

Artificial intelligence facilitates decision-making in the treatment of lumbar disc herniations

André Wirries, Florian Geiger, Ahmed Hammad, Ludwig Oberkircher, Ingmar Blümcke, Samir Jabari

Angaben zur Veröffentlichung / Publication details:

Wirries, André, Florian Geiger, Ahmed Hammad, Ludwig Oberkircher, Ingmar Blümcke, and Samir Jabari. 2021. "Artificial intelligence facilitates decision-making in the treatment of lumbar disc herniations." *European Spine Journal* 30 (8): 2176–84.
<https://doi.org/10.1007/s00586-020-06613-2>.



Artificial intelligence facilitates decision-making in the treatment of lumbar disc herniations

André Wirries¹ · Florian Geiger¹ · Ahmed Hammad¹ · Ludwig Oberkircher² · Ingmar Blümcke³ · Samir Jabari³

Received: 25 April 2020 / Revised: 5 September 2020 / Accepted: 22 September 2020 / Published online: 13 October 2020
© The Author(s) 2020

Abstract

Purpose Apart from patients with severe neurological deficits, it is not clear whether surgical or conservative treatment of lumbar disc herniations is superior for the individual patient. We investigated whether deep learning techniques can predict the outcome of patients with lumbar disc herniation after 6 months of treatment.

Methods The data of 60 patients were used to train and test a deep learning algorithm with the aim to achieve an accurate prediction of the ODI 6 months after surgery or the start of conservative therapy. We developed an algorithm that predicts the ODI of 6 randomly selected test patients in tenfold cross-validation.

Results A 100% accurate prediction of an ODI range could be achieved by dividing the ODI scale into 12% sections. A maximum absolute difference of only 3.4% between individually predicted and actual ODI after 6 months of a given therapy was achieved with our most powerful model. The application of artificial intelligence as shown in this work also allowed to compare the actual patient values after 6 months with the prediction for the alternative therapy, showing deviations up to 18.8%.

Conclusion Deep learning in the supervised form applied here can identify patients at an early stage who would benefit from conservative therapy, and on the contrary avoid painful and unnecessary delays for patients who would profit from surgical therapy. In addition, this approach can be used in many other areas of medicine as an effective tool for decision-making when choosing between opposing treatment options, despite small patient groups.

Keywords Artificial intelligence · Lumbar disc herniation · Supervised machine learning · Outcome prediction · Conservative vs. Operative

Introduction

The decision for operative or conservative treatment of spinal disorders is often difficult as the evidence for treatment options is insufficient. Particularly in the case of lumbar disc

herniations, the decision between surgical and conservative therapy is often a challenging and physician-dependent decision. If the patients do not suffer from neurological deficits, usually a conservative therapy is started for at least 6 weeks up to months [1]. Only when this fails, surgery is proposed. Most patients undergoing surgery therefore have a history of failed conservative therapy. This delay can lead to longer incapacity to work, higher treatment costs and more frequent chronification of pain. Thus, it would be beneficial to decide as soon as possible whether a conservative therapy is promising or an early surgical intervention would provide the superior result.

The current literature on this topic usually describes comparable results one year after the onset of symptoms, regardless of how the patients were treated [2–6]. However, the known studies are frequently difficult to generalize as the design often allows only a limited application of the results into practice. A minimum duration of symptoms of

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00586-020-06613-2>) contains supplementary material, which is available to authorized users.

✉ André Wirries
Andre.Wirries@Hessing-Stiftung.de

¹ Spine Center, Hessing Foundation, Hessingstrasse 17, 86199 Augsburg, Germany

² Center of Orthopaedic and Trauma Surgery, Philipps University of Marburg, Baldingerstrasse, 35043 Marburg, Germany

³ Neuropathological Institute, University Hospitals Erlangen, Schwabachanlage 6, 91054 Erlangen, Germany

6–8 weeks is usually required, which means that all patients with severe pain who are operated on within this period are excluded [1, 2, 4, 7]. In this selection process, only those patients who have undergone conservative therapy over several weeks are included in studies. Moreover, conservative therapy is not precisely defined and the maximum duration of symptoms is rarely fixed, so that chronic and acute symptoms are equated. Together with high drop-out and cross-over rates in randomized controlled trials, this leads to a significant statistical weakness [7, 8]. As a consequence of the missing evidence in current literature it is recommended to start with a conservative therapy of about 6 weeks up to 4 months and then to decide for or against surgery [1]. This gives a distorted picture of this pathology and its practical implications [5]. Thus, other decision supporting tools should be used to decide earlier who will benefit from surgical therapy.

In recent years it has been shown that artificial intelligence (AI) can be used to make more and more precise predictions about patient outcomes [9–11]. AI is a generic term for various applications of complex, self-learning algorithms. This ranges from adapted, intensified statistical evaluations to self-learning neural networks. However, the previous AI approaches in spinal therapy have not made use yet of the possibilities of deep learning as presented in this work [9]. We investigated the potential role of supervised deep learning to support objective decision-making in the treatment of lumbar disc herniations.

Patients and methods

Patient population

The data of 60 patients of an ongoing observational study at the Spine Center of the Hessing Foundation in Augsburg were used in this feasibility study and analysed by a digital pathology and AI working group at University Hospitals Erlangen. All of the patients provided informed written consent to the use of their data in the study and agreed to publishing. Of these 60 patients 31 were male and 29 female. Thirty-three patients were treated surgically and 27 patients conservatively. These records did not include the data of 3 patients who withdrew their agreement in the course, nor of 6 patients who initially declined participation. Thus, the initial inclusion rate was 91% and the follow-up rate at 6 months was 95%. The baseline demographic data are shown in Table 1. The approval of the local ethics committee has been obtained in advance (Ethics Committee No. 16098 of the Bavarian Medical Association, Germany).

The patients were at least 18 years old and presented with radicular pain caused by a herniated disc of the lumbar spine, which was confirmed by an MRI scan. At the time of enrolment, the radicular symptoms did not last longer than 12 weeks. No minimum time for the duration of the symptoms was defined to include patients in the study. Exclusion criteria were instability or scoliosis in the

Table 1 Patient data (demographic and clinical scores) on admission

	<i>n</i> = 60				Surgery <i>n</i> = 33		Conservative <i>n</i> = 27	
	Mean	Std	Min	Max	Mean	Std	Mean	Std
Age	53.6	11.9	31.0	80.0	55.2	10.8	51.6	13.1
Days from onset of symptoms to 1st treatment	41.0	23.9	3.0	82.0	43.4	24.6	38.1	23.1
BMI	27.5	6.6	17.9	54.7	26.8	4.4	28.5	8.6
Patient standing height	173.2	9.9	157.0	196.0	172.5	8.7	173.9	11.3
Pack years (smoking)	8.1	13.3	0.0	47.0	10.7	15.2	4.9	10.0
SF 36 physical functioning	23.4	21.2	0.0	95.0	26.7	23.6	19.4	17.6
SF 36 limit. physical health	16.3	30.8	0.0	100.0	14.4	31.9	18.5	29.9
SF 36 limit. emotional problems	50.6	47.7	0.0	100.0	49.5	50.1	51.9	45.6
SF 36 energy fatigue	39.8	19.5	5.0	90.0	41.1	21.2	38.3	17.5
SF 36 emotional well being	61.6	19.2	20.0	100.0	61.5	21.3	61.8	16.6
SF 36 social functioning	51.9	28.0	0.0	100.0	56.4	31.4	46.3	22.4
SF 36 pain	26.3	26.9	0.0	100.0	26.5	30.0	26.0	23.1
SF 36 general health	59.8	18.0	10.0	100.0	63.5	17.2	55.2	18.2
ODI percentage	55.3	18.4	14.0	100.0	55.7	17.2	54.7	20.2
HADS anxiety	6.4	3.8	0.0	15.0	6	3.9	6.8	3.6
HADS depression	6.5	3.4	1.0	15.0	6.1	3.4	7.0	3.4
Back pain (VAS)	5.3	3.4	0.0	10.0	4.8	3.7	5.9	3.1
Leg pain (VAS)	7.1	2.2	1.9	10.0	7.6	1.7	6.6	2.6

segment of the herniated disc, advanced degeneration (e.g. spondylogenic spinal stenosis), a recurrent herniated disc and previously performed surgery in the affected segment.

Therapy

Conservative therapy consisted of analgesics, periradicular or peridural infiltrations, balneophysical measures and physiotherapy over a period of several days in an inpatient setting, followed by outpatient therapy after discharge. Surgical therapy was standardly performed as microscopically or endoscopically assisted interlaminar sequestrectomy.

Content and structure of learning and test group

Basic demographic data, as well as the MOS 36-Item Short Form Survey (SF-36) [12], the Oswestry Disability Index (ODI) [13] along with leg and back pain, each measured on a 100 mm visual analogue scale, and the Hospital Anxiety and Depression Scale (HADS) [14] for every patient were assessed on the day of admission. The ODI was defined as the target variable and re-assessed 6 months later (Table 1).

Special attention was paid from the beginning to the completeness of the data. As the success of a neural network training depends on complete data sets, any insufficient data were completed promptly together with the patients.

Stratified tenfold shuffled splitting of the 60 patients data (scikit-learn v.0.21.2) [15] resulted in a training set of 54 patients and a test set of 6 patients. Care was taken that gender and treatment (operative vs. conservative) were distributed evenly throughout the two sets.

Artificial intelligence-based prediction model

The data collected from the 60 Patients were stored in a comma separated value formatted file (csv). This file was read by the pandas python package (pandas v.0.23.1; python 3.6.7) [16]. Plotting of correlation matrix (matplotlib v.2.1.2 and seaborn v.0.8.1) [17], density distributions and histograms of various parameters as well as basic statistical operations were performed on the dataset.

For further machine learning processing, we defined the ODI score 6 months after the start of treatment as the target value for prediction, so the machine learning problem is a linear regression problem.

By applying recursive feature elimination, weighing of feature importance and analysis of intercorrelating features (feature selector v.1.0.0), many of the parameters within the csv file for a given patient were dropped in order to reduce complexity resulting in the final features fed to the model [18]. The parameters finally used to train the neural network after recursive feature elimination was applied are shown in Table 2.

Table 2 Variables identified and used for final training of the neural network

Categorical variables	Continuous variables	Target variable
Gender	Age	ODI after 6 month
Treatment option	Days from onset of symptoms to 1st treatment	
	SF 36 physical functioning	
	Patient standing height	
	SF 36 energy fatigue	
	SF 36 social functioning	
	SF 36 general health	
	Back pain	
	Leg pain	
	BMI	
	HADS depression	

Table 3 Performance results of various machine learning algorithms obtained in tenfold cross-validation

Machine learning algorithm	Mean absolute error rate	SD
Linear regression	11.3	4.0
ElasticNet	11.6	4.5
Nearest neighbour	12.7	4.5
Decision trees	14.2	4.2
Stochastic gradient descent regressor	10.9	
Support vector regressor	11.8	4.5

After identification of categorical variables and continuous variables, categorical variables were encoded using scikit-learns “LabelEncoder”. Various machine learning algorithms were cross-checked regarding their performance in tenfold cross-validation (Table 3). A simple but deep neural network architecture (three layers) was identified to be most promising (Fig. 1). This approach was further targeted and evaluated as follows.

Each categorical variable was now fed separately into the network via an embedding layer, while the rest of the continuous variables were collected in an additional array and fed into the network via one separate input. In total, the model used had two categorical inputs. All inputs were concatenated and processed through two additional hidden layers with rectified linear activation functions and a subsequent linear output at the last layer. The Keras framework (v. 2.2.4) with tensorflow backend (v.1.12.0) was used to model the network architecture and perform network training [19]. Grid searching of various parameters was performed and all training was evaluated by tenfold cross-validation. The training was performed for 1000 epochs while early stopping revealed best model performance at epoch 488. Cycling the learning rate within each epoch between 0.001 and 0.1 with

Fig. 1 Model architecture: the categorical data were introduced into the model via respective embedding layers. The continuous data were introduced into the model via a separate input. All data were then linked together and processed via three additional layers

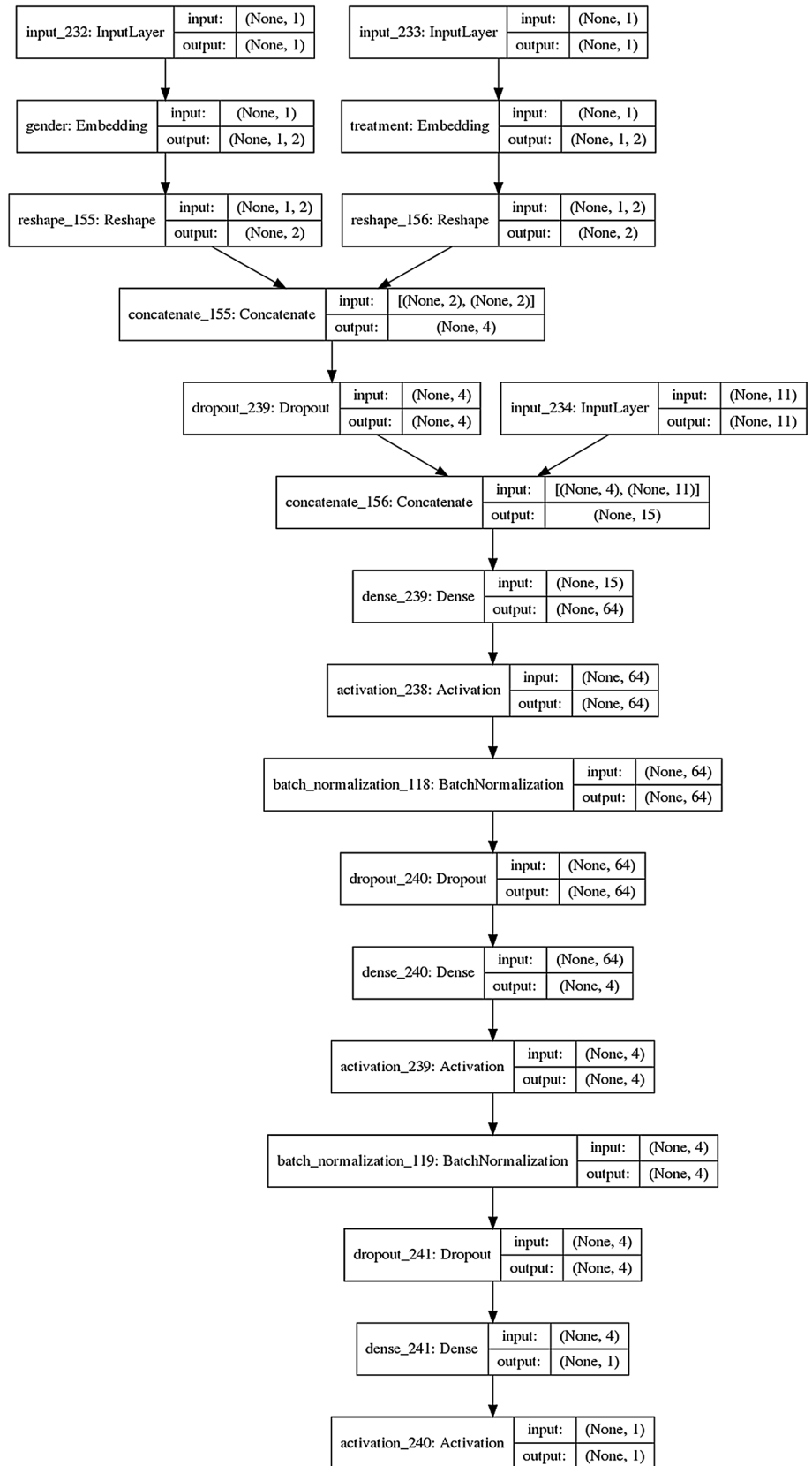


Table 4 Mean absolute error rate and mean and standard deviation of the performed tenfold cross-validation for the prediction of the ODI after 6 months using our neural network

Fold	1	2	3	4	5	6	7	8	9	10	Mean	SD
Mean absolute error	7.1	8.5	3.4	6.8	6.9	3.8	1.5	8.4	5.6	7.11	5.9	2.3

Table 5 Actual and AI predicted ODI values for conservatively treated test patients

Actual ODI after 6 months	AI-predictions for ODI in 6 months	
	Surgery	Conservative
Conservative		
10	2.0	8.4
12	2.1	9.1

Table 6 Actual and AI predicted ODI values for surgically treated test patients

Actual ODI after 6 months	AI-predictions for ODI in 6 months	
	Surgery	Conservative
Surgery		
30	29.9	12.8
46.9	50.3	31.5
2	2.1	13.2
12	11.2	7.9

a batch size of 54 (all training samples at once) showed the best results. All values were saved and loss curves as well as mean absolute error rates were plotted. Model evaluation was additionally performed by binning the regression output variable into bins of 12% ranges hence turning the final continuous regression prediction into a categorical problem.

Finally, after training was complete, we evaluated the model predictions of the ODI 6 month after start of treatment. Therefore, we compared the model-prediction for the patients in the test set with the real life values of these patients. Next we compared the predictions for both kind of therapies for each patient in the test set. In this way, the prediction values for the actually applied therapy could be compared with the prediction values for the other, not applied, form of therapy (Tables 5, 6).

Results

Prediction accuracy

The complete data sets of 60 patients were used and all model evaluations were performed in a tenfold cross-validation. The mean absolute error in the cross-validation was 5.9% on the test data set. Our best-performing neural network had a surprisingly low mean absolute error of 1.5% (Table 4). The mean absolute error of our worst performing model was 8.5% and with under 10% still better performing than every other tested machine learning algorithm (Table 3). All further results refer to our best performing model.

Using our deep learning algorithm, a maximum difference between individually predicted and actual ODI after 6 months of therapy of only 3.4% could be achieved. From our point of view, this is a sufficient correlation between the prediction and the real values to be able to use the predictions for therapy prognoses.

Dividing the ODI (with a percentage value from 0 to 100) as a target value into ranges of 12%, a 100% accurate prediction of the individual percentage range at the time of the 6 month evaluation could be achieved.

Comparison of the AI-predictions for different forms of therapy

In our best performing model the test data set consisted of 2 conservatively and 4 surgically treated patients. Table 5 and 6 present the ODI results of these patients with the corresponding AI predictions. The first column shows the ODI values actually achieved after 6 months, the shaded 2nd and 3rd columns show the AI predictions for both therapy forms. Thus, the conservative and operative therapy predictions can be compared for the same patient.

The conservative patients in Table 5 reached ODI values of 10% and 12% after 6 months of therapy, shown in the first column. For the same patients the AI prediction for surgical therapy (2nd column) gave low values of 2% and 2.1%. The AI prediction for conservative therapy (3rd column) showed with values of 8.4% and 9.1% a good approximation to the actually achieved ODI.

Table 6 presents the surgically treated test-patients. The first column shows the actually achieved ODI 6 months after surgery. There are pronounced inter-individual differences

with values between 2% and 46.9%. The AI predictions for the operative (2nd column) and conservative (3rd column) forms of therapy also show individual differences, some of them considerable. The first test patient has a real ODI of 30% 6 months after surgery. The corresponding AI prediction showed with 29.9% almost the same value for the operative, but significantly better 12.8% for the conservative therapy. The second test patient showed an unsatisfactory result 6 months after surgery with an actual ODI of 46.9%, the prediction for the operative therapy would have been even worse with 50.3%, while a slightly better result was predicted for the conservative therapy with 31.5%. The remaining 2 test patients achieved very good ODI values of 2% and 12% 6 months after operative care. The AI prediction for this surgical therapy (2nd column) was very accurate at 2.1% and 11.2%, respectively. The AI prediction for conservative therapy (3rd column) was only moderately different in these patients with a slightly worse ODI of 13.2% for the third and slightly better ODI of 7.9% for the fourth test patient. (For a more detailed presentation of the individual patient data presented here, we refer to the additional material in the Supplementary Tables 1.1 and 1.2.)

The deviations of the predictions for different treatment options (2nd and 3rd column in Tables 5 and 6) in the test patients ranged from 3.3 to 18.8%.

Discussion

Our approach to AI—supervised deep learning

Overall, our presented model shows good convergence and surprisingly good predictive power. Considering the fact that the minimal clinically important difference (MCID) in the German version of the Oswestry Disability Index is reported to be around 9% (with 95% confidence) [13], the exact prediction of a 12% range within the ODI can be regarded as sufficient to derive an individual therapy recommendation.

It is important to clarify that the supervised deep learning approach presented here differs significantly from an unsupervised approach, where large amounts of unfiltered data are processed. This is the most important reason why we are able to achieve sufficient training of our AI algorithm with a data set of only 60 patients. Furthermore, we assume that the good predictive power of this supervised deep learning approach is made possible by the fact that the data acquisition was designed from the beginning to be processed with supervised deep learning. In particular, time-consuming repeated interviews with individual patients ensured that the data sets were as accurate and complete as possible. From our point of view, this enables a much more efficient learning process, since insufficient data do not have to be compensated by quantity.

In contrast to the unsupervised big data approach (trying to find patterns not yet known within the data), our supervised deep learning approach predicts a specific clinical parameter, the ODI. This means that our model tries to make a prediction for a specific value based on the collected patient-related parameters, rather than grouping the data. Our artificial neural network learns the aspects of a disease pattern by repeated learning processes on complete data sets. The resulting AI is able to perform an increasingly good prediction of therapy outcomes for previously unknown data sets. In the neural network used here, the establishment of a good predictive power can be seen after training with about 50 patient data sets (Fig. 2).

Interpretation of results

It is questionable whether differences in the predictions for therapy options that are smaller than the MCID, which is described at about 9% for the ODI [13], are individually noticeable. It should be assumed, that these prediction differences remain below the perception threshold. Thus, the results seem to be similar for both types of therapy in several cases, which would correspond to the current literature.

The partly noteworthy differences in the actually achieved ODI 6 months after the operation with values up to 46.9% show a pronounced individuality of the operative success. We show cases that in reality reach unsatisfactory ODI values after an operation. Differences in the corresponding AI prediction, clearly exceeding the MCID, are worth a detailed examination. Using these AI predictions could have led to better results under conservative therapy. On the other hand, the purpose of surgery is questionable if the predicted results are comparable. Moreover, there are also

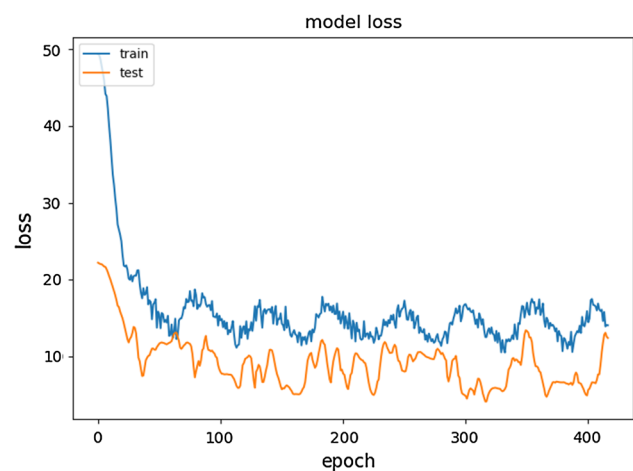


Fig. 2 Loss curve: the image shows loss curves for the training and respective validation data. Convergence of both validation and training loss towards the bottom line before the beginning of the 100th epoch indicate a successful learning process

cases where conservative therapy produces worse results in the AI predictions. This proves that there are also patients who benefit from early surgical intervention.

Interestingly, the maximum deviation of the predictions for different treatment options of a patient ranged from 3.3 to 18.8%. As they do not always exceed the MCID these different AI predictions would not always be perceived differently. These results confirm to a certain extent the existing literature, where conservative and operative therapies often lead to comparable results after a follow-up period of 6 months. However, our results also show that in some cases the decision for one way of therapy has a noticeable impact on the outcome. If AI predictions were made at an early stage, the selection of the individually best therapy would be facilitated and suboptimal outcomes could be avoided.

Artificial intelligence in spine therapy

Despite many valuable studies, objective decision-making in the individual case of a lumbar disc herniation is only possible to a limited extent. Most studies provide generalized statements, which mainly recommend to refrain from surgical therapy if possible [2, 8, 20]. All those studies generalize and do not take the individual fate into account. In daily clinical practice, however, a high percentage of patients choose and benefit from surgery despite this approach [2] and patients who receive intensive conservative therapy may develop chronic pain or decide for surgery at a later stage. As most studies have a cross over rate of more than 30% a generalized recommendation for conservative or surgical therapy is not justified. Therefore, the possibility to predict the optimal therapy at an early stage would be a valuable aid for individual decision-making.

The concept of individualized decision-making has been discussed repeatedly, but so far has not been consistently implemented in the field of spine therapy. AI techniques based purely on extensive statistics were already used by McGirt et al. in 2015 to predict various outcome variables together with the ODI one year after spinal surgery. This prediction model was able to achieve an accuracy between 72 and 84% for more than 40 different values; however, it was not based on machine learning [21]. Kim et al. went a step further in 2018 and developed a prediction of various complications after spondylodesis using logistic regression and combining it with a shallow artificial neural network. They already achieved comparatively better prediction results than the clinical scores usually used [11]. With an overall accuracy of 87.6%, it was recently shown that even a combination of decision algorithms could predict the most important intra- or perioperative complications following spinal deformity surgery in adults. However, no real deep learning has been applied [22].

Another way to enable individualized decisions is the establishment of a DST (Decision Support Tool). This attempts to support clinical decisions by providing personalized predictions. There is already a DST available for spinal treatment, the Nijmegen Decision Tool for Chronic Low Back Pain. This tool recommends a surgical or conservative therapy, but also a discontinuation of therapy, taking into account various patient-related characteristics [23, 24]. However, this DST is currently under further development and the technical implementation still needs to be finalized. Only recently it could be shown that the use of unsupervised artificial intelligence is suitable to achieve good risk detection by hierarchical grouping of a large patient population. However, in contrast to our presented supervised AI approach, the necessity to acquire and process huge amounts of data is a challenge [10].

Comparing the role of AI in clinical decision-making in spinal therapy with other applications of AI and machine learning, there is still much need for development. Generally speaking, machine learning methods still do not play a relevant role in decision-making in spinal surgery. The predictive models presented so far are generally not based on modern techniques such as deep learning as applied in our study [9]. According to our current state of knowledge, this is the first work that provides a state-of-the-art design of neural network that can be successfully used to predict the outcome of treatment for lumbar disc herniations. As these first results are encouraging, we plan to adapt it to other problems, where a statistical analysis or double blind studies were not successful in the past.

Restrictions to the presented model

The number of patients with whom our model has been trained was small and further patient data will be included on an ongoing basis to further improve the prediction results. However, this is exactly one of the advantages of supervised deep learning, namely the ability to carry out an effective learning process and make good individual predictions possible with a small number of cases. Furthermore, it will also be necessary to objectively verify whether the predictions of artificial intelligence are in accordance with the results of clinical practice. A validation of these predictions is already planned to be carried out.

In addition, we observed that when there were significant deviations in the predictions for conservative and operative therapy (and thus greater relevance for therapy decisions), the deviations from the actual ODI value also tended to be greater, albeit still within the MCID—we might see a further improvement by increasing the number of training-patients.

In this study no correlation with MRI imaging was made. It is known that MRI imaging only correlates with clinical symptoms to a certain degree [25], but an appropriate implementation in the further development of the prediction makes sense

and is planned in the ongoing study to ensure the completion of patient-related data.

Conclusion

We believe that the approach of a supervised artificial intelligence will improve the predictability of a therapy outcome and thus help to individualize therapy recommendations for patients such as those with a disc herniation.

This approach (especially the presented supervised deep learning version) can serve as a basis for further developments of AI, not only in the field of spinal therapy, but also in many other areas of medicine where randomization or inclusion of high patient numbers is not feasible.

Funding No funding has been utilized for this study.

Data availability The algorithms used in the present study can be reviewed upon reasoned request via the corresponding author.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Consent for publication Informed written consent was provided by every participant.

Ethics approval Approval was obtained from the Ethics Committee of the Bavarian Medical Association, Germany (Ethics Approval Number 16098). The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bailey CS, Rasoulinejad P, Taylor D, Sequeira K, Miller T, Watson J, Rosedale R, Bailey SI, Gurr KR, Siddiqi F, Glennie A, Urquhart JC (2020) Surgery versus conservative care for persistent sciatica lasting 4 to 12 months. *N Engl J Med* 382:1093–1102. <https://doi.org/10.1056/NEJMoa1912658>
- Ramakrishnan A, Webb KM, Cowperthwaite MC (2017) One-year outcomes of early-crossover patients in a cohort receiving nonoperative care for lumbar disc herniation. *J Neurosurg Spine* 27:391–396. <https://doi.org/10.3171/2017.2.SPINE16760>
- Pearson AM, Lurie JD (2014) Surgical versus nonoperative treatment: how do we choose the right approach to lumbar disk herniation? *Pain Manag* 4:247–249. <https://doi.org/10.2217/pmt.14.28>
- Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Hanscom B, Skinner JS, Abdu WA, Hilibrand AS, Boden SD, Deyo RA (2006) Surgical vs nonoperative treatment for lumbar disk herniation: the Spine patient outcomes research trial (SPORT): a randomized trial. *JAMA* 296:2441–2450
- Mazanec D, Okereke L (2007) Interpreting the spine patient outcomes research trial. Medical vs surgical treatment of lumbar disk herniation: implications for future trials. *Cleve Clin J Med* 74:577–583
- Peul WC, van den Hout, WB, Brand R, Thomeer RT, Koes BW, and Leiden-The Hague Spine Intervention Prognostic Study Group. (2008). Prolonged conservative care versus early surgery in patients with sciatica caused by lumbar disc herniation: two year results of a randomised controlled trial. *Bmj* 336:1355–1358
- Peul WC, van Houwelingen HC, van den Hout WB, Brand R, Eekhof JA, Tans JT, Thomeer RT, Koes BW (2007) Surgery versus prolonged conservative treatment for sciatica. *N Engl J Med* 356:2245–2256
- Kerr D, Zhao W, Lurie JD (2015) What are long-term predictors of outcomes for lumbar disc herniation? a randomized and observational study. *Clin Orthop Relat Res* 473:1920–1930. <https://doi.org/10.1007/s11999-014-3803-7>
- Galbusera F, Casaroli G, Bassani T (2019) Artificial intelligence and machine learning in spine research. *JOR spine* 2:e1044
- Ames CP, Smith JS, Pellisé F, Kelly M, Alanay A, Acaroglu E, Pérez-Grueso FJS, Kleinstück F, Obeid I, Vila-Casademunt A (2019) Artificial intelligence based hierarchical clustering of patient types and intervention categories in adult spinal deformity surgery: towards a new classification scheme that predicts quality and value. *Spine* 44:915–926
- Kim JS, Merrill RK, Arvind V, Kaji D, Pasik SD, Nwachukwu CC, Vargas L, Osman NS, Oermann EK, Caridi JM, Cho SK (2018) Examining the ability of artificial neural networks machine learning models to accurately predict complications following posterior lumbar spine fusion. *Spine* 43:853–860. <https://doi.org/10.1097/brs.0000000000002442>
- Ware JE Jr, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care* 30:473–483
- Mannion A, Junge A, Grob D, Dvorak J, Fairbank J (2006) Development of a German version of the Oswestry disability index. part 2: sensitivity to change after spinal surgery. *Eur Spine J* 15:66–73
- Hinz A, Brähler E (2011) Normative values for the hospital anxiety and depression scale (HADS) in the general German population. *J Psychosom Res* 71:74–78
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V (2011) Scikit-learn: machine learning in Python. *J mach learn res* 12:2825–2830
- Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J (2020) SciPy fundamental algorithms for scientific computing in Python. *Nat Methods* 17:1–12
- Hunter JD (2007) Matplotlib: A 2D graphics environment. *Comput Sci Eng* 9:90–95
- Koehrsen W Feature-selector. GitHub repository. 2020. <https://github.com/WillKoehrsen/feature-selector>. Accessed 4 Oct 2020
- Chollet F (2015) Keras. GitHub repository. 2020. <https://github.com/fchollet/keras-resources>. Accessed 4 Oct 2020
- Lurie JD, Tosteson TD, Tosteson AN, Zhao W, Morgan TS, Abdu WA, Herkowitz H, Weinstein JN (2014) Surgical versus nonoperative treatment for lumbar disc herniation: eight-year results for the spine patient outcomes research trial. *Spine (Phila Pa 1976)* 39:3–16

21. McGirt MJ, Sivaganesan A, Asher AL, Devin CJ (2015) Prediction model for outcome after low-back surgery: individualized likelihood of complication, hospital readmission, return to work, and 12-month improvement in functional disability. *Neurosurg Focus* 39:E13
22. Scheer JK, Smith JS, Schwab F, Lafage V, Shaffrey CI, Bess S, Daniels AH, Hart RA, Protopsaltis TS, Mundis GM (2017) Development of a preoperative predictive model for major complications following adult spinal deformity surgery. *J Neurosurg Spine* 26:736–743
23. van Hooff ML, Mannion AF, Staub LP, Ostelo RW, Fairbank JC (2016) Determination of the Oswestry disability index score equivalent to a “satisfactory symptom state” in patients undergoing surgery for degenerative disorders of the lumbar spine—a Spine tango registry-based study. *Spine J* 16:1221–1230
24. Coupé VM, van Hooff ML, de Kleuver M, Steyerberg EW, Ostelo RW (2016) Decision support tools in low back pain. *Best Pract Res Clin Rheumatol* 30:1084–1097
25. Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, Ross JS (1994) Magnetic resonance imaging of the lumbar spine in people without back pain. *N Engl J Med* 331:69–73

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.