



Article

Performance of Artificial Intelligence-Based Algorithms to Predict Prolonged Length of Stay after Lumbar Decompression Surgery

Babak Saravi ^{1,2,3,*} , Alisia Zink ², Sara Ülkümen ², Sebastien Couillard-Despres ^{3,4} , Frank Hassel ^{2,†} 
and Gernot Lang ^{1,2,†} 

- ¹ Department of Orthopedics and Trauma Surgery, Medical Center—University of Freiburg, Faculty of Medicine, University of Freiburg, 79108 Freiburg, Germany; gernotmichaellang@gmail.com
- ² Department of Spine Surgery, Loretto Hospital, 79108 Freiburg, Germany; alisia.zink@gmail.com (A.Z.); sara.uelkuemen@jupiter.uni-freiburg.de (S.Ü.); frank.hassel@rkk-klinikum.de (F.H.)
- ³ Institute of Experimental Neuroregeneration, Spinal Cord Injury and Tissue Regeneration Center Salzburg (SCI-TReCS), Paracelsus Medical University, 5020 Salzburg, Austria; s.couillard-despres@pmu.ac.at
- ⁴ Austrian Cluster for Tissue Regeneration, 1200 Vienna, Austria
- * Correspondence: babak.saravi@jupiter.uni-freiburg.de
- † These authors contributed equally to this work.

Abstract: Background: Decompression of the lumbar spine is one of the most common procedures performed in spine surgery. Hospital length of stay (LOS) is a clinically relevant metric used to assess surgical success, patient outcomes, and socioeconomic impact. This study aimed to investigate a variety of machine learning and deep learning algorithms to reliably predict whether a patient undergoing decompression of lumbar spinal stenosis will experience a prolonged LOS. Methods: Patients undergoing treatment for lumbar spinal stenosis with microsurgical and full-endoscopic decompression were selected within this retrospective monocentric cohort study. Prolonged LOS was defined as an LOS greater than or equal to the 75th percentile of the cohort (normal versus prolonged stay; binary classification task). Unsupervised learning with K-means clustering was used to find clusters in the data. Hospital stay classes were predicted with logistic regression, RandomForest classifier, stochastic gradient descent (SGD) classifier, K-nearest neighbors, Decision Tree classifier, Gaussian Naive Bayes (GaussianNB), support vector machines (SVM), a custom-made convolutional neural network (CNN), multilayer perceptron artificial neural network (MLP), and radial basis function neural network (RBNN) in Python. Prediction accuracy and area under the curve (AUC) were calculated. Feature importance analysis was utilized to find the most important predictors. Further, we developed a decision tree based on the Chi-square automatic interaction detection (CHAID) algorithm to investigate cut-offs of predictors for clinical decision-making. Results: 236 patients and 14 feature variables were included. K-means clustering separated data into two clusters distinguishing the data into two patient risk characteristic groups. The algorithms reached AUCs between 67.5% and 87.3% for the classification of LOS classes. Feature importance analysis of deep learning algorithms indicated that operation time was the most important feature in predicting LOS. A decision tree based on CHAID could predict 84.7% of the cases. Conclusions: Machine learning and deep learning algorithms can predict whether patients will experience an increased LOS following lumbar decompression surgery. Therefore, medical resources can be more appropriately allocated to patients who are at risk of prolonged LOS.

Keywords: length of stay; spine surgery; decompression; spinal stenosis; machine learning; deep learning; artificial intelligence; prediction



Citation: Saravi, B.; Zink, A.; Ülkümen, S.; Couillard-Despres, S.; Hassel, F.; Lang, G. Performance of Artificial Intelligence-Based Algorithms to Predict Prolonged Length of Stay after Lumbar Decompression Surgery. *J. Clin. Med.* **2022**, *11*, 4050. <https://doi.org/10.3390/jcm11144050>

Academic Editor: Akihiko Hiyama

Received: 17 May 2022

Accepted: 11 July 2022

Published: 13 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In patients over 65, lumbar spinal stenosis (LSS) is the most commonly occurring reason for spinal surgery, and decompression of the lumbar spine is among the most common

surgical procedures [1]. In theory, surgical decompression should relieve discomfort and improve function. In addition to laminectomy and laminotomy, fusion procedures may also be used as decompression methods. The purpose of spinal surgery is to reduce the pressure on the nerves afflicted by spinal stenosis by improving the cross-sectional area of the spinal canal [2]. Nearly a third of the general population suffers from lower back pain. Approximately 80 to 100 billion USD are spent annually on treating these patients [3,4]. The annual incidence of LSS surgery in the United States is approximately 1,2 per 1000; however, the prevalence of LSS in adults is expected to rise 59% by the year 2025 [5,6]. In terms of hospital expenses alone, the total cost of surgical intervention amounted to 1.65 billion USD [4]. An estimated 306 million USD have been spent on lumbar spine surgeries in the United States in the last few years among patients aged 65 and older [4,7,8]. In particular, by 2050, the elderly population is projected to increase from 43.1 million people to 83.7 million people, which will result in a significant increase in healthcare expenditures on degenerative spine diseases. An average of 8–10% of all patients who undergo spinal decompression need to undergo the procedure again, resulting in higher hospital costs [9,10]. Furthermore, these surgical procedures carry a 3.1% risk of cardiac problems or stroke, as well as a 0.4% risk of death within one month [4].

An important factor to consider when evaluating socioeconomic costs, surgical success, and patient outcomes is the length of stay (LOS). LOS is a measure that is frequently used to reduce healthcare costs, especially in light of escalating healthcare expenditures. When patients are undergoing spinal surgery, the length of stay variable is critical from the patient's point of view and is a major factor in determining health care expenditures. It is estimated that each additional day in the hospital costs approximately 1000 USD [11], and inpatient hospital expenditures (excluding instrumentation and surgical procedures) strongly correlate with the length of stay [12]. It is estimated that hospitalization costs for spine surgery increased nearly threefold between 1992 and 2003 [13,14]. Hospitalization costs are significant and should not be underestimated. In order to define value-based care for spinal illnesses, it is crucial to determine the precise length of stay. Ensuring that patients are not kept in the hospital for longer than necessary is an important goal. An analysis conducted by Boylan and colleagues on the costs associated with adolescents' scoliosis surgery found that each additional day of hospitalization costs in excess of 1100 USD in insurance expenditures and in excess of 5200 USD in terms of hospital charges [15]. Additionally, long-term hospital patients can incur an additional 19,000 USD in costs in comparison to those with a shorter hospital stay [16]. For other spinal fusion surgeries, similar results have been observed, indicating the need for further research either to reduce LOS significantly or to prevent unnecessary extensions [17].

Predictive analysis and machine learning have emerged as valuable tools for predicting patient outcomes based on pertinent feature characteristics variables [18]. Developing patient-centered outcome prediction models, including those for patient-related outcome measures and length of stay, can contribute to improving society's utilization of healthcare resources [19,20]. In doing so, policymakers and clinicians could compare treatments across disciplines to determine how best to allocate budget resources among different approaches.

Many studies have already shown that certain characteristics are associated with a longer hospital stay following spinal surgery. A longer length of stay (LOS) was associated with an increase in operating time during adult spinal deformity surgery, according to Phan et al. [15]. The authors of a multicenter study found that age, the number of levels fused, infection, and comorbidities are risk factors for a higher death rate [16]. Despite their ability to identify individuals likely to have a longer LOS, risk indicators cannot predict whether a patient will have a longer LOS. The use of machine learning and deep learning algorithms can enhance the knowledge provided by these studies and predict whether a particular patient will require a normal length of hospital stay or an extended hospital stay based on a wide range of clinicopathological characteristics. In particular, these algorithms can be integrated into the hospital's software environment, resulting in continuous monitoring of at-risk patients and the achievement of precision medicine goals.

Considering the socioeconomic impact of prolonged LOS, the present study sought to investigate various artificial intelligence-based algorithms that might be able to predict whether lumbar decompression surgery patients will experience a short or long LOS. Further, we aim to find the most relevant feature variables which are important to solve this classification task.

2. Materials and Methods

2.1. Study Design

We performed a retrospective cohort study including consecutive patients treated with microsurgical decompression or full-endoscopic interlaminar decompression of lumbar spinal stenosis between 2016 and 2021 at the Department of Spine Surgery, Loretto-Hospital Freiburg in Germany, an affiliated hospital of the University Medical Center Freiburg, Germany. This retrospective observational study was approved by the local Ethics Committee Freiburg, Germany [Number: 116/200]. Written informed consent to participate in observational studies was obtained from each patient.

The main inclusion criterion involved patients with lumbar spinal stenosis treated with either microsurgical or full-endoscopic decompression in the aforementioned period. The iLESSYS[®] system (Joimax GmbH, Karlsruhe, Germany) was utilized for the endoscopic group. After collecting all data from patients fulfilling our inclusion criteria, we applied our exclusion criteria for filtering the initial dataset. Exclusion criteria included: <18-years-old patients, patients with tumors of the spine, patients having spinal fusion, and patients who have declined the usage of their data for research purposes.

2.2. Data Handling and Statistical Analysis

Patients were collected from the in-house patient information system and extracted into a predefined datasheet. Data was pseudonymized utilizing a code generated with the “encode” command in Stata Statistical Software Release 15 (StataCorp. 2011, College Station, TX, USA). Variables were included in the study extraction form according to our previous study [20] and a literature search to only consider previously identified significant variables for hospital length of stay. The surgery-related and clinical factor variable group included surgical technique (microsurgical versus full-endoscopic decompression), number of targeted levels, operation time (OT), hospital length of stay (LOS), the American Society of Anesthesiologists (ASA) physical status classification, and perioperative and postoperative complications. Patient data were screened for the following complications: residual sensorimotor deficits or new-onset sensorimotor deficits, hematomas requiring revision, persisting stenosis requiring revision, postoperative instability, and fracture. This group also contained the demographic data for descriptive statistics (sex, age, BMI), alcohol and nicotine use data, and German insurance type (private or public insurance). The laboratory variable group included preoperative C-reactive protein (CRP) levels.

The target classes were classified as prolonged when hospital length of stay was $\geq 75\%$ percentile and normal when $< 75\%$ percentile (binary classification task) [21,22]. Unsupervised learning with K-means clustering was used to evaluate whether data can be clustered according to the features in the dataset. Auto clustering was used according to Schwarz’s Bayesian Criterion (BIC). Comparison of cluster variables was made with the Mann–Whitney U test or Chi-square test where applicable. Further, a decision tree with the Chi-square automatic interaction detection algorithm was applied. CHAID is a statistical tool to find the difference between child and parent nodes. The difference between observed and expected counts of the target variable for each node and the squared sum of these standardized differences will give the Chi-square value. CHAID algorithm was applied with cross-validation ($n = 5$), maximum tree depth of 3, minimum cases in parent node 10, and minimum cases in child node of 5, resulting in $n = 19$ nodes, $n = 11$ terminal nodes, and a depth of 3. Multiple imputations were applied utilizing linear regression for scale variables and logistic regression for categorical variables to impute missing values for the CHAID algorithm. Supervised machine learning and deep learning

techniques were applied to predict the target LOS classes. The following algorithms were used: logistic regression, RandomForest classifier, stochastic gradient descent (SGD) classifier, K-nearest neighbors, Decision Tree classifier, Gaussian Naive Bayes (GaussianNB), support vector machines (SVM), a custom-made convolutional neural network (CNN), multilayer perceptron artificial neural network (MLP), and radial basis function neural network (RBNN). The hardware and software environment specifications were as follows:

- CPU: AMD Ryzen 9 5950X 16-Core Processor (Santa Clara, CA, USA)
- RAM: 64 GB
- GPU: NVIDIA Geforce RTX 3090 (Santa Clara, CA, USA)
- Python version: 3.10.4 (64-bit) (Wilmington, DE, USA)
- OS: Windows 10 (Redmond, WA, USA)

Statistical analyses were conducted in Python and SPSS v26 (IBM, Armonk, NY, USA).

3. Results

We first performed unsupervised learning via K-means cluster analysis to cluster the dataset into two classes and evaluate which features are important for clustering. The first cluster analysis was done with the hospital length of stay (LOS) as raw data type (continuous scale). The cluster analysis resulted in two clusters (74.5% in the high risk group and 25.5% in the low risk group) (Figure 1). The data distribution showed that all cluster variables (BMI, number of levels, age, preoperative CRP, and LOS) were significantly higher for the high risk cluster than low risk cluster ($p < 0.0001$ for all comparisons) (Table 1). Overall, it was possible to find two clusters indicating patients' risk levels (i.e., the high risk group showed higher BMI, number of targeted levels, age, LOS, and preoperative CRP).

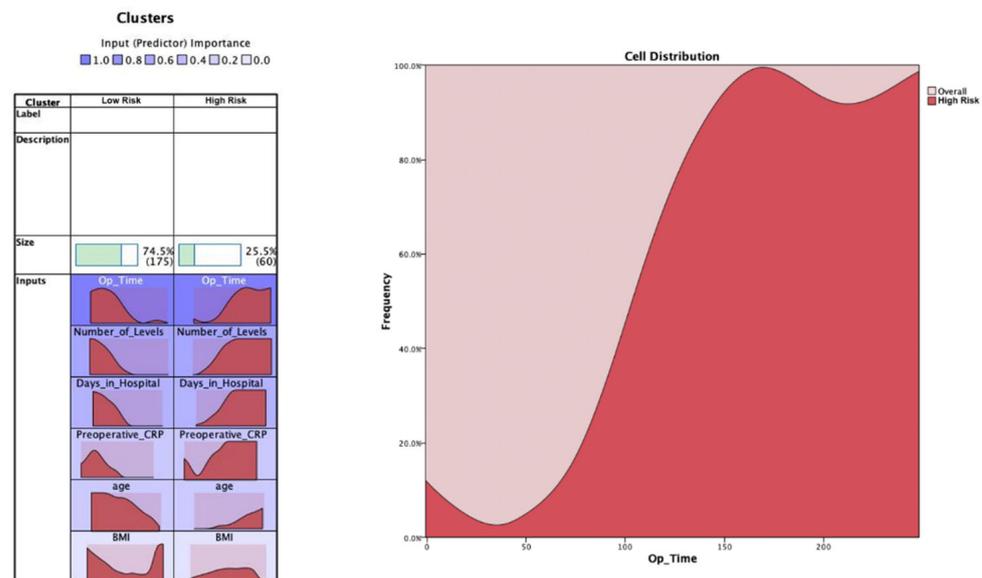


Figure 1. K-means cluster analysis for the raw hospital length of stay target variable (continuous scale). The most important feature to classify the data was the operation time, followed by the number of levels, LOS, preoperative CRP, age, and BMI. The distribution charts show the distribution of the features for both clusters. An example feature (operation time) is shown to help interpretation. The selected feature shows that the distribution of operation time is right-shifted for the high risk cluster, whereas it is left-shifted for the low risk cluster.

Table 1. Descriptive statistics and pairwise comparisons of the clusters were obtained through the K-means cluster analysis, including the continuous scale feature hospital length of stay (LOS). **** $p < 0.0001$.

		BMI		Number_of_Levels		Op_Time		Preoperative_CRP		Age		Days_in_Hospital	
		Mean	Std. De- viation	Mean	Std. De- viation	Mean	Std. De- viation	Mean	Std. De- viation	Mean	Std. De- viation	Mean	Std. De- viation
Cluster	Low risk	27.39	5.670	1.04	0.242	55.04	29.086	4.45	13.325	57.2366	16.37208	11.70	5.482
	High risk	28.75	5.145	1.92	0.973	137.63	51.662	40.90	66.902	72.2849	11.97116	24.27	14.429
	Combined	27.74	5.562	1.27	0.656	76.13	51.034	13.76	38.912	61.0787	16.69396	14.91	10.247
	<i>p</i> -value	****	****	****	****	****	****	****	****	****	****	****	****

The second cluster analysis was done with the hospital length of stay (LOS) as a binary class type utilizing the 75% percentile as the cut-off point. The cluster analysis resulted in two clusters (26.8% in cluster 1 and 73.2% in cluster 2) (Figure 2). The data distribution showed that all cluster variables (BMI, number of levels, age, preoperative CRP) were significantly higher for cluster number 1 than cluster number 2 ($p < 0.0001$ for all comparisons) (Table 2). Further, the distribution of LOS class was significantly different between clusters 1 and 2, with all prolonged cases being in cluster 1 and normal cases being in cluster 2 ($p < 0.0001$).

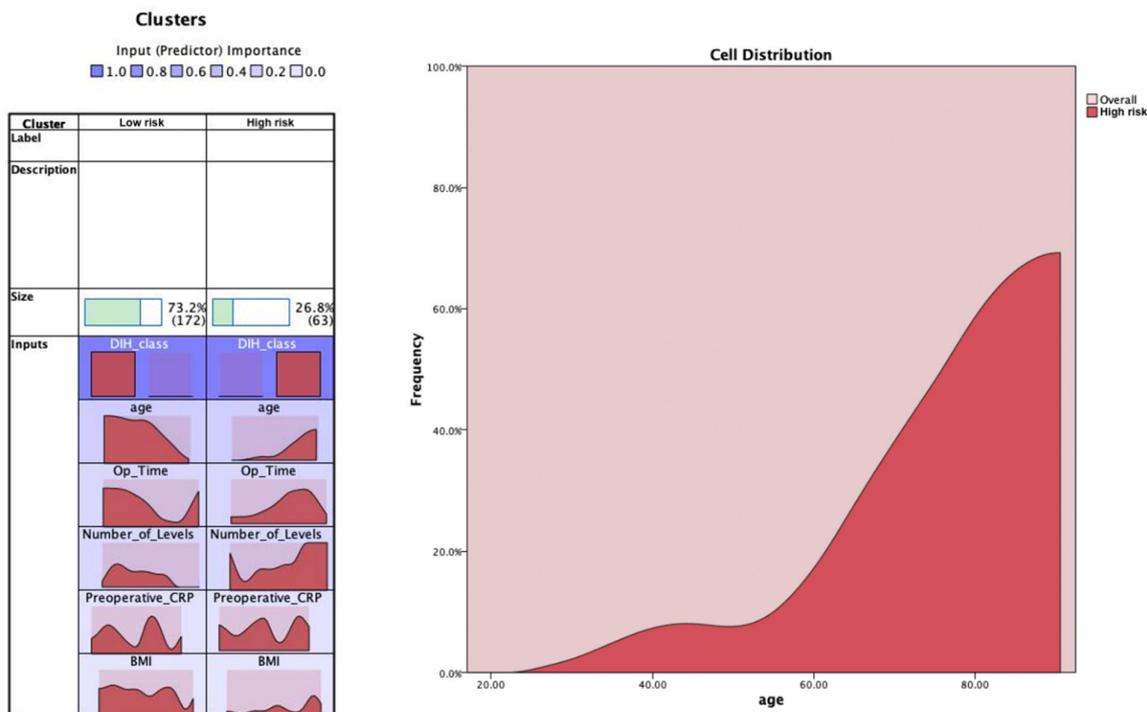


Figure 2. K-means cluster analysis for the binary hospital length of stay class (normal versus prolonged). The most important feature to classify the data was the hospital length of stay class (DIH_class), followed by age, operation time, number of levels, preoperative CRP, and BMI. The distribution charts show the distribution of the features for both clusters. An example feature (age) is shown to help interpretation. The selected feature shows that age distribution is right-shifted for the high risk cluster (including the prolonged hospital stay class), whereas it is left-shifted for the low risk cluster.

Table 2. Showing the descriptive statistics and pairwise comparison of the clusters obtained through the K-means cluster analysis, including the binary class feature of hospital length of stay. **** $p < 0.0001$.

	Cluster	BMI		Number_of_Levels		Op_Time		Preoperative_CRP		Age		Frequencies	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	Normal	Prolonged
	High risk	28.31	5.649	1.62	1.022	104.92	58.253	23.72	51.250	73.2094	12.37884	0	63
	Low risk	27.53	5.531	1.14	0.382	65.58	43.776	10.11	32.706	56.6354	15.86603	172	0
	Combined	27.74	5.562	1.27	0.656	76.13	51.034	13.76	38.912	61.0787	16.69396	172	63
	<i>p</i> -value	****	****	****	****	****	****	****	****	****	****	****	****

We applied various machine learning and deep learning algorithms to predict the LOS classes in the next step. The AUC for the multilayer perceptron model reached the highest AUC with 0.873 (Figure 3). The feature importance analysis revealed that operation time was the most important feature for the classification task, followed by BMI, preoperative CRP, and age (Figure 4). 0: normal LOS; 1: prolonged LOS.

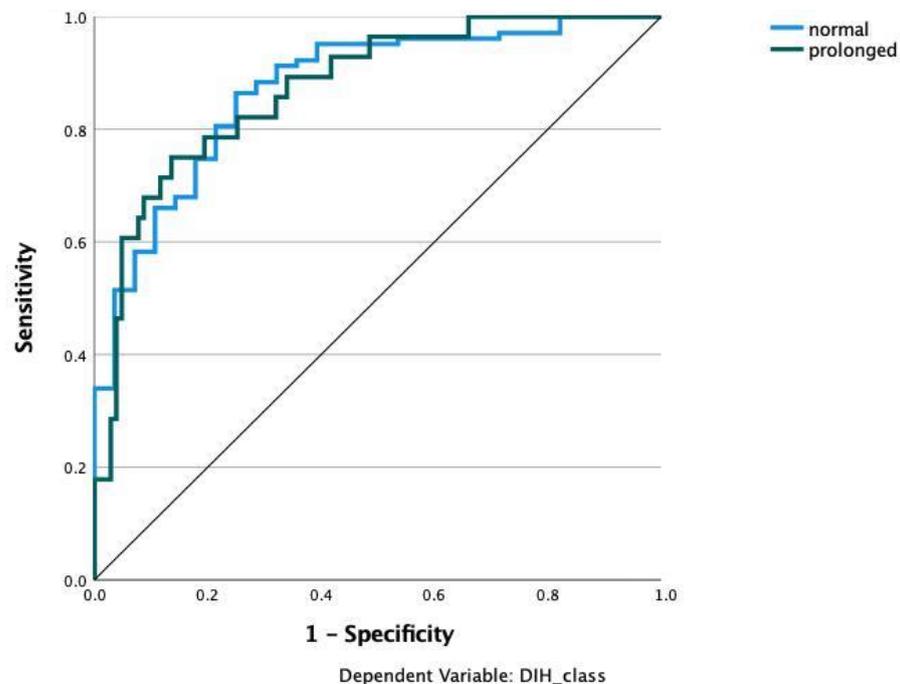


Figure 3. Prediction of hospital stay with multilayer perceptron (MLP). Input layer: feature variables (22 units). Hidden Layer: 2 units, activation function: hyperbolic tangent. Output layer: dependent variable Hospital stay (2 units), activation function: softmax, error function cross-entropy. The number of units in the hidden layer was determined by the testing data criterion: the best number of hidden units is the one that yields the smallest error in the testing dataset. Train/Test/Validation split: 80/10/10. Percent incorrect predictions on training set: 14.5%; percent incorrect predictions on testing set: 14.3%; percent incorrect predictions on holdout set: 10.0%; 0 refers to the prediction of the normal class; 1 refers to the prediction of the prolonged LOS class. AUC: 0.873.

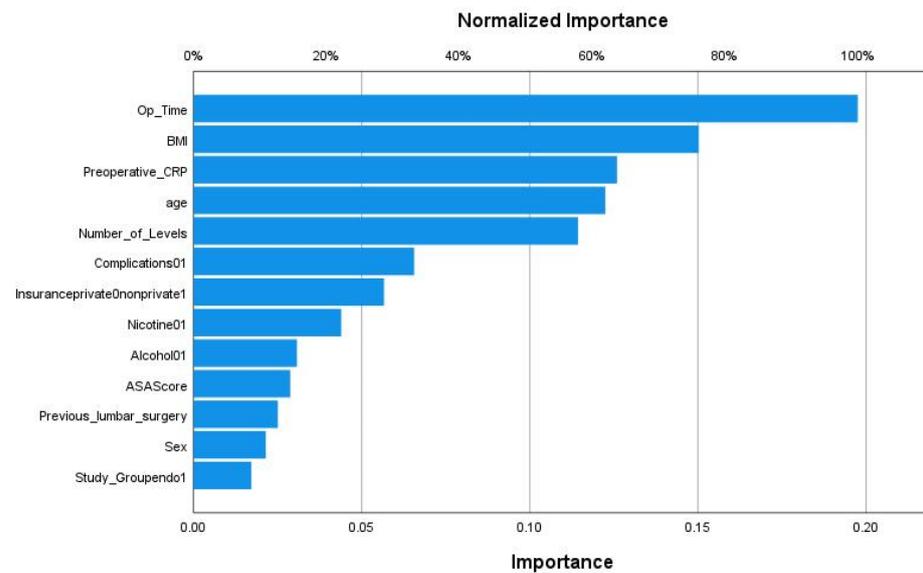


Figure 4. Feature importance analysis for predicting the hospital length of stay classes in the multi-layer perceptron model.

The AUC for the radial basis function neural network model (RBNN) reached an AUC of 0.847 (Figure 5). The MLP model’s feature importance analysis revealed that operation time was the most important feature for the classification task, followed by BMI and preoperative CRP. However, the number of levels reached higher importance than age with this model age (Figure 6).

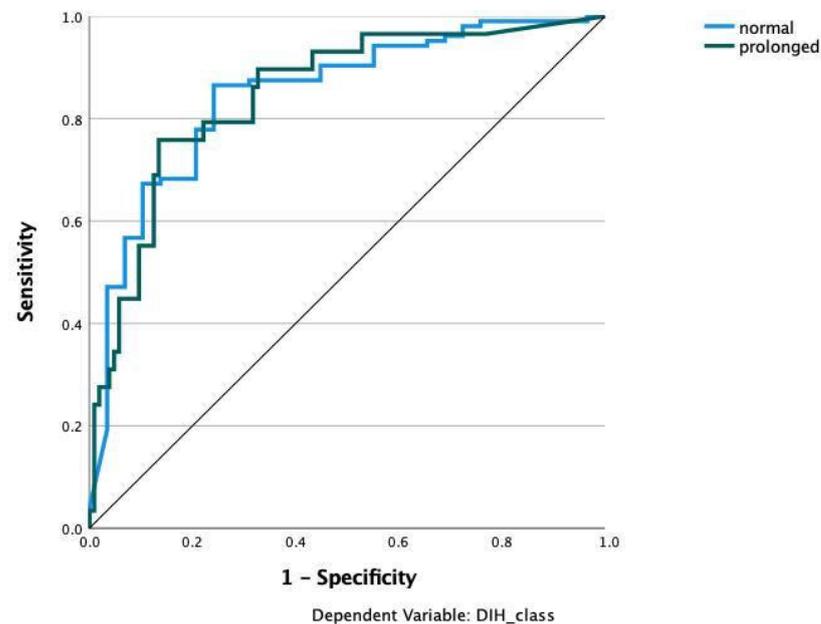


Figure 5. Prediction of hospital stay with radial basis function neural network (RBNN). Input layer: feature variables (22 units). Hidden Layer: 5 units, activation function softmax. Output layer: dependent variable Hospital stay (2 units), activation function: identity, error function sum of squares. The number of units in the hidden layer was determined by the testing data criterion: the best number of hidden units is the one that yields the smallest error in the testing dataset. Train/Test/Validation split: 80/10/10. Percentage incorrect predictions on the training set 19.0%; percent incorrect predictions on testing set: 11.8%; percent incorrect predictions on holdout set: 16.7%. AUC: 0.847.

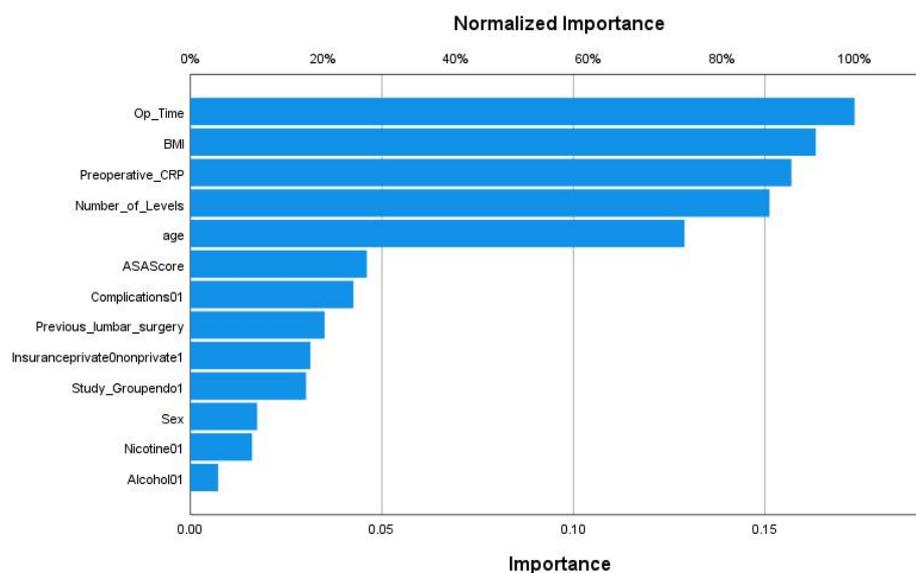


Figure 6. Feature importance analysis for predicting the hospital length of stay classes in the radial basis function neural network model (RBNN).

The accuracies and AUC of the machine learning algorithms and the custom-made convolutional neural network (CNN) is shown in Table 3. The logistic regression reached the highest accuracy, whereas the highest AUC was reached for the custom-made CNN (Table 3).

Table 3. Performance measures for the machine learning and deep learning algorithms to predict the length of hospital stay class (binary classification task). Analysis was done with k-fold cross-validation (k = 5). AUC: area under the curve; Accuracy: (TP + TN)/(TP + TN + FP + FN).

Algorithm	Accuracy	AUC
LogisticRegression	0.814	0.814
Random Forest classifier	0.808	0.826
SGD classifier	0.771	0.804
K-nearest neighbors	0.755	0.769
Decision Trees classifier	0.739	0.675
GaussianNB	0.755	0.799
SVM	0.771	0.794
Custom CNN	0.771	0.862

Finally, we developed a decision tree utilizing the CHAID growing method for simple interpretation in clinics (Figure 7). The algorithm could detect 84.7% of the cases correctly. How to read the CHAID decision tree: for example, an age ≤ 49.829 resulted in 94.4% of normal and only 5.6% of prolonged LOS cases. An absence of complication after this node further results in all cases having normal LOS. In addition, an age between 49.829–67.811 results in 83.1% of cases having normal LOS. Of these cases who underwent endoscopic decompression, all patients had normal LOS. In contrast, an age of > 83.170 results in most cases having prolonged LOS (69.6%). Most of these cases had prolonged LOS with a non-private insurance type.

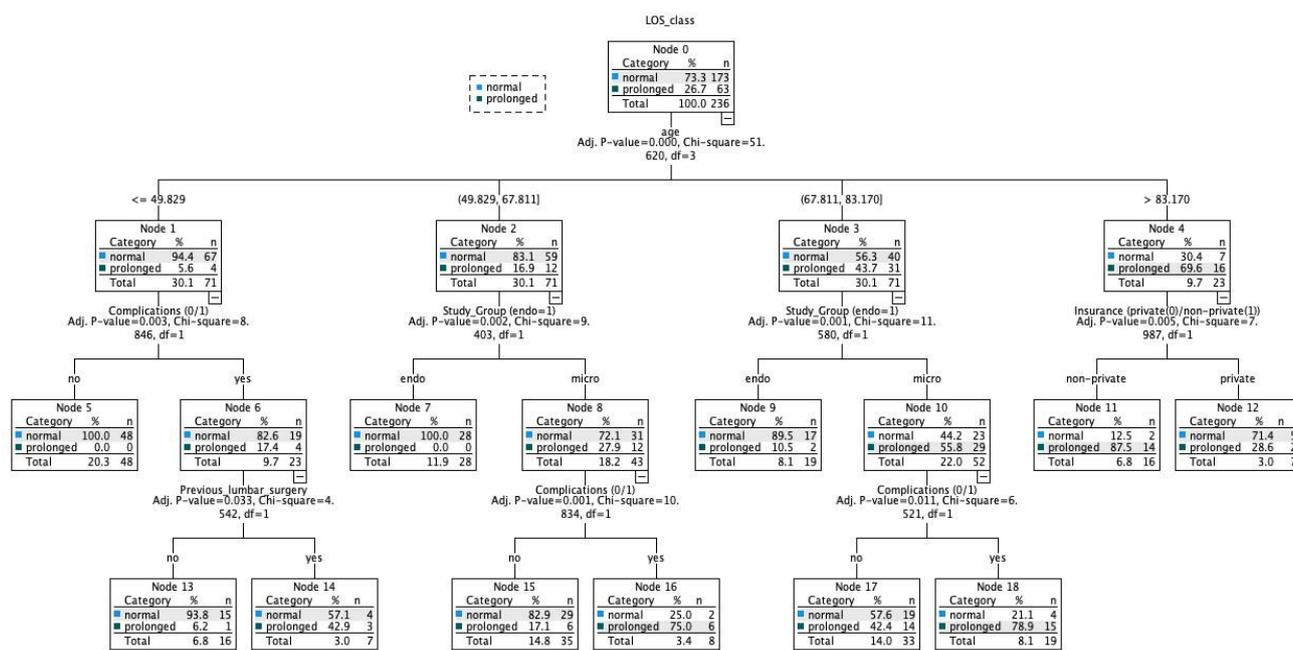


Figure 7. Decision tree with CHAID growing method. Five-fold cross-validation. Percent correct predictions with this algorithm: 84.7%.

4. Discussion

Lower back pain (LBP) is reported to be one of the most serious diseases affecting both health and function, in addition to being among the most expensive to treat [23]. There has been an increase in the use of various treatments over the past 20 years, as well as an increase in the cost of spinal care [6,24–26]. Despite these findings, studies demonstrate that the number of people suffering from disabling back pain has increased over the same period [26,27]. These developments are taking place in a context where health care is increasingly focused on providing value-based treatment to patients. Researchers, providers, and policymakers have an opportunity to design a uniform, value-driven, and digitalized spine care paradigm built on precision medicine. In the absence of studies pertaining to the length of stay, one of the most relevant patient-centered outcome measures, we investigated whether the prolonged length of stay could be predicted using artificial intelligence-based techniques. Our findings suggest that LOS can be predicted using data from lumbar decompression surgery patients. Although we only used a small number of cases, the algorithms showed satisfactory accuracies for the prediction task. As a result, our findings can serve as a foundation for larger multi-center prospective studies that collect data in order to develop more accurate models based on our findings.

Both the patient and the healthcare system can be significantly affected by a longer LOS following lumbar decompression surgery [28,29]. A recent study has shown that longer hospital stays are associated with higher complication rates and higher hospital costs, especially in neurosurgical and orthopedic spine cases. In particular, more prolonged LOS cases were more likely to be readmitted to the hospital. There is, however, the possibility that this may be caused by underlying comorbidities rather than LOS as such [30]. Further, there is evidence that prolonged hospital stay increases the risk of anemia requiring transfusion, impaired mental status, pneumonia, readmission, and hardware issues resulting in reoperation in patients who have undergone spine surgery [31]. One study of neurosurgical patients found that both physical therapy consultations and discharges to a specialist nursing facility were related to 2.4 days and 6.2 days longer length of stay, respectively [32].

Even though several risk factors have been identified for longer LOS following lumbar decompression surgery, no effective systematic model has been proposed to predict

whether a patient will have a short or long LOS following this procedure. We only found one study applying machine learning techniques to predict prolonged LOS after lumbar spinal stenosis surgery [33]. Nevertheless, the authors reported an AUC of 0.54, which is insufficient, considering that an AUC of 0.5 translates to no classification capacity (not much better than a coin toss). In addition to identifying statistically significant clusters in our cohort, which support our hypothesis that prolonged hospital stays are associated with other features, the present study demonstrated that machine learning and deep learning techniques could be used to effectively predict short hospital stays versus long hospital stays. Along with contributing to a deeper understanding of a patient's particular risk profile, these findings can assist in better resource allocation and discharge planning. The variables used in the machine learning method can be used to create an open-source web-based LOS prediction tool. Such a procedure would facilitate external validation of the provided models.

This study found that all algorithms in this study were able to predict and discriminate between classes with AUC values of 0.675–0.873. In this study, the findings compared favorably with those from other studies examining machine learning in orthopedic subspecialties such as spine surgery. Prior research utilizing machine learning has used tree-based methods or neural networks to predict intraoperative blood loss, prolonged hospital stay, patient-reported outcomes, and discharge dispositions in the field with similar or inferior results [15,34–38]. As the current study builds on previous work in this area, the results from the current study add to the growing body of evidence supporting the use of machine learning in orthopedic surgery. Also, according to Kobayashi et al., higher ASA class and longer operating time were associated with a prolonged LOS after spinal surgery, although our findings suggest that operation time contributes more significantly to the prolonged LOS [39]. According to Adogwa et al., surgeon practice style and preferences appear to be risk factors for prolonged LOS, which is consistent with our present finding that operating time is correlated with an increase in LOS [40]. It is important to acknowledge that operating time cannot be attributed solely to surgeon preference or style, as patient comorbidities may also contribute to longer surgical duration [41]. It is also important to consider the surgeon's learning curves when assessing surgical prediction models. Overall, the operation time can be associated with multiple parameters which might also be of importance for LOS. The degree of invasiveness of spinal procedures is generally correlated with the outcome of the operation (blood loss, operation time, and complication rate). As such, the reduction of feature importance into a few parameters may be viewed as problematic as long as not all other potentially important features are included in an assessment, which is simply not feasible in a retrospective study design.

This risk assessment tool (RAPT) consists of six questions that have been validated for predicting outcomes of patients undergoing joint replacement surgery. In addition, it assists patients and physicians in identifying hurdles to discharge and can simplify the discharge process [42]. Likewise, the findings of this study could be incorporated into an equivalent tool. Quantifying the risks associated with surgery would be very helpful to both patients and physicians before surgery. In addition to more accurately setting patient expectations, this will allow the entire patient care team to develop an individualized care plan that is intended to reduce the length of stay while increasing patient safety and satisfaction.

Our results also indicate that preoperative CRP levels might be relevant for assessing the hospital length of stay in future prediction models. CRP is an acute-phase inflammatory cytokine that is released by the liver in response to interleukin-6 and other inflammatory factors [43]. It binds to phospholipids on bacterial surfaces, damaged or apoptotic eukaryotic cells, and nuclear debris in cells. When bound to Fc-gamma receptors, CRP activates the classical complement cascade and promotes phagocytosis [44]. CRP is a well-recognized marker of systemic inflammation. It is also used to predict cardiovascular events in nonsurgical populations based on different CRP risk cut-offs [45]. An increase in CRP levels may be caused by infection, inflammation, trauma, malignancy, or tissue infarction. In addition, an increase in CRP may be observed earlier in the course of a

disease process than other non-specific markers (e.g., fever), and it may fall rapidly during the course of the recovery process [44]. CRP may be helpful as a screening test to detect inflammation early in the course of a disease process, as well as a monitoring tool to assess the effectiveness of treatment. There is evidence that CRP rises as a consequence of surgical trauma, and peaks 48 h after surgery [46,47]. In some patients, the CRP response may only be incomplete or may not occur at all [47]. Validated CRP risk categories were able to predict future cardiovascular events in apparently healthy individuals [48]. In previous studies of cardiac surgery patients, higher preoperative levels of CRP were associated with a greater risk of short- and long-term morbidity and mortality [49–51]. Further, a link has been reported between the peak postoperative CRP response and the degree of surgical trauma [52]. Although not all studies agree, minimal-invasive surgical procedures generally result in lower CRP levels than open procedures [53–55]. It appears that patients who have a high preoperative CRP are more likely to have a higher and later peak postoperative CRP than those who have a normal CRP [56]. The severity of inflammation seems to be not related to surgical operation grade only, suggesting that it may be due to both inflammatory co-morbidity and surgical trauma [57]. Overall, evidence regarding the impact of preoperative CRP in spinal surgery patients is limited. Although it might be intuitive that increased preoperative inflammatory states might affect the length of stay, this phenomenon needs to be validated in future prospective studies. In our algorithms, the inclusion of CRP as a laboratory marker led to an increase in diagnostic accuracy. Thus, it is recommended to include laboratory markers in addition to other data types (clinical data, imaging, genetics, histology) in future prediction models to evaluate whether the models' diagnostic capabilities might improve. Further, this approach is helpful to determine the most impactful features in the dataset [15].

Despite the strength of the algorithms presented, the novelty of the approach, and the promising predictive results, the study does have some potential limitations. The algorithms must be based on a sample that is representative of the population of patients receiving spinal decompression surgery in order to be therapeutically beneficial. Because our data comes from a single, small study, its applicability to other institutions may be limited. As a result, external validation of the models given is required. Furthermore, data were collected retrospectively, lowering the evidence grade because the data may not be as reliable as data collected prospectively. We were also unable to increase the selection of variables to include other possibly important variables that could improve the model further due to the retrospective methodology. However, based on what we found in our prior work and a search of the literature [20], we set out to incorporate important variables.

5. Conclusions

Decompression of the lumbar spine is one of the most common procedures performed in spine surgery. The associated costs are highly dependent on the hospital length of stay. Patient-centered outcome prediction models to predict the length of stay can allow using the available societal healthcare resources effectively. This enables policymakers and providers to compare treatment strategies among different disciplines and to identify the relative priorities for optimal resource allocation among various interventions. Our results indicate that machine learning and deep learning techniques can effectively predict whether patients will have a prolonged length of stay. Implementing the provided algorithms into open-source software and external validation through large-scale prospective studies are warranted to introduce the provided prediction tools in clinics.

Author Contributions: Conceptualization: B.S., S.C.-D., F.H. and G.L.; Data curation: B.S., A.Z. and S.Ü.; Formal analysis: S.C.-D., F.H. and G.L.; Investigation: B.S., A.Z. and S.Ü.; Methodology: B.S., A.Z. and S.Ü.; Resources: S.C.-D., F.H. and G.L.; Validation, S.C.-D., F.H. and G.L.; Visualization: B.S.; Writing—original draft: B.S. and S.Ü.; Writing—review & editing: A.Z., S.C.-D., F.H. and G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee Freiburg, Germany [Number: 116/200]. Written informed consent to participate in observational studies was obtained from each patient.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The raw data are pseudonymized and available from the corresponding author on reasonable request. The python code and machine learning algorithm structures are available from: <https://github.com/Freiburg-AI-Research> (accessed on 16 May 2022).

Acknowledgments: The article processing charge was funded by the Baden-Wuerttemberg Ministry of Science, Research and Art, and the University of Freiburg in the funding program Open Access Publishing. Joimax GmbH provided fellowship support for B.S.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Lurie, J.; Tomkins-Lane, C. Management of Lumbar Spinal Stenosis. *BMJ* **2016**, *352*, h6234. [[CrossRef](#)] [[PubMed](#)]
- Deer, T.; Sayed, D.; Michels, J.; Josephson, Y.; Li, S.; Calodney, A.K. A Review of Lumbar Spinal Stenosis with Intermittent Neurogenic Claudication: Disease and Diagnosis. *Pain Med.* **2019**, *20*, S32–S44. [[CrossRef](#)] [[PubMed](#)]
- Dagenais, S.; Caro, J.; Haldeman, S. A Systematic Review of Low Back Pain Cost of Illness Studies in the United States and Internationally. *Spine J.* **2008**, *8*, 8–20. [[CrossRef](#)]
- Deyo, R.A.; Mirza, S.K.; Martin, B.I.; Kreuter, W.; Goodman, D.C.; Jarvik, J.G. Trends, Major Medical Complications, and Charges Associated with Surgery for Lumbar Spinal Stenosis in Older Adults. *JAMA* **2010**, *303*, 1259–1265. [[CrossRef](#)]
- Raad, M.; Reidler, J.S.; El Dafrawy, M.H.; Amin, R.M.; Jain, A.; Neuman, B.J.; Riley, L.H.; Sciubba, D.M.; Kebaish, K.M.; Skolasky, R.L. US Regional Variations in Rates, Outcomes, and Costs of Spinal Arthrodesis for Lumbar Spinal Stenosis in Working Adults Aged 40–65 Years. *J. Neurosurg. Spine* **2018**, *30*, 83–90. [[CrossRef](#)]
- Bae, H.W.; Rajee, S.S.; Kanim, L.E. Nationwide Trends in the Surgical Management of Lumbar Spinal Stenosis. *Spine* **2013**, *38*, 916–926. [[CrossRef](#)]
- Luo, X.; Pietrobon, R.; Sun, S.X.; Liu, G.G.; Hey, L. Estimates and Patterns of Direct Health Care Expenditures among Individuals with Back Pain in the United States. *Spine* **1976**, *29*, 86. [[CrossRef](#)] [[PubMed](#)]
- Emanuel, E.J.; Fuchs, V.R. The Perfect Storm of Overutilization. *JAMA* **2008**, *299*, 2789–2791. [[CrossRef](#)] [[PubMed](#)]
- Modhia, U.; Takemoto, S.; Braid-Forbes, M.J.; Weber, M.; Berven, S.H. Readmission Rates after Decompression Surgery in Patients with Lumbar Spinal Stenosis among Medicare Beneficiaries. *Spine* **2013**, *38*, 591–596. [[CrossRef](#)]
- Martin, B.I.; Mirza, S.K.; Flum, D.R.; Wickizer, T.M.; Heagerty, P.J.; Lenkoski, A.F.; Deyo, R.A. Repeat Surgery after Lumbar Decompression for Herniated Disc: The Quality Implications of Hospital and Surgeon Variation. *Spine J.* **2012**, *12*, 89–97. [[CrossRef](#)]
- Yeom, J.S.; Buchowski, J.M.; Shen, H.X.; Liu, G.; Bunmaprasert, T.; Riew, K.D. Effect of Fibrin Sealant on Drain Output and Duration of Hospitalization after Multilevel Anterior Cervical Fusion: A Retrospective Matched Pair Analysis. *Spine* **2008**, *33*, E543–E547. [[CrossRef](#)] [[PubMed](#)]
- Epstein, N.E.; Schwall, G.; Reilly, T.; Insinna, T.; Bahnken, A.; Hood, D.C. Surgeon Choices, and the Choice of Surgeons, Affect Total Hospital Charges for Single-Level Anterior Cervical Surgery. *Spine* **2011**, *36*, 905–909. [[CrossRef](#)] [[PubMed](#)]
- Weinstein, J.N.; Lurie, J.D.; Olson, P.; Bronner, K.K.; Fisher, E.S.; Morgan, M.T.S. United States Trends and Regional Variations in Lumbar Spine Surgery: 1992–2003. *Spine* **2006**, *31*, 2707. [[CrossRef](#)] [[PubMed](#)]
- Katz, J.N.; Lipson, S.J.; Lew, R.A.; Grobler, L.J.; Weinstein, J.N.; Brick, G.W.; Fossel, A.H.; Liang, M.H. Lumbar Laminectomy Alone or with Instrumented or Noninstrumented Arthrodesis in Degenerative Lumbar Spinal Stenosis: Patient Selection, Costs, and Surgical Outcomes. *Spine* **1997**, *22*, 1123–1131. [[CrossRef](#)] [[PubMed](#)]
- Phan, K.; Kim, J.; Di Capua, J.; Lee, N.J.; Kothari, P.; Dowdell, J.; Overley, S.C.; Guzman, J.Z.; Cho, S.K. Impact of Operation Time on 30-Day Complications after Adult Spinal Deformity Surgery. *Glob. Spine J.* **2017**, *7*, 664–671. [[CrossRef](#)]
- Klineberg, E.O.; Passias, P.G.; Jalai, C.M.; Worley, N.; Sciubba, D.M.; Burton, D.C.; Gupta, M.C.; Soroceanu, A.; Zebala, L.P.; Mundis, G.M. Predicting Extended Length of Hospital Stay in an Adult Spinal Deformity Surgical Population. *Spine* **2016**, *41*, 798–805. [[CrossRef](#)]
- Saravi, B.; Lang, G.; Ülkümen, S.; Südkamp, N.; Hassel, F. Case-Matched Radiological and Clinical Outcome Evaluation of Interlaminar versus Microsurgical Decompression of Lumbar Spinal Stenosis. In Proceedings of the German Congress of Orthopaedics and Traumatology, Berlin, Germany, 26–29 October 2021. [[CrossRef](#)]
- Saravi, B.; Hassel, F.; Ülkümen, S.; Zink, A.; Shavlokhova, V.; Couillard-Despres, S.; Boeker, M.; Obid, P.; Lang, G.M. Artificial Intelligence-Driven Prediction Modeling and Decision Making in Spine Surgery Using Hybrid Machine Learning Models. *J. Pers. Med.* **2022**, *12*, 509. [[CrossRef](#)]
- Gellman, D.D. Cost-Benefit in Health Care: We Need to Know Much More. *Can. Med. Assoc. J.* **1974**, *111*, 988–989.

20. Dagenais, S.; Roffey, D.M.; Wai, E.K.; Haldeman, S.; Caro, J. Can Cost Utility Evaluations Inform Decision Making about Interventions for Low Back Pain? *Spine J.* **2009**, *9*, 944–957. [[CrossRef](#)]
21. Krell, R.W.; Girotti, M.; Dimick, J.B. Extended Hospital Stay after Surgery: A Marker of Hospital Quality or Efficiency? *J. Surg. Res.* **2013**, *179*, 219. [[CrossRef](#)]
22. Bottle, A.; Middleton, S.; Kalkman, C.J.; Livingston, E.H.; Aylin, P. Global Comparators Project: International Comparison of Hospital Outcomes Using Administrative Data. *Health Serv. Res.* **2013**, *48*, 2081–2100. [[CrossRef](#)] [[PubMed](#)]
23. Hoy, D.; March, L.; Brooks, P.; Blyth, F.; Woolf, A.; Bain, C.; Williams, G.; Smith, E.; Vos, T.; Barendregt, J.; et al. The Global Burden of Low Back Pain: Estimates from the Global Burden of Disease 2010 Study. *Ann. Rheum. Dis.* **2014**, *73*, 968–974. [[CrossRef](#)]
24. Deyo, R.A.; Mirza, S.K.; Turner, J.A.; Martin, B.A. Overtreating Chronic Back Pain: Time to Back Off? *J. Am. Board Fam. Med.* **2009**, *22*, 62–68. [[CrossRef](#)]
25. Friedly, J.; Standaert, C.; Chan, L. Epidemiology of Spine Care: The Back Pain Dilemma. *Phys. Med. Rehabil. Clin. N. Am.* **2010**, *21*, 659–677. [[CrossRef](#)] [[PubMed](#)]
26. Martin, B.I.; Turner, J.A.; Mirza, S.K.; Lee, M.J.; Comstock, B.A.; Deyo, R.A. Trends in Health Care Expenditures, Utilization, and Health Status among US Adults with Spine Problems, 1997–2006. *Spine* **2009**, *34*, 2077–2084. [[CrossRef](#)] [[PubMed](#)]
27. Freburger, J.K.; Holmes, G.M.; Agans, R. The Rising Prevalence of Chronic Low Back Pain. *Arch. Intern. Med.* **2009**, *169*, 251–258. [[CrossRef](#)]
28. McCarthy, I.M.; Hostin, R.A.; O'Brien, M.F.; Fleming, N.S.; Ogola, G.; Kudyakov, R.; Richter, K.M.; Saigal, R.; Berven, S.H.; Ames, C.P. Analysis of the Direct Cost of Surgery for Four Diagnostic Categories of Adult Spinal Deformity. *Spine J.* **2013**, *13*, 1843–1848. [[CrossRef](#)]
29. Shields, L.B.; Clark, L.; Glassman, S.D.; Shields, C.B. Decreasing Hospital Length of Stay Following Lumbar Fusion Utilizing Multidisciplinary Committee Meetings Involving Surgeons and Other Caretakers. *Surg. Neurol. Int.* **2017**, *8*, 5. [[CrossRef](#)]
30. Ansari, S.F.; Yan, H.; Zou, J.; Worth, R.M.; Barbaro, N.M. Hospital Length of Stay and Readmission Rate for Neurosurgical Patients. *Neurosurgery* **2018**, *82*, 173–181. [[CrossRef](#)]
31. Gruskay, J.A.; Fu, M.; Bohl, D.D.; Webb, M.L.; Grauer, J.N. Factors Affecting Length of Stay after Elective Posterior Lumbar Spine Surgery: A Multivariate Analysis. *Spine J.* **2015**, *15*, 1188–1195. [[CrossRef](#)]
32. Linzey, J.R.; Kahn, E.N.; Shlykov, M.A.; Johnson, K.T.; Sullivan, K.; Pandey, A.S. Length of Stay Beyond Medical Readiness in Neurosurgical Patients: A Prospective Analysis. *Neurosurgery* **2019**, *85*, 60–65. [[CrossRef](#)] [[PubMed](#)]
33. Siccoli, A.; de Wispeleere, M.P.; Schröder, M.L. Machine Learning–Based Preoperative Predictive Analytics for Lumbar Spinal Stenosis. *Neurosurg. Focus* **2019**, *46*, 5. [[CrossRef](#)] [[PubMed](#)]
34. Biron, D.R.; Sinha, I.; Kleiner, J.E. A Novel Machine Learning Model Developed to Assist in Patient Selection for Outpatient Total Shoulder Arthroplasty. *J. Am. Acad. Orthop. Surg.* **2019**, *28*, 580–585. [[CrossRef](#)] [[PubMed](#)]
35. Navarro, S.M.; Wang, E.Y.; Haeberle, H.S. Machine Learning and Primary Total Knee Arthroplasty: Patient Forecasting for a Patient-Specific Payment Model. *J. Arthroplasty* **2018**, *33*, 3617–3623. [[CrossRef](#)] [[PubMed](#)]
36. Durand, W.M.; DePasse, J.M.; Daniels, A.H. Predictive Modeling for Blood Transfusion after Adult Spinal Deformity Surgery: A Tree-Based Machine Learning Approach. *Spine* **2018**, *43*, 1058–1066. [[CrossRef](#)] [[PubMed](#)]
37. Fontana, M.A.; Lyman, S.; Sarker, G.K.; Padgett, D.E.; MacLean, C.H. Can Machine Learning Algorithms Predict Which Patients Will Achieve Minimally Clinically Important Differences from Total Joint Arthroplasty? *Clin. Orthop. Relat. Res. Lippincott Williams Wilkins* **2019**, *477*, 1267–1279. [[CrossRef](#)]
38. Malik, A.T.; Khan, S.N. Predictive Modeling in Spine Surgery. *Ann. Transl. Med.* **2019**, *7*, 173. [[CrossRef](#)]
39. Kobayashi, K.; Ando, K.; Kato, F.; Kanemura, T.; Sato, K.; Hachiya, Y.; Matsubara, Y.; Kamiya, M.; Sakai, Y.; Yagi, H. Predictors of Prolonged Length of Stay after Lumbar Interbody Fusion: A Multicenter Study. *Glob. Spine J.* **2019**, *9*, 466–472. [[CrossRef](#)]
40. Adogwa, O.; Lilly, D.T.; Khalid, S.; Desai, S.A.; Vuong, V.D.; Davison, M.A.; Ouyang, B.; Bagley, C.A.; Cheng, J. Extended Length of Stay after Lumbar Spine Surgery: Sick Patients, Postoperative Complications, or Practice Style Differences among Hospitals and Physicians? *World Neurosurg.* **2019**, *123*, 734–739. [[CrossRef](#)] [[PubMed](#)]
41. Kim, B.D.; Hsu, W.K.; De Oliveira, G.S.; Saha, S.; Kim, J.Y.S. Operative Duration as an Independent Risk Factor for Postoperative Complications in Single-Level Lumbar Fusion: An Analysis of 4588 Surgical Cases. *Spine* **2014**, *39*, 510–520. [[CrossRef](#)]
42. Dibra, F.F.; Silverberg, A.J.; Vasilopoulos, T.; Gray, C.F.; Parvataneni, H.K.; Prieto, H.A. Arthroplasty Care Redesign Impacts the Predictive Accuracy of the Risk Assessment and Prediction Tool. *J. Arthroplast.* **2019**, *34*, 2549–2554. [[CrossRef](#)] [[PubMed](#)]
43. Vigushin, D.M.; Pepys, M.B.; Hawkins, P.N. Metabolic and Scintigraphic Studies of Radioiodinated Human C-Reactive Protein in Health and Disease. *J. Clin. Investig.* **1993**, *91*, 1351–1357. [[CrossRef](#)]
44. Pepys, M.B.; Hirschfield, G.M. C-Reactive Protein: A Critical Update. *J. Clin. Investig.* **2003**, *111*, 1805–1812. [[CrossRef](#)] [[PubMed](#)]
45. Ridker, P.M. Clinical Application of C-Reactive Protein for Cardiovascular Disease Detection and Prevention. *Circulation* **2003**, *107*, 363–369. [[CrossRef](#)] [[PubMed](#)]
46. Colley, C.M.; Fleck, A.; Goode, A.W.; Muller, B.R.; Myers, M.A. Early Time Course of the Acute Phase Protein Response in Man. *J. Clin. Pathol.* **1983**, *36*, 203–207. [[CrossRef](#)]
47. White, J.; Kelly, M.; Dunsmuir, R. C-Reactive Protein Level after Total Hip and Total Knee Replacement. *J. Bone Jt. Surg. Br.* **1998**, *80*, 909–911. [[CrossRef](#)]

48. Perry, T.E.; Muehlschlegel, J.D.; Liu, K.-Y.; Fox, A.A.; Collard, C.D.; Body, S.C.; Shernan, S.K. CABG Genomics Investigators Preoperative C-Reactive Protein Predicts Long-Term Mortality and Hospital Length of Stay after Primary, Nonemergent Coronary Artery Bypass Grafting. *Anesthesiology* **2010**, *112*, 607–613. [[CrossRef](#)]
49. Nielsen, H.J.; Christensen, I.J.; Sørensen, S.; Moesgaard, F.; Brünner, N. Preoperative Plasma Plasminogen Activator Inhibitor Type-1 and Serum C-Reactive Protein Levels in Patients with Colorectal Cancer. The RANX05 Colorectal Cancer Study Group. *Ann. Surg. Oncol.* **2000**, *7*, 617–623. [[CrossRef](#)]
50. Nozoe, T.; Matsumata, T.; Kitamura, M.; Sugimachi, K. Significance of Preoperative Elevation of Serum C-Reactive Protein as an Indicator for Prognosis in Colorectal Cancer. *Am. J. Surg.* **1998**, *176*, 335–338. [[CrossRef](#)]
51. Fransen, E.J.; Maessen, J.G.; Elenbaas, T.W.; van Aarnhem, E.E.; van Dieijen-Visser, M.P. Enhanced Preoperative C-Reactive Protein Plasma Levels as a Risk Factor for Postoperative Infections after Cardiac Surgery. *Ann. Thorac. Surg.* **1999**, *67*, 134–138. [[CrossRef](#)]
52. Brewster, N.; Guthrie, C.; McBirnie, J. CRP Levels as a Measure of Surgical Trauma: A Comparison of Different General Surgical Procedures. *J. R. Coll. Surg. Edinb.* **1994**, *39*, 86–88. [[PubMed](#)]
53. Grande, M.; Tucci, G.F.; Adorisio, O.; Barini, A.; Rulli, F.; Neri, A.; Franchi, F.; Farinon, A.M. Systemic Acute-Phase Response after Laparoscopic and Open Cholecystectomy. *Surg. Endosc.* **2002**, *16*, 313–316. [[CrossRef](#)] [[PubMed](#)]
54. Hildebrandt, U.; Kessler, K.; Plusczyk, T.; Pistorius, G.; Vollmar, B.; Menger, M.D. Comparison of Surgical Stress between Laparoscopic and Open Colonic Resections. *Surg. Endosc.* **2003**, *17*, 242–246. [[CrossRef](#)] [[PubMed](#)]
55. Saravi, B.; Ülkümen, S.; Couillard-Despres, S.; Lang, G.; Hassel, F. One-Year Clinical Outcomes of Minimal-Invasive Dorsal Percutaneous Fixation of Thoracolumbar Spine Fractures. *Medicina* **2022**, *58*, 606. [[CrossRef](#)]
56. Cappabianca, G.; Paparella, D.; Visicchio, G.; Capone, G.; Lionetti, G.; Numis, F.; Ferrara, P.; D’Agostino, C.; de Luca Tupputi Schinosa, L. Preoperative C-Reactive Protein Predicts Mid-Term Outcome after Cardiac Surgery. *Ann. Thorac. Surg.* **2006**, *82*, 2170–2178. [[CrossRef](#)] [[PubMed](#)]
57. Cole, D.S.; Watts, A.; Scott-Coombes, D.; Avades, T. Clinical Utility of Peri-Operative C-Reactive Protein Testing in General Surgery. *Ann. R. Coll. Surg. Engl.* **2008**, *90*, 317–321. [[CrossRef](#)] [[PubMed](#)]