient to create an optimal data string model. In the case of Elman neural networks, the context layer exhibits features of signal feature clustering, with some similarity to the mechanism as we observe in Kohonen neural networks, which allows it to generalize the model much better while still being able to store long data strings. In each learning step, this internal representation, together with the hidden layer, processes new information, creating increasingly subtle relationships between neurons. Elman recurrent neural network used.

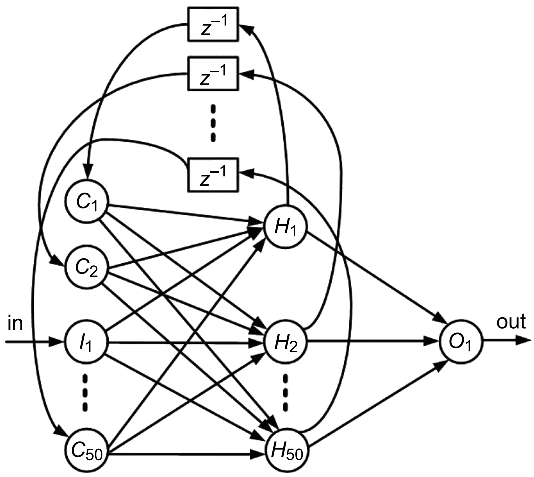

Fig. 1. Elman network

Based on the research carried out, it can be concluded that the use of a recurrent neural network, for the detection of anomalies, allows for more predictable results, and, at the same time, the speed of the network is satisfactory. Not without significance is also the possibility of easier adaptation of the network to new requirements, by changing the number of neurons in the contextual and hidden layers, as well as adjusting the number of learning steps.

Experimental sample. The table shows the results of the comparison of the two techniques, for the same speech samples for the control and pathological group (laryngeal polyp).

Table 1

Results of the comparison of the two techniques, for the same speech samples for the control and pathological group (laryngeal polyp)

Women from the control group			Women with laryngeal polyp		
ID	SDA – error	NN – error	ID	SDA – error	NN – error
WCG1	0.311	0.00059	WP1	0.084	0.0024
WCG2	0.159	0.00032	WP2	0.138	0.0141
WCG3	0.167	0.00068	WP3	0.147	0.0039
WCG4	0.012	0.0003	WP4	0.84	0.00082
WCG5	0.139	0.00072	WP5	0.2	0.0007
WCG6	0.205	0.00041	WP6	0.333	0.0021
WCG7	0.118	0.00073	WP7	0.169	0.0022
WCG8	0.127	0.00065	WP8	0.219	0.0024

Source: based on data from SZKOLA et al. (2010a).

We can clearly see that for both the control and pathological group, the SDA coefficient can return high (WCG1, WP4, WP6) as well as low values (WCG4, WP1) . In the case of the results obtained from the neural network, we see a much better separation of results, the control group containing an error practically an order of magnitude smaller than the pathological group. When using statistical methods, we cannot be sure which class the sample belongs to.

Recurrent neural networks in computerized clinical decision support for laryngopathy: an experimental study

The paper SZKOŁA, PANCERZ and WARCHOŁ (2010) introduces significant improvements to the Elman neural network, which was first used to analyze speech samples for anomaly detection, as outlined in the paper SZKOLA, PANCERZ and WARCHOL (2010a). Elman's and Jordan's neural networks, well known until now, were used disjointly for different tasks, in the main this was due to the nature of the data, and the ability to teach the chosen network. In the figure below, we have presented two classical neural networks, the Elman neural network and the Jordan neural network.

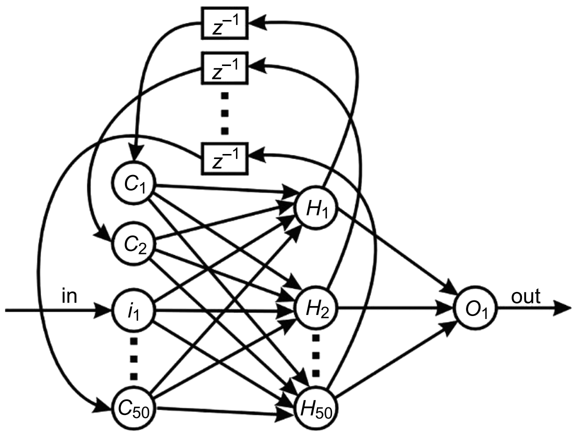

Fig. 2. Elman neural network

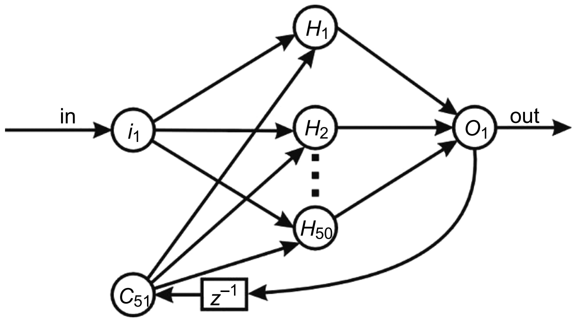

Fig. 3. Jordan neural network

A characteristic feature of the Jordan neural network is feedback from the context layer to the output layer. In the case of the Elman network, feedback to the hidden layer is used. In this way, the Jordan network has a smaller contextual memory capacity and therefore a lower capacity for data generalization. An important advantage of the Jordan network is that it takes as input of the next iteration, the value directly available at the output of the previous learning iteration of the neural network, with the context value thus set. Dealing with the Elman neural network, the context layer is a kind of abstraction resulting from the connection to the hidden layer. Therefore It was therefore decided to exploit the advantages of both networks, and in this way the first Elman-Jordan neural network presented in the article SZKOŁA, PANCERZ and WARCHOŁ (2010) was created (The article includes an erroneous drawing of the extended Elman--Jordan neural network, instead of the first version of this network).

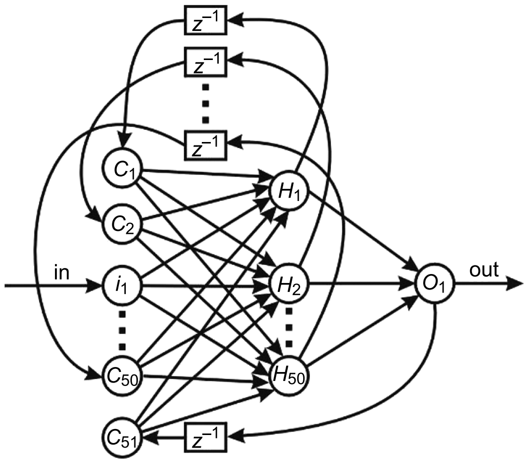

Fig. 4. Elman-Jordan network

Experimental sample. For the new architecture of the recurrent neural network, tests were carried out, in terms of detecting anomalies in speech samples, on a slightly extended group of patients. From the available pool of speech samples, 100 samples were selected for each class. 90% of the samples were used as the training set, and 10% of the samples were used for model testing. The learning process continued until an accuracy of less than 0.001 was achieved. If the network exceeds 2,300 epochs without achieving an accuracy of less than 0.001, the learning process is stopped.

In the following tables, n_Er is denoted as mean squared error, and n_Ep is denoted as the average number of epochs needed to achieve an error less than or equal to 0.001. The following results were obtained from the simulations.

Table 2

Results for test datasets for Elman network

Control group			Pathological group		
ID	n_Er	n_Ep	ID	n_Er	n_Ep
wCG1	0.0163	109	wP1	0.1832	110
wCG2	0.0196	114	wP2	0.3386	241
wCG3	0.0228	95	wP3	0.0630	284
wCG4	0.0196	87	wP4	0.0314	118
wCG5	0.0400	91	wP5	0.0486	122
wCG6	0.0237	91	wP6	0.2238	79
wCG7	0.0223	90	wP7	0.2238	791
wCG8	0.0163	125	wP8	0.0421	96
wCG9	0.0190	157	wP9	0.0459	121
wCG10	0.0189	93	wP10	0.1440	90

Table 3

Results for test datasets for Elman-Jordan network

Control group			Pathological group		
ID	n_Er	n_Ep	ID	n_Er	n_Ep
wCG1	0.0138	68	wP1	0.1480	53
wCG2	0.0193	69	wP2	0.2184	65
wCG3	0.0219	68	wP3	0.0551	100
wCG4	0.0182	61	wP4	0.0306	69
wCG5	0.0376	65	wP5	0.0416	54
wCG6	0.0235	62	wP6	0.2237	92
wCG7	0.0283	67	wP7	0.0844	66
wCG8	0.0162	67	wP8	0.0364	87
wCG9	0.0193	78	wP9	0.0374	65
wCG10	0.0206	70	wP10	0.1214	71

Based on the results obtained, it can be concluded that the combination of the Elman and Jordan neural networks into a single network has resulted in the new recurrent neural network showing even better performance than the usual Elman neural network. It is worth noting that the new network shows better class separation, for healthy samples and pathological samples, as well as a reduction in the learning time of the Elman-Jordan neural network. The behavior of the network is as predicted, as the modifications made mean that the new network has more information than the classic Elman neural network, in addition to the data from the hidden layer, we still have an exact copy of the output value, which is taken into account in the learning process at each subsequent epoch. More information allows faster convergence of the model for the same data as the Elman neural network.

Improving the learning capacity of recurrent neural networks – experiments on the speech signals of patients with laryngopathies

In the paper Szkoła, Pancerz and Warchoł (2011a) further modifications to the Elman-Jordan neural network are presented. Compared to the first version of the Elman-Jordan neural network, a modification has been introduced with an additional output layer feedback.

The additional feedback introduced improves the linearity of the learning process, and at the same time speeds up the whole process. The figure shows the difference in the learning process of a typical Elman neural network, an extended Elman-Jordan neural network.

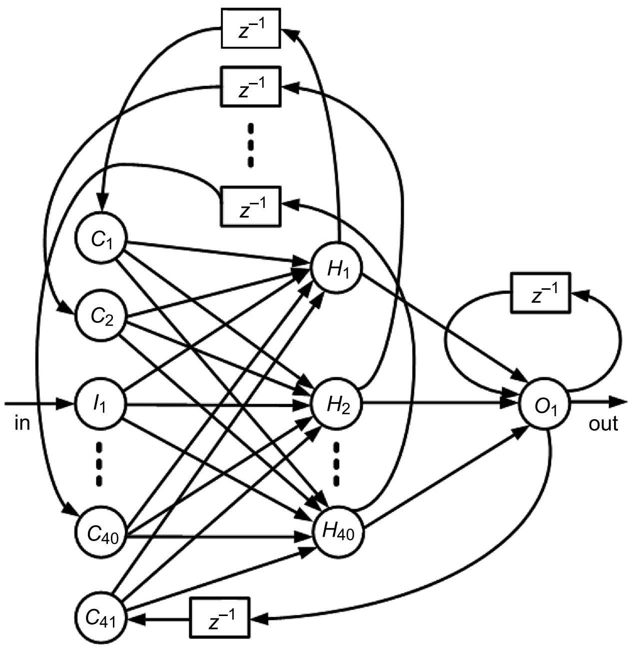

Fig. 5. Extended Elman-Jordan network

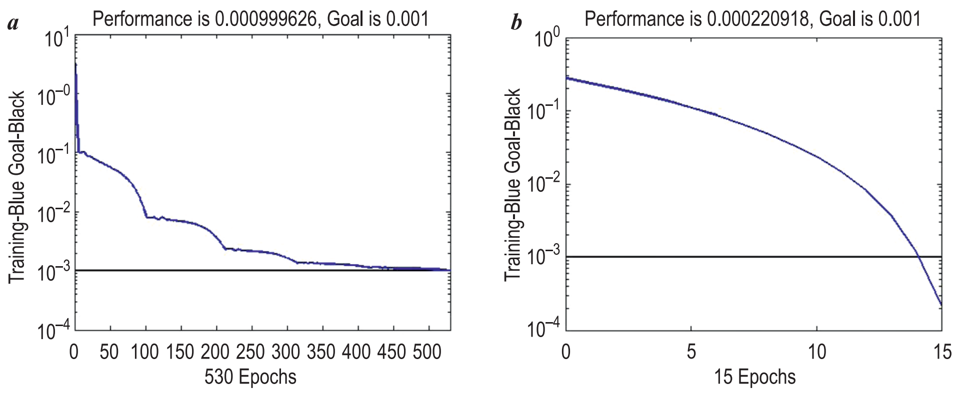

Fig. 6. Comparison of the learning curve for the Elman network (*a*) and the Extended Elman-Jordan network (*b*)

Relative to the Elman neural network, a significant acceleration of the learning process was achieved, with no deterioration in the quality of anomaly detection, relative to the Elman neural network or the first version of Elman--Jordan. As indicated in the title of the article SZKOŁA, PANCERZ and WARCHOŁ (2011a), the aim of the paper was to present an improvement of the Elaman-Jordan neural network presented earlier and the objective was achieved. Comparative trials were carried out on the datasets that were used to study the E-J neural network. The new network was named ‘modified Elman-Jordan’. The result achieved is significant, as it allows, with the same resources, to carry out analyses on larger sets, or for sets of similar size, it allows a reduction in simulation time.

The paper SZKOŁA, PANCERZ and WARCHOŁ (2010) presents a new approach to data analysis, by using two of the same Elman-Jordan neural networks, with the second network obtaining data derived from the data fed to the first neural network.

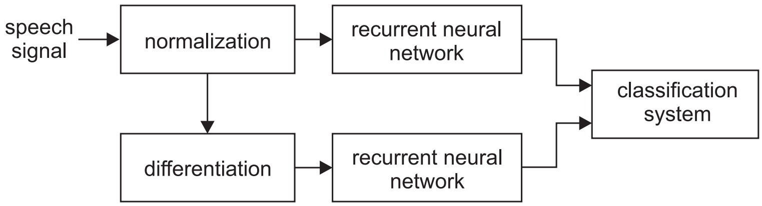

Fig. 7. The Blok diagram of the process of the experiment

Ultimately, the decision is made by a rule-based expert system, based on the values obtained from both neural networks. This system can be called a hybrid system, as it contains elements of preprocessing, two neural networks and a decision-making system based on approximate methods. The study showed that such a system could be effective in distinguishing healthy from diseased samples, but failed to achieve satisfactory results against two pathological groups, i.e. laryngeal polyp and Reinke’s sign. One significant problem, as often arises in sensory data acquisition, is the level of noise and interference. In the case of speech samples, for which even subtle changes in signal amplitude and frequency can be significant, even a small level of noise is an extremely damaging phenomenon. When recording analogue data and converting it to digital form, interference and noise cannot be avoided. Therefore, for sensitive signals, preprocessing techniques should be used to extract only useful information from the signal, and background components or noise are removed. There are many DSP techniques for cleaning up audio waveforms, from filters to gating and signal regeneration systems. Each of these techniques has advantages and disadvantages. Each technique should be selected according to the data, and in the case of data that are specific speech samples (only vowels repeated continuously with a similar intonation level), dedicated signal processing methods can be used.

A Bezier curve approximation of the speech signal in the classification process of laryngopathies

In the paper Szkoła, Pancerz and Warchoł (2011b) a technique for pre-regenerating speech signals using Bezier curves was proposed. Bezier curves are well-known for vector computer graphics, but can be successfully applied to other tasks. Bezier curves are derived from Bernstein polynomials, as shown below.

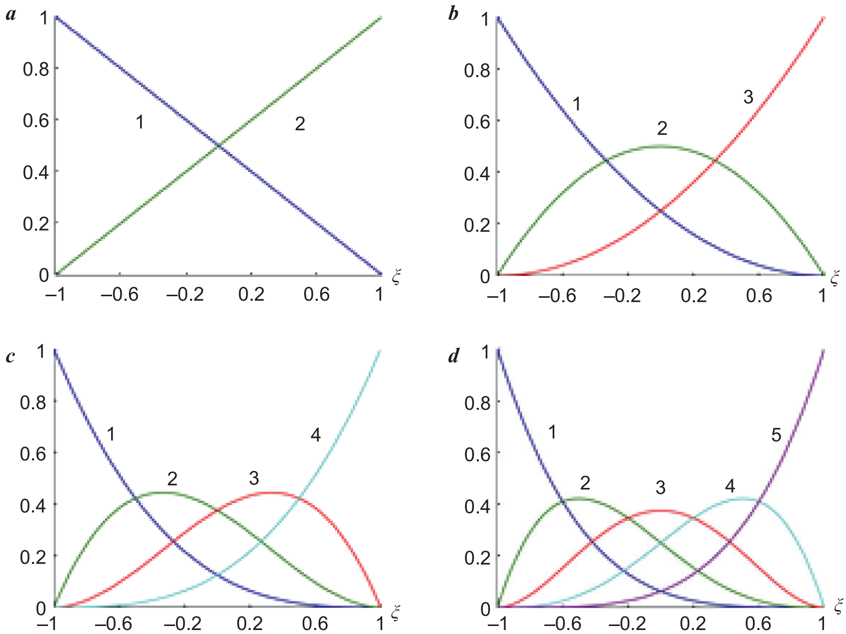

Fig. 8. Bernstein basis function for polynominal degree p: $a - p$=1, $b - p$=2, $c - p$=3, $d - p$=4

The unique properties of Bezier curves applied to the regeneration of speech samples: Bezier curves are always smooth (the curve must not contain noise), Bezier curves are affine invariant; No loss of quality due to scaling; Possibility of signal compression due to point coding of the curve.

This paper presents an algorithm for converting an audio sample into a sequence of points forming a set of connected Bezier curves in the form of so-called S-plines. The data processed in this way was used to train an Elman-Jordan neural network. On the basis of the tests carried out, it was noted that the results were not as good as for the modified Elaman-Jordan neural network. There may be several reasons for this, one of which may be an insufficiently good algorithm for converting audio waveforms into the form of Bezier curve control points. There is no deterministic algorithm that allows such an operation, the solution developed is not an approximation algorithm that examines the average distance of the Bezier curve points from the actual input waveform. This is a very simple approach, but no other methods have been found or developed in this area, so once a high quality conversion algorithm has been developed, the results that can be achieved should be much better.

Conclusions

In the paper, different approaches to speech signal analysis using recurrent neural networks (RNNs) for the diagnosis of laryngopathy, including Reinke's oedema and laryngeal polyps, are presented. Among other things, the focus was on creating a combine of Elman's neural network with Jordan's, achieving a better learning ability of the neural network to distinguish between normal and disease states at the same level. The combination provides the opportunity for accelerated network learning and to achieve a level of real-time decision-making, what is important in clinical practice. The modification of this hybrid presented in this review offers the possibility of achieving target patterns even faster. In addition, one study used Bézier curve preprocessing of the speech signal to reduce the amount of data and eliminate noise from the signal. Despite the fact that the quality of the models with Bézier curve in distinguishing between normal and pathological categories is not fully satisfactory, the presented review summarizes the research thread pointing in the direction of further research. The presented review may provide guidance in the selection of appropriate methods for the development of tools for computer-assisted diagnosis of laryngeal diseases.

References

Bakator M., Radosav D. 2018. *Deep Learning and Medical Diagnosis: A Review of Literature.* Multimodal Technologies and Interaction, 2(3):47. https://doi.org/10.3390/mti2030047

Elman J. 1993. *Distributed Representations, Simple Recurrent Networks, and Grammatical Structure.* Machine Learning, 7: 195-225. https://doi.org/10.1007/BF00114844

Jordan M.I. 1986. *Serial Order: A Parallel Distributed Processing Approach.* ICS Report 8604. Institute for Cognitive Science University of California, San Diego.

Mehrish A., Majumder N., Bharadwaj R., Mihalcea R., Poria S. 2023. *A Review of Deep Learning Techniques for Speech Processing.* Information Fusion, 99: 101869. https://doi.org/10.1016/j.inffus.2023.101869

Sfayyih A.H., Sabry A.H., Jameel S.M., Sulaiman N., Raafat S.M., Humaidi A.J., Kubaiaisi Y.M.A. 2023a. *Acoustic-Based Deep Learning Architectures for Lung Disease Diagnosis: A Comprehensive Overview.* Diagnostics, 13(10):1748. https://doi.org/10.3390/diagnostics13101748

Sfayyih A.H., Sulaiman N., Sabry A.H. 2023b. *A Review on Lung Disease Recognition by Acoustic Signal Analysis with Deep Learning Networks*. Journal of Big Data, 10(1), Article number 101. https://doi.org/10.1186/s40537-023-00762-z

Szkola J., Pancerz K., Warchol J. 2010a. *Computer Diagnosis of Laryngopathies Based on Temporal Pattern Recognition in Speech Signal*. Bio-Algorithms and Med-Systems, 6(12): 75-80.

Szkoła J., Pancerz K., Warchoł J. 2010b. *Computer-Based Clinical Decision Support for Laryngopathies Using Recurrent Neural Networks*. 10^{th} International Conference on Intelligent Systems Design and Applications, Cairo, Egypt, p. 627-632. https://doi.org/10.1109/ISDA.2010.5687195

Szkoła J., Pancerz K., Warchoł J. 2011a. *Improving Learning Ability of Recurrent Neural Networks: Experiments on Speech Signals of Patients with Laryngopathies*. In: F. Babiloni, A. Fred, J. Filipe, H. Gamboa (Eds.) *Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing* (BIOSIGNALS '11). SciTePress – Science and Technology Publications, Setúbal, p. 360-364. https://doi.org/10.5220/0003292603600364

Szkoła J., Pancerz K., Warchoł J. 2011b. *A Bezier Curve Approximation of the Speech Signal in the Classification Process of Laryngopathies*. Federated Conference on Computer Science and Information Systems (FedCSIS), Szczecin, p. 141-146. Retrieved from https://annals-csis.org/proceedings/2011/pliks/221.pdf

Tanveer M., Rastogi A., Paliwal V., Ganaie M.A., Malik A.K., Del Ser J., Lin C.-T. 2023. *Ensemble Deep Learning in Speech Signal Tasks: A Review*. Neurocomputing, 550: 126436. https://doi.org/10.1016/J.NEUCOM.2023.126436

Zaidi B., Selouani S., Boudraa M., Sidi Yakoub M. 2021. *Deep Neural Network Architectures for Dysarthric Speech Analysis and Recognition*. Neural Computing and Applications, 33: 9089-9108.