Feasibility and Assessment of a Machine Learning-Based Predictive Model of Outcome After Lumbar Decompression Surgery

Arthur André, MD, MSc[1,2,3] ⓘ, Bruno Peyrou, PhD[3],
Alexandre Carpentier, MD, PhD[2],
and Jean-Jacques Vignaux, MSc[3] ⓘ

Abstract

Study design: Retrospective study at a unique center.

Objective: The aim of this study is twofold, to develop a virtual patients model for lumbar decompression surgery and to evaluate the precision of an artificial neural network (ANN) model designed to accurately predict the clinical outcomes of lumbar decompression surgery.

Methods: We performed a retrospective study of complete Electronic Health Records (EHR) to identify potential unfavorable criteria for spine surgery (predictors). A cohort of synthetics EHR was created to classify patients by surgical success (green zone) or partial failure (orange zone) using an Artificial Neural Network which screens all the available predictors.

Results: In the actual cohort, we included 60 patients, with complete EHR allowing efficient analysis, 26 patients were in the orange zone (43.4%) and 34 were in the green zone (56.6%). The average positive criteria amount for actual patients was 8.62 for the green zone (SD+/- 3.09) and 10.92 for the orange zone (SD 3.38). The classifier (a neural network) was trained using 10,000 virtual patients and 2000 virtual patients were used for test purposes. The 12,000 virtual patients were generated from the 60 EHR, of which half were in the green zone and half in the orange zone. The model showed an accuracy of 72% and a ROC score of 0.78. The sensitivity was 0.885 and the specificity 0.59.

Conclusion: Our method can be used to predict a favorable patient to have lumbar decompression surgery. However, there is still a need to further develop its ability to analyze patients in the "failure of treatment" zone to offer precise management of patient health before spinal surgery.

Keywords

machine learning, lumbar decompression surgery, retrospective study, synthetic electronic medical record, ROC curve

Abbreviations

ACC, accuracy; ACS-NSQIP, American College of Surgeons National Surgical Quality Improvement Program; ANN, artificial neural networks; AUC, area under the receiver operating characteristic curve; COPD, chronic obstructive pulmonary disease; DNN, deep neural networks; HER, electronic health records; GBM, gradient boosting machine; GLM, generalized linear model; GLMnet, elastic-net GLM; LSS, lumbar spinal stenosis; MCID, minimum clinically important difference; ML, machine learning; NPV, negative predictive value; NRS, numeric rating scale; NRS-BP, NRS for back pain; NRS-LP, NRS for leg pain; ODI, Oswestry Disability Index; PHC, predictive hierarchical clustering; PPV, positive predictive value; PROMs, patient-reported outcome measures; RF, random forest; ROC, receiver operating characteristic

[1] Ramsay santé, Clinique Geoffroy Saint-Hilaire, Paris, France
[2] Neurosurgery Department, Pitié-Salpêtrière University Hospital, Paris, France
[3] Cortexx Medical Intelligence, Paris, France

Corresponding Author:
Arthur André, Cortexx Medical Intelligence, 156 Boulevard, Haussmann 75008, Paris.
Email: arthur.andre@neurochirurgie.fr

Introduction

Lumbar spinal disorders are among the most disabling conditions, particularly in developed countries, due to the increase in sedentary lifestyles and aging populations.[1]

When conservative treatment is insufficient or pharmaceutical options show too many secondary effects (dependency, misuse), surgery is a valid option to relieve pain and improve function.[2-4]

However, patient selection remains very complex and the benefits of surgical interventions sometimes uncertain.[5] Indeed, between 2 and 23% of patients having back surgery will present an adverse event or a complication after surgery.[6,7]

Around 30% to 50% of patients will not be—or only slightly—relieved—by the surgical act, and will maintain their intake of morphine, with the side effects and the costs that this entails[8]

Surgery success is well evaluated by validated indicators such as patient-reported outcomes measures (PROMS).[9] This protocol is based on the standardized collection of patient well-being and health status after a surgical procedure. It is used on large cohorts to study a set of factors participating in clinical outcomes after surgical treatment (see Table 1.).

Most of these studies are based on the analysis of electronic medical records (EHR) in single-institution or in large national Database, describing statistically relevant risk factors of adverse event or surgery failure on a population.[5,55] There is a growing interest about predictive factors influencing individual response after surgery, especially in terms of individual PROM. Furthermore, some promising predictive models in disk herniation recurrence or fusion[50,56,57] exist but there is a lack of practical models for lumbar spine decompression in general.

"4P" (predictive, preventive, personalized and participative) medicine benefits from the support of artificial intelligence[58] (AI) machine learning and synthetic patient models.[59,60] Regarding spine surgery, tools are already capable of improving the quality of the spine diagnosis.[61]

Some algorithms allow to determine the average duration of sick leave,[62] the risks of opioids dependence for prolonged periods post-operatively[63] and to predict postoperative adverse events up to 30 days after spinal surgery[64-66] (see Table 2.).

Among these machine learning methods, we found multivariate logistic regression, stochastic gradient boosting or support vector machine methods and recently artificial neural networks and their improvement in deep neural networks[60,77] to support decision-making activities.

Despite the current focus using EHR as the standard for development of machine learning algorithms, it can be very difficult to gather all the data needed to train such models. Likewise, for technical reasons (interoperability, data exchange, and ability of the operator to use information technologies) or legal and ethical issues,[82] it is difficult to access the full records in academic and industrial research.

The generation of synthetic patients from the exploitation of EHR solves many problems related to the processing of real patients data.[83] Therefore data-driven methods were developed based on synthetic EHR[84] in 3 different ways: using synthetic health data records to help overcome confidentiality issues,[62,85] modeling disease progression and interventions for prospective analysis of large scale virtual cohorts[86]; and completing EHR data for imbalanced cohorts (cf. Table 3).

Objective

The aim of this study is twofold, to develop a virtual patients model for lumbar decompression surgery and to evaluate the precision of an artificial neural network (ANN) model designed to accurately predict the clinical outcomes of lumbar decompression surgery.

Materials and Methods

A transparent reporting of a multivariable prediction model for individual prognosis was used for reporting our model of machine learning in Biomedical Research.

Institutional Review Board

The EHR screening was approved by the department review board from the Department of Neurosurgery, Pitié-Salpêtrière University Hospital, all other data was anonymously reported and there is no specific approval.

Population

Any patient who underwent lumbar decompression surgery from January 2019 to April 2019 in the Department of Neurosurgery, Pitié-Salpêtrière University Hospital was included. We exploited retrospectively the local EHR.

Data Collection

Data collection was carried out through the automated request of EHR patients from our center (Orbis, Agfa Healthcare).

Pre-operative criteria were collected, including the patient's age, sex, body mass index (BMI), demographic, radiological criteria, as well as the presence of comorbidities (diabetes, sleep apnea syndrome, kidney disease.), the type of work and the duration of sick leave, socio-professional problems, psychological disorders (anxiety or depressive syndrome) drugs consumption (NSAIDs, opioids), and immediate post-operative criteria such as: radiological criteria, sleep or food improvement, return to work, or rehabilitation inpatients center.

Patients were classified into 3 categories according to their surgery outcome: Green (significant improvement of pain and function without level 2 or 3 analgesics or other symptom) Orange (no significant improvement and/or significant medication intake anxiety-depression and/or persistent lumbar pain) and Red (early adverse event or complication)

Table 1. Predictors.

Author	Year	Significant predictor	Positive predictive factor	Negative predictive factor	Area
Katz et al[10]	1999	Low cardiovascular comorbidity	*		GREEN ZONE
Hägg et al[11]	2003	Severe disc degeneration, Neuroticism, Pre-operative sick leave		*	ORANGE ZONE
Kohlboeck et al[12]	2004	Straight leg raise test, Depression, Sensory pain		*	ORANGE ZONE
Trief et al[13]	2006	Better emotional health	*		GREEN ZONE
Slover et al[14]	2006	Active compensation case, Self-rated poor health, Smoking, Headaches, Depression, Nervous system disorders		*	ORANGE ZONE
Braybrooke et al[15]	2007	Time to surgery		*	ORANGE ZONE
Mannion et al[16]	2007	Pain duration, Re-operations, Multilevel surgery, Depression, FABQ Score		*	ORANGE ZONE
Park et al[17]	2008	Minimally invasive surgery	*		GREEN ZONE
Park et al[17]	2008	Age, BMI > 25, Hypertension, Coronary artery diseases, Diabetes		*	RED ZONE
Garcia et al[18]	2008	Weight reduction program	*		GREEN ZONE
Vaidya et al[19]	2009	Obesity, Multiple level fusions		*	RED ZONE
Chen et al[20]	2009	Diabetes		*	RED ZONE
Abbott et al[21]	2011	Catastrophizing, Pain intensity, Bad expectations		*	ORANGE ZONE
Senker et al[22]	2011	Minimally invasive surgery	*		GREEN ZONE
Chaichana et al[23]	2011	Depression, Decreased perception scale anxiety		*	ORANGE ZONE
Sinikallio et al[24]	2011	Depression		*	ORANGE ZONE
Kalanithi et al[25]	2012	Morbid obesity		*	RED ZONE
Sørlie et al[26]	2012	MODIC type 1 smoking		*	ORANGE ZONE
Hellum et al[28]	2012	Long duration Low back pain high fear avoidance for work, MODIC changes		*	ORANGE ZONE
Gaudelli and Thomas[29]	2012	Instrumented fusion		*	RED ZONE
Mehta et al[30]	2012	Obesity		*	RED ZONE
Sharma et al[31]	2013	Diabetes		*	RED ZONE
Takahashi et al[32]	2013	Diabetes of more than 20 years		*	RED ZONE
Bekelis et al[33]	2014	Age, Extensive operations, Medical deconditioning (weight loss, dialysis, peripheral vascular disease) BMI, Neurologic deficit, Bleeding disorders		*	RED ZONE
Lee et al[34]	2014	Opioid consumption, Modified somatic perception, Depression		*	ORANGE ZONE
Pakarinen et al[27]	2014	Depression		*	ORANGE ZONE
Kim et al[35]	2018	Back pain, Pain sensitivity		*	ORANGE ZONE
Coronado et al[36]	2015	Increased pain sensitivity Increased pain catastrophizing		*	ORANGE ZONE
McGirt et al[37]	2015	Functional score opioid use, Hypertension, Atrial fibrillation, extremity pain, myocardial infarction, Diabetes, Osteoporosis, Smoking		*	ORANGE ZONE
Anderson et al[38]	2015	Chronic opioid therapy, Additional lumbar surgery, depression, work loss		*	ORANGE ZONE
Chotai et al[39]	2015	Insurance status, Functional score, BP/NP Scores		*	ORANGE ZONE
Schöller et al[40]	2016	Re-operation, Duration of pain, Spondylisthesis, Smoking, gender, Age, BMI		*	ORANGE ZONE
Archer et al[41]	2016	Cognitive-behavioral based physical therapy (CBPT)	*		GREEN ZONE
Asher et al[42]	2017	ASA score, disability, education, Unemployment, Insurance status		*	ORANGE ZONE
Mummaneni et al[43]	2017	Open surgery		*	ORANGE ZONE
Crawford et al[44]	2017	Discopathy			ORANGE ZONE
Suri et al[45]	2017	Smoking, Depression		*	ORANGE ZONE
McGirt et al[5]	2017	Education, Employment status, Baseline EQ5D, Fusion		*	ORANGE ZONE
Sharma et al[46]	2018	Prior opioid dependence, Younger age		*	ORANGE ZONE
Dunn et al[47]	2018	Catastrophizing, depression		*	ORANGE ZONE
Chan et al[48]	2018	Symptom duration		*	ORANGE ZONE
O'Donnell et al[49]	2018	Opioid use, Time to surgery, Legal representation, Psychiatric comorbidity		*	ORANGE ZONE
Khor et al[50]	2018	Age, Gender, Ethnic, Insurance Status, ASA Score, functional score		*	ORANGE ZONE
Dobran et al[51]	2019	Age, BMI		*	RED ZONE
Staub et al[52]	2020	Obesity, Re-operation, insurance status		*	ORANGE ZONE
Mauro et al[53]	2020	BMI	*		ORANGE ZONE
Rudolfsen et al[54]	2020	Quality of life score, Functional score	*		GREEN ZONE

Table 2. Predictive Model for Spine Surgery.

Author	Year	Data collection (center)	Number of patients	Classifier used	Prediction / AUC
Azimi et al[67]	2014	Database (single-center)	168	ANN, Logistic regression analysis	2-year surgical satisfaction (AUC 0.80)
Azimi et al[68]	2014	Database (single-center)	203	ANN, Logistic regression analysis	Successful surgery outcome for disk herniation (AUC 0.82)
Azimi et al[69]	2015	Database (single-center)	402	ANN, Logistic regression analysis	Successful ANN model to predict recurrent lumbar disk herniation (AUC 0.84)
Ratliff et al[70]	2016	Database (National)	279 135	LASSO (GLMnet), multivariate logistic regression	Adverse events (AUC 0.61)
Azimi et al[56]	2017	Database (single-center)	346	ANN	Optimal treatment choice for LSCS patients (AUC 0.89)
Oh et al[71]	2017	Database (Multi-center)	234	C5.0 algorithm (type of decision tree model)	Post-operative improvement AUC (0.96)
Scheer et al[72]	2017	Database (Multi-center)	557	C5.0 algorithm (type of decision tree model)	Major intra- or perioperative complications (AUC 0.89)
Staarjes et al[73]	2018	Registry (single-center)	422	TensorFlow ANN	Favorable outcome (AUC 0.87)
Khor et al[50]	2018	Database (Multi-center)	1 965	Multivariate analysis	Predicting lower ODI: nonprivate insurance workers' compensation (0.20), current smoking (0.43) or previous smoking (0.66), asthma (0.54), and a lower baseline score (1.05)
Iderberg et al[62]	2018	Registry (Multi-center)	19 131	Multivariate, regression analysis / GLM	Predicting Clinical outcomes: Odds ratios: Social welfare (1.34) / Living Alone (1.14) / Educational level (-2.39) / Disposable income (-2.58)
Kim et al[35]	2018	Registry (Multi-center)	22 629	ANNs and multivariate logistic regression	Wound complications and mortality (AUC 0.6 to 0.71)
Karhade et al[74]	2018	Registry (Multi-center)	26 364	SVM, ANN	Prediction of anormal discharges (AUC 0.82)
Kuo et al[75]	2018	Database (Single-center)	532	SVMs, logistic regression, C4.5 decision tree	Medical costs (AUC 0.90)
Kalagara et al[65]	2018	Registry (Multi-center)	26 869	R Foundation for statistical computing/ GBM	Readmission (AUC 0.69)
Goyal et al[76]	2019	Registry (Multi-center)	59 145	GLM/ GMB/ ANN/ RF / pLDA/ VarBayes	Discharge to non-home facility (AUC >0.80)
Han et al[66]	2019	MarketScan & Medicaid Databases (Multi-center)	1 106 234	Multivariate logistic regression analysis	Predicting the risk of a pulmonary complication (AUC 0.76)
Siccoli et al[64]	2019	Registry	635	Random forests, extreme gradient boosting (XGBoost), Bayesian generalized linear models (GLMs), boosted trees, k-nearestneighbor, simple GLMs, artificial neural networks with a single hidden layer	Extended hospital stay with an accuracy of 77% (AUC 0.58)
Shah et al[77]	2019	Database (single-center)	367	Logistic regression analysis, Stochastic gradient boosting, Random Forest, Support Vector machine	Failure of nonoperative management. Random Forest (AUC 0.56) Logistic Regression (AUC 0.79)
Karhade et al[78]	2019	Database (single-center)	1 053	Logistic regression analysis, Stochastic gradient boosting, Random Forest, Support Vector machine	Prediction of 90-day mortality in spinal epidural abscess (AUC 0.89)
Hopkins et al[79]	2019	Registry (Multi-center)	23 264	ANN (7 layers)	Readmissions (AUC > 0.60)
Nelson et al[80]	2019	Database (Single-center)	22 318 appointments	ANN, Logistic regression analysis, Support vector machine, Random Forest	Scheduled appointment attendance in healthcare ANN AUC (0.81)
Karhade et al[63]	2019	Database (Multi-center)	5 413	Logistic regression analysis, Stochastic gradient boosting, Random Forest, Support Vector machine	Prolonged postoperative opioid prescription (AUC 0.81)
Hopkins et al[81]	2020	Database (single-center)	4046	ANN (9 layers deep neural network)	Prediction of infections (AUC 0.78)

Notes: ACC = accuracy; ACS-NSQIP = American College of Surgeons National Surgical Quality Improvement Program; ANN = artificial neural networks; AUC = area under the receiver operating characteristic curve; COPD = chronic obstructive pulmonary disease; DNN = deep neural networks; EHR = electronic health records; GBM = gradient boosting machine; GLM = generalized linear model; GLMnet = elastic-net GLM; LSS = lumbar spinal stenosis; MCID = minimum clinically important difference; ML = machine learning; NPV = negative predictive value; NRS = numeric rating scale; NRS-BP = NRS for back pain; NRS-LP = NRS for leg pain; ODI = Oswestry Disability Index; PHC = predictive hierarchical clustering; PPV = positive predictive value; PROMs = patient-reported outcome measures; RF = random forest; ROC = receiver operating characteristic

Table 3. Synthetic Patient Models.

Study	Authors	Patient synthetic model and technology	Keypoint
He et al[87]	2008	Adaptive Synthetic Sampling Method for Imbalanced Data (ADASYN)	Reducing the bias introduced by the class imbalance, and promote recognition of complex patients
Teutonico et al[88]	2015	Discrete re-sampling and multivariate normal distribution (MVND) methodologies in the creation of virtual patient population	The multivariate distribution method produces realistic covariate correlations, comparable to the real population. Moreover, it allows simulation of patient characteristics beyond the limits of inclusion and exclusion criteria in historical protocols.
McLachlan et al[89]	2016	The CoMSER method takes a constraint-based approach involving: (1) formalizing clinical practice guidelines into the CareMap constraint and the CareMap into the State Transition Machine (STM), (2) incorporating published Health Incidence Statistics based constraints into the STM, and (3) exploiting domain expertise in verifying domain knowledge and creating the reusable library of clinical notes	Production of synthetic EHR that is considered realistic. The main contribution of this work is the approach that uses a CareMap for generating synthetic EHR with neither access to the real EHR nor using anonymized EHR. .
Kim et al[90]	2018	ADASYN	Adaptive synthetic sampling approach to imbalanced learning (ADASYN) was used to generate positive synthetic complications for training model
Kim et al[35]	2018	ADASYN	ADASYN utilizes examples from the minority class that are difficult to learn and generates synthetic new cases based on these examples to improve model learning and generalizability
Baowaly et al[83]	2019	MedWGAN / MedBGAN (modified Generating Adversarial network)	Learn the distribution of real-world EHRs and exhibit remarkable performance in generating realistic synthetic EHRs for both binary and count variables.
Pollack et al[91]	2019	5 Steps Generating Synthetic Patient Data*	Steps to generate EHR for testing and evaluation of Health information technology

Predictors

The potential predictive factors were identified based on a comprehensive literature review (see Table 1.) on PubMed central library using the following MESH terms combined to the screening of preoperative data available in our EHRs (see Table 4.):

From Predictors to Criteria Tables

The potential predictors had to be usable in a neural network algorithm (see part Training and validation of the model). In the input table each criterion was a binary value (1 or 0) that represents the presence or absence. So, each predictor was transformed into discrete criterium to fill the binary values tables.

"Machine Learning"[Mesh] OR "Artificial Intelligence"[Mesh] OR "Natural Language Processing"[Mesh] OR "Neural Networks (Computer)"[Mesh] OR "Support Vector Machine"[Mesh] OR Machine learning[Title/Abstract] OR Artificial Intelligence[Title/Abstract] OR Neural network[Title/Abstract] OR Neural networks[Title/Abstract] OR Natural language processing[Title/Abstract] OR deep learning[Title/Abstract] OR machine intelligence[Title/Abstract] OR computational intelligence[Title/Abstract] OR computer reasoning[Title/Abstract])) AND ((("Neurosurgery"[Mesh] OR "Neurosurgical Procedures"[Mesh] OR "Intervertebral Disc Displacement"[Mesh] OR "Spinal Stenosis"[Mesh] OR neurosurgery[Title/Abstract] OR neurosurgeries[Title/ Abstract] OR neurosurgical[Title/Abstract] OR neurosurgically[Title/Abstract] OR spinal [Title/Abstract] OR lumbar[Title/Abstract] AND ("Surgical Procedures, operative"[Mesh] OR "Postoperative Complications"[Mesh] OR "surgery" [Subheading] OR "Postoperative Period"[Mesh] OR "Perioperative Period"[Mesh] OR "Preoperative Period"[Mesh] OR surgery[Title/Abstract] OR surgeries[Title/Abstract] OR surgical[Title/Abstract] OR post-operative*[Title/Abstract] OR post-operative*[Title/Abstract] OR preoperative*[Title/Abstract] OR preoperative*[Title/Abstract] OR perioperative*[Title/Abstract] OR peri-operative*[Title/Abstract] OR operative procedure*[Title/Abstract])))) NOT (Comment[Publication Type] OR editorial[Publication Type] OR letter[Publication Type] OR case reports[Publication Type])."

Table 4. Patient Baseline Predictors.

Variable	Binary criteria (1;0)	Baseline Strength established
Day of surgery	Same day; day before	0%
Length of stay (LOS)	> 4 days: < 4 days	10%
Timing for procedure (1st,2nd,3rd, 4th, 5th positioning in the day)	3rd, 4th, 5th in the day; 1st, 2nd,3rd	10%
Type of job: sedentary	Presence; absence	30%
Type of job: heavy worker	Presence; absence	30%
Work stopping duration before surgery-sedentary >1, 0	< 1 day	10%
Work stopping duration before surgery-heavy worker >3, 0	< 3 days	10%
Work stopping duration before surgery-moderate >14, 0	< 14 days	10%
Work stopping duration before surgery-light worker >35, 0	< 35 days	10%
Sleep disorder	Presence; absence	15%
Professional conflict	Presence; absence	30%
Family conflict	Presence; absence	15%
Specific physical activity	Presence; absence	30%
General physical activity	Absence; presence	30%
Appetite	Absence; presence	5%
Age	> 65 ans	15%
BMI	> 30	50%
Smoking	> 10 pack-year	10%
Pre-operative walking distance reduction	Presence; absence	15%
Prior to surgery opioid consumption	Presence; absence	20%
Cauda equina syndrome	Presence; absence	30%
Transit disorders	Presence; absence	5%
Pre-operative motor deficit	Presence; absence	20%
Pre-operative sensitive deficit	Presence; absence	Indication
Impulsive movement or pushing effort	Presence; absence	30%
Pre-operative inflammatory pain	Presence; absence	30%
Limp	Presence; absence	10%
Acute lumbar pain	Presence; absence	5%
Chronic lumbar pain	Presence; absence	30%
Lumbar stifness	Presence; absence	20%
Sphincter dysfunction	Presence; absence	40%
Diabete	Presence; absence	10%
Pre-operative anxiety or depressive syndrome	Presence; absence	20%
Sleep apnea syndrome	Presence; absence	10%
COPD	Presence; absence	5%
Pneumopathy	Presence; absence	20%
Liver disorder	Presence; absence	15%
Atheroma	Presence; absence	15%
Kidney Disease	Presence; absence	5%
Pre-operative MODIC Images	Presence; absence	30%
Pre-operative Calcification	Presence; absence	30%
Pre-operative stenosis	Presence; absence	Indication
Pre-operative protrusion	Presence; absence	0%
Pre-operative excluded disc herniation	Absence; presence	50%
Pre-operative disc herniation	Presence; absence	Discrete
L1L2 Level	Presence; absence	30%
L2L3 Level	Presence; absence	30%
L3L4 Level	Presence; absence	30%
Pre-operative arthritis	Presence; absence	0%
Pre-operative hypertrophic facet disease	Presence; absence	0%
Pre-operative osteophyte	Presence; absence	0%
Pre-operative spondylolysis	Presence; absence	0%
Explicit pre-operative explanations	Absence; Presence	50%
Favorable operator experience	Absence;presence	70%
Food intake improvement	> 3 days	10%
Sleep improvement	> 2 days	20%
Return to work sedentary >42	> 42 days	30%
Return to work light >42	> 42 days	30%

(continued)

Table 4. (continued)

Variable	Binary criteria (1;0)	Baseline Strength established
Return to work moderate >75	> 75 days	30%
Return to work heavy workers >90	> 90 days	30%
Infection	Presence; absence	15%
Autonomous walking recovery	> 2 days	20%
Anti-inflammatory drugs post-operatively	Presence; absence	10%
Post-operative anxiety or depressive syndrome	Presence; absence	20%
Post-operative disc calcification	Presence; absence	20%
Post-operative stenosis	Presence; absence	40%
Post-operative fibrosis	Presence; absence	50%
Rehabilitation inpatients center	Convalescent home; home	20%
Operative recurrence	Presence; absence	50%

Statistical Analysis

Criteria for real and synthetic patients were compared. The mean percentage of presence for each criterion for each zone (green and orange), as well as the mean number of criteria for each category of patients and each zone were reported.

Synthetic Patient Model

Our synthetic patient model allows us to generate as many virtual patients as we desire in order to train the classifier without the need of real patients. The model that we propose can help in bootstrapping a new model without long and costly data collection, it could also be used to boost under represented categories in classification problem.[35]

It is a statistical approach designed to create a virtual model, statistically representative of real patients' population. Our method was to create patients that fall in the 2 zones that we defined (orange or green). To do so, we generated tables of random pre-op symptoms based on the input data defined before. Each input data (criteria) has a probability of presence, either 1 or 0 (present or not) based on a uniform distribution.

Then, each criterium was associated with a strength. The strength of each criteria was determined by a cross-professional group including spine surgeons, clinical register experts and statisticians.

In the input table, each criterium strength was added to the total strength of the table. This total strength was compared to a threshold, classifying patient in the orange zone (superior to the threshold) or the green zone (inferior)

$$tot_{strength} = \sum_{i=0}^{nb_{symptoms}} S_i * P_i$$

Tables are generated for 10000 virtual patients, of which 5000 are green and 5000 are orange.

Artificial Neural Network Architecture

Our classifier is an artificial neural network, which architecture is based on our criteria (see Figure 1). Each input neuron
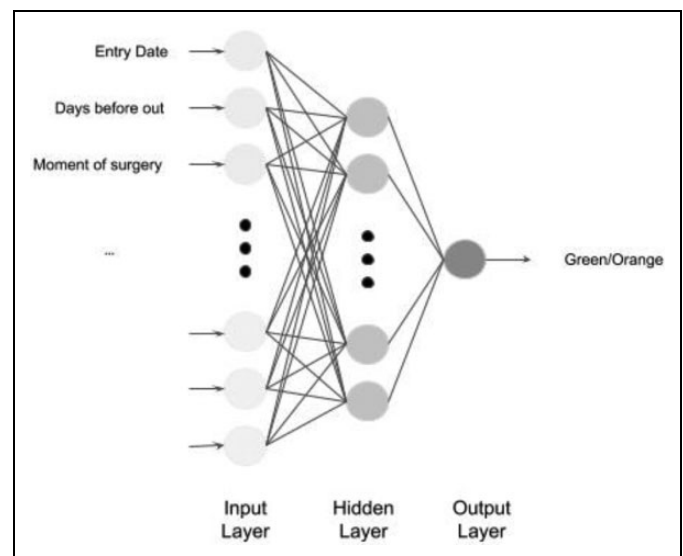

Figure 1. Architecture of our artificial neural network.

represents a pre-operative criterium and the value associated is the presence or the absence of it.

Activation functions for input and hidden layers are Rectified Linear Unit (ReLU). The activation function of the output layer is a sigmoid, the output value is then a Boolean: 1 if green, 0 if orange (See Figure 1). We use Keras Tensorflow framework for the construction and training of our model.

Training and Validation of the Model

The training of the classifier is done using 80% of the data set of virtual patients and 20% were used for testing purposes. The sets are randomly chosen in the virtual patient's dataset, but we keep the 50% green and orange repartition. The algorithm chosen for loss calculation is binary cross entropy and Adam optimizer for back propagation.

The indicator that we use for real data is twofold: accuracy of the model—i.e. classification in either green or orange zone for a given table, and the ROC curve—i.e. the percentage of true positive on false positive at different thresholds.
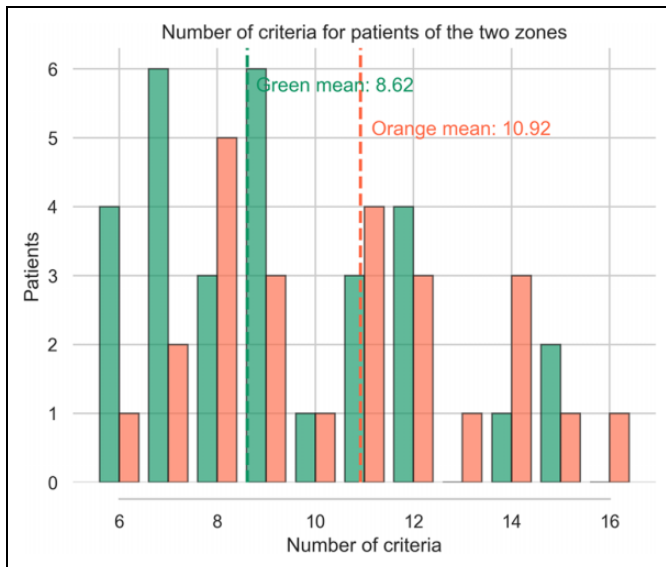
Figure 2. Real patient distribution according the number of pre operative criteria and their outcome (green: success/orange: failure).
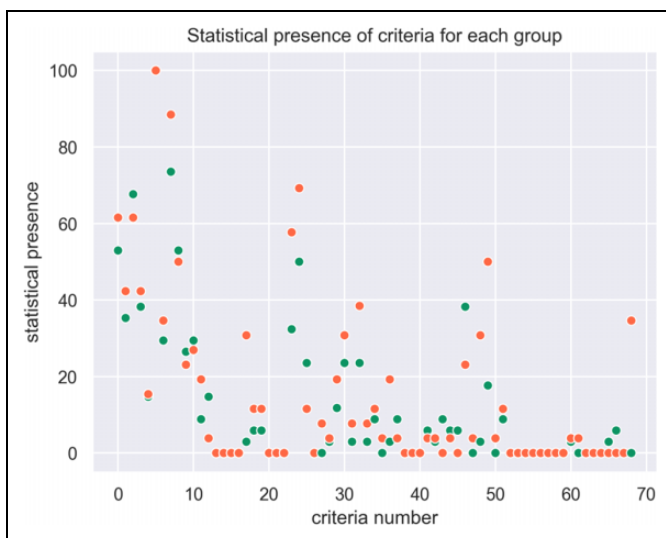

Figure 3. Statistical presence of criteria for each group orange / green (EHR).

Validation of the ANN is done against real patient tables using the Receiver Operating Characteristic Curve (AUC).

Results

Population and EHR Data Set

In the actual cohort, we included 60 patients, with complete EHR allowing sufficient analysis, 26 patients are in the orange zone constituting (43.4%) and 34 are in the green zone (56.6%) (See Figure 2). The average positive criteria amount for actual patients is 8.5 for the green zone (SD+/- 3.09) and 10.47 for the orange zone (SD 3.38). Results are presented in Figures 2 and 3.

Table 5. Patient's Clinical Outcomes (orange zone).

Clinical characteristic evaluated	Binary criteria (1;0)	Area
Walking distance still limited at 1 month	Presence; absence	Orange zone
Partial recovery from post-operatively motor deficit at 1 month	Presence; absence	Orange zone
Partial recovery from post-operatively sensory deficit at 1 month	Presence; absence	Orange zone
Post-operative neuropathic pain at 1 month	Presence; absence	Orange zone
Post-operative anxiety-depression syndrome at 1 month	Presence; absence	Orange zone

Predictors

A total of 68 unfavorable predictors were collected and included in the initial training of the predictive model (See Table 4.). Those 68 criteria are used (58 "type of criteria" and their variants). Among the 68 criteria, 54 are pre-operative criteria and 14 are peri-operative criteria (from surgery to 1-month follow-up). Missing criteria are also counted.

5 other criteria are related to Patient-Related Outcome and allow us to assess the improvement of the quality of life (See Table 5.). The presence of one of these criteria defines the patient's outcome as falling into the orange zone. Our machine learning model was then evaluated through the correct patient classification in the orange zone.

Synthetic Data Set

We generated 10000 virtual patients for training our classifier, 5000 were allocated to the green zone, 5000 to the orange zone. We chose a 50/50 split in order not to introduce a bias of distribution between the 2 zones during the algorithm training. We also generated 2000 tables for testing (20% of the training set).

Figure 4 shows a Gaussian distribution of the number of criteria for the 2 zones.

For patients in the green zone we found a mean of 7.92 symptoms per table, (median: 9, SD +/- 1.71), for patients in the orange zone the mean is 10.93, (median: 11, SD +/- 1.81). These numbers are coherent with what we observe in real patient distributions (see Figure 2.). Submitting the number of criteria to a Welch's test we get a value of -71.31715 with a p-value of 0.0, confirming that the difference in number of criteria for the 2 zones is significantly different.

Indeed, patients in the orange zone tend to have more criteria. Moreover, the higher the strength of a criteria the higher the probability of presence is for that symptom in the orange category. For instance, the predictor "BMI >30" is more represented in orange tables (16.88%) than in green ones (1.84%). Conversely, most of the criteria with low strength are represented with nearly the same proportion in the 2 categories

(<2%): age, appetite, COPD, transit disorders, Sleep apnea work stopping duration before surgery-light worker>35, kidney disease and diabetes.

The statistical presence of each criteria in each zone is plotted in Figure 5.

The combination of several criteria leads from green to orange zone, i.e, the presence of 1 or 2 criteria is not significant in itself to classify the patient outcome. In our synthetic population, 5 criteria are present more than 20% of the time, but these criteria alone do not determine the zone.

Comparison of Criteria Between Real Patient and Synthetic Patient

The criteria proportions in each cohort are compared in Table 6. In order to assess the relevance of the virtually generated patients and their representativeness, we used an open-clustering approach.

As we are conscious of the lack of exhaustive data in the real patients cohort criteria, we presume that several non-significantly different criteria could be finally relevant if
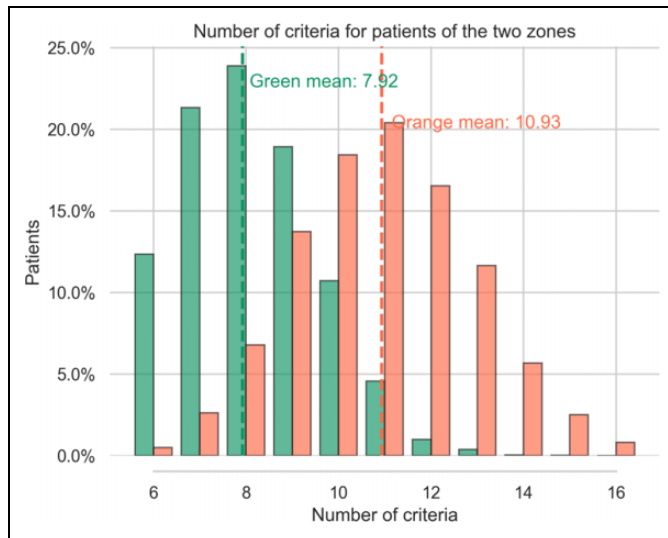

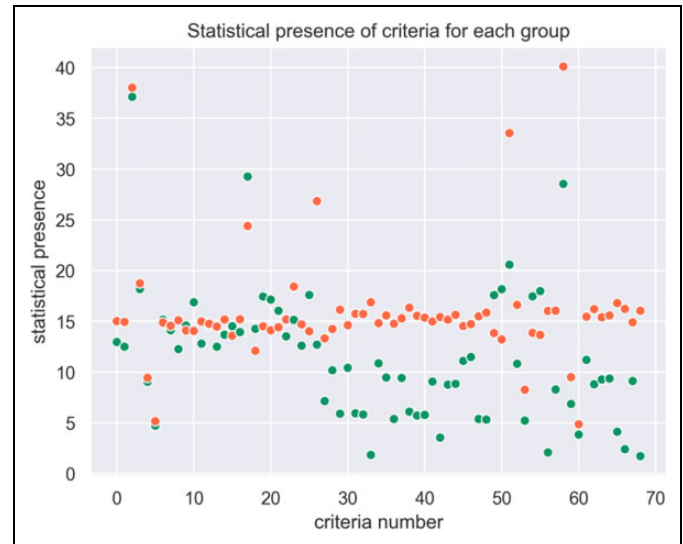**Figure 4.** Number of patient criteria for the 2 zones (syn-EHRS).

Figure 5. Statistical presence of criteria for each group (syn-EHRs).

Table 6. Real and Synthetics Patient's Predictors Distribution (%).

	Criteria	Green_real (%)	Orange_real (%)	Green_synth (%)	Orange_synth (%)
0	Day of surgery	52.94	61.54	17.6	14.02
1	Length of stay (LOS)	35.29	42.31	12.96	15.02
2	Timing for procedure (1st, 2nd,3 rd, 4th, 5th in the day)	67.65	61.54	12.5	14.94
3	Type of job sedentary	8.82	19.23	12.7	26.84
4	Type of job worker	14.71	3.85	7.14	13.32
5	Work stopping duration before surgery-sedentary>1	0	0	37.12	38.02
6	Work stopping duration before surgery-heavy worker>3	0	0	18.18	18.74
7	Work stopping duration before surgery-moderate>14	0	0	9.04	9.44
8	Work stopping duration before surgery-light worker>35	0	0	4.72	5.16
9	Sleep disorder	2.94	30.77	10.18	14.24
10	Professional conflict	5.88	11.54	5.9	16.14
11	Family conflict	5.88	11.54	10.42	14.62
12	Specific physical activity	0	0	5.94	15.74
13	General physical activity	0	0	5.82	15.72
14	Appetite	0	0	15.16	14.88
15	Age	32.35	57.69	14.12	14.56
16	BMI	50	69.23	1.84	16.88
17	Smoking	23.53	11.54	12.26	15.1
18	Pre-operative walking distance	38.24	42.31	10.86	14.82
19	Prior to surgery opioid consumption	0	0	9.46	15.58
20	Cauda equina syndrome	0	7.69	5.38	14.76
21	Transit disorders	2.94	3.85	14.58	14.1
22	Pre-operative motor deficit	11.76	19.23	9.42	15.3

(continued)

Table 6. (continued)

	Criteria	Green_real (%)	Orange_real (%)	Green_synth (%)	Orange_synth (%)
23	Pre-operative sensitive deficit	23.53	30.77	16.88	14.06
24	Impulsive movement or pushing effort	14.71	15.38	6.1	16.34
25	Pre-operative inflammatory pain	2.94	7.69	5.72	15.54
26	Limp	100	100	12.8	14.98
27	Acute lumbar pain	29.41	34.62	14.64	14.76
28	Chronic lumbar pain	73.53	88.46	5.78	15.36
29	Lumbar stiffness	23.53	38.46	9.06	14.98
30	Sphincter dysfunction	2.94	7.69	3.54	15.42
31	Diabetes	8.82	11.54	12.5	14.48
32	Pre-operative anxiety or depressive syndrome	0	3.85	8.76	15.16
33	Sleep apnea syndrome	2.94	19.23	13.68	15.18
34	COPD	8.82	3.85	14.52	13.58
35	Pneumopathy	0	0	8.84	15.64
36	Liver disorder	0	0	11.1	14.54
37	Atheroma	0	0	11.48	14.72
38	Kidney Disease	5.88	3.85	13.94	15.2
39	Pre-operative MODIC Images	2.94	3.85	5.38	15.5
40	Pre-operative Calcification	8.82	0	5.32	15.86
41	Pre-operative stenosis	52.94	50	17.58	13.84
42	Pre-operative protrusion	5.88	3.85	18.16	13.22
43	Pre-operative excluded disc herniation	5.88	0	29.26	24.4
44	Pre-operative disc herniation	38.24	23.08	14.26	12.1
45	L1L2 Level	0	3.85	20.58	33.54
46	L2L3 Level	2.94	30.77	10.82	16.62
47	L3L4 Level	17.65	50	5.22	8.26
48	Pre-operative arthrosis	26.47	23.08	17.44	14.5
49	Pre-operative hypertrophic facet disease	29.41	26.92	17.14	14.12
50	Pre-operative osteophyte	0	3.85	17.46	13.86
51	Pre-operative spondylolysis	8.82	11.54	17.98	13.66
52	Explicit pre-operative explanations	0	0	2.08	16.02
53	Operator experience (years of practice)	0	0	16.04	14.42
54	Food intake improvement	0	0	13.52	15.18
55	Sleep improvement	0	0	8.28	16.04
56	Return to work sedentary >42	0	0	28.54	40.1
57	Return to work light >42	0	0	15.14	18.42
58	Return to work moderate >75	0	0	6.86	9.5
59	Return to work heavy workers >90	0	0	3.84	4.86
60	Infection	2.94	3.85	11.2	15.46
61	Autonomous walking recovery	0	3.85	8.8	16.2
62	Anti-inflammatory drugs	0	0	12.6	14.7
63	Post-operative anxiety or depressive syndrom	0	0	9.28	15.4
64	Post-operative disc calcification	0	0	9.36	15.58
65	Post-operative stenosis	2.94	0	4.12	16.8
66	Post-operative fibrosis	5.88	0	2.4	16.22
67	Rehabilitation inpatients center	0	0	9.12	14.9
68	Operative recurrence	0	34.62	1.72	16.04

correctly assessed. Therefore, we preserve them to keep a maximum of meaningful data for the training of our machine learning and increase the reliability of our synthetic population.

Training and Validation of the Model (ANN Results)

The classifier is trained using 10000 patients from the training set and 2000 patients from the test set. The batch size is 2000 and the model is trained for 100 epochs. The loss decreases rapidly, and the accuracy is growing also quickly. After 50 epochs the model is already close to convergence (see Figure 6.).

The test set is also synthetic and does not provide a solid way of stopping the model before overfitting because it has the same convergence as the training set. Thus, we use the real data to test our model and stop training.

After 100 epochs the test on real data gives an accuracy of 72% and the ROC curve is as follows with a ROC score of 0.78 (See Figure 6). The sensitivity of our model is then 88,5%, specificity is 58%, PPV is 62% an NPV 87%, these numbers for each zone are reported in Table 7.
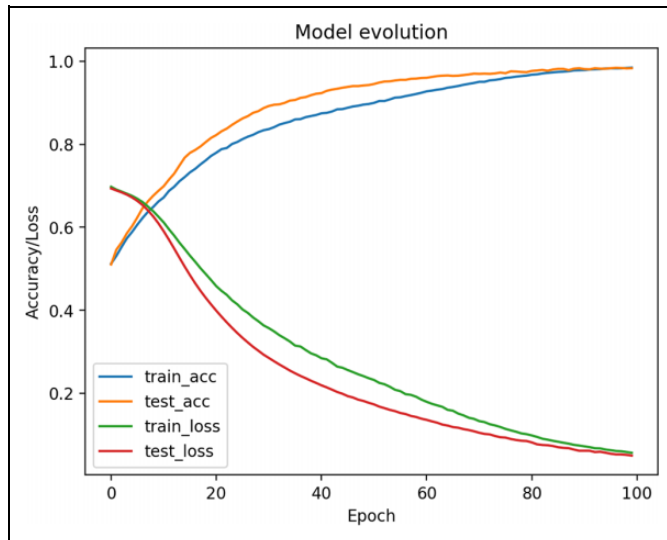
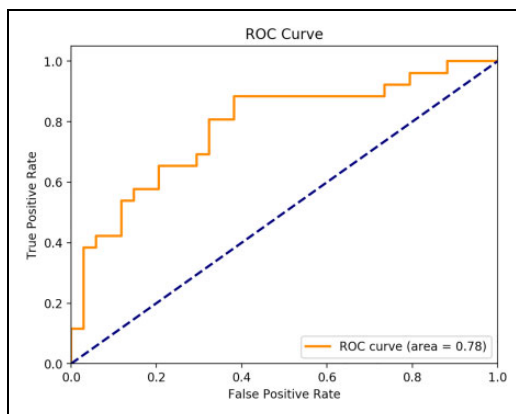**Figure 6.** Training model evolution (Accuracy and loss / Number of epochs).

Figure 7. AUC of our ANN-models using EHRs and syn-EHRs.

Discussion

Our results show similar risk factors identified in other cohorts.[92] In our real patients cohort, age > 65 years, BMI> 30, surgery same day of hospital entry, chronic low back pain are strongly predictive of the orange zone. In our virtual cohorts, sedentary job, L1L2 level, return to work to sedentary job >42 days, work stopping duration before surgery-sedentary>1, are the strongest predictors for the orange zone, ie. treatment failure or poor improvement.

However, on their own, they cannot determine one outcome or the other. This illustrates the need for an individual predictive tool based on several predictors, having multiple degrees of influence (strength) on the outcome.

Our model was statistically representative of the real data. We also used the real data as the validation set of the classifier, in order to better fit the real world.

Table 7. ANN Model for Predict Successful Spine Surgery.

	Precision	Recall	f1-score	Support
Orange Zone	0.62	0.885	0.73	26
Green Zone	0.87	0.59	0.70	34
Accuracy			0.72	60
Macro average	0.75	0.74	0.72	60
Weighted avg	0.76	0.72	0.71	60
ANN Model global performance				
ROC AUC Score	Sensitivity	Specificity	PPV	NPV
0.78	0.885	0.59	0.62	0.87

Notes: PPV = Positive Predictive Value; NPV = Negative Predictive Values

Our machine learning model can classify the orange population in 88,5% of cases, whereas our green zone is correctly classified in 59% of the cases. The overall precision, calculated by the area under the ROC curve (AUC) is 0.78 (see Figure 7).[35,56,63,74,67-76,78-81] This model is particularly suitable for screening patients who react negatively to lumbar surgery, with similar sensitivity to other predicting tools recently published. Nevertheless, there is still a lack of specificity, maybe due to the 23 missing criteria from the database, which prevent our model to evaluate their impact as clinical predictors. Although ANNs show very promising performance, it was trained using virtual patients generated by our model, thus limiting the precision of the response in real cases. Moreover the study sample of real patients was small, and therefore this study will need to be repeated with larger, multicentre datasets and external validation to convincingly demonstrate its validity and predictive power.

The goal of our method was to obtain a reproducible, repeatable, and usable tool, that can fit with various databases, deal with missing data and can be applied to similar stakes. Indeed, the missing complete electronic patient data, the difficulty to access it and the inability to standardize and exploit this data make the development of an omniscient prediction tool challenging.

Thus, we increase the number of exploitable variables (below the significance threshold) to obtain an individual response, we generate virtual patients to increase the size of our training cohort, and we use medical know-how as a tool for architecture of our virtual patients to answer a data quantity problem.

Our algorithm is based on deep learning, which goal is to use as much data as possible to increase its accuracy and precision. The more intensive the use of the algorithm, the better the accuracy in cases statistically farther and farther from the center of the Gaussian. Indeed, the amount of data influences the variability of this data. This increases the number of "rare" cases far from the median value, making it less necessary to use techniques to boost their number (data augmentation). The real cases collected by retro-analysis of the data will gradually replace the data augmentation of the training set and the model

will increase its robustness. This method is used in all machine learning algorithms whose training is supervised. Successive versions are improved by increasing the dataset as the actual data is captured.[93]

As we move toward personalized medicine and value-based care, there is an increasing need to collect and use PRO scores not just in research settings, but also in routine clinical care or quality improvement activities.[50] The progressive digital transformation in the healthcare facilities should allow us to collect more precise and valuable clinical data.

Conclusion

Our method can be used to predict outcome lumbar decompression surgery. There is still a need to further develop its ability to analyze patients in the "failure of treatment" zone in order to offer precise management of patient health before spinal surgery. Through the exploitation of a larger database more representative over time, we think that our model will be capable of improving classification of the orange zone. This model is in concordance with already published machine-learning tools in spine surgery, successfully allowing to predict the improvement of post-operative symptomatology[64,94] and reduction of drug consumption.[38,95,96] Thus, it will be possible to administer the patient's health monitoring to reduce the post-operative risks and above all to promote its recovery after surgery with appropriate therapies. In addition, a software suite could help surgical practice by reducing the surgical gesture to its anatomical usefulness by avoiding the psychological or iatrogenic undesirable effects inherent in the medico-social framework of the intervention.

ORCID iD

Arthur André, MD, MSc https://orcid.org/0000-0002-0075-5257
Jean-Jacques Vignaux, MSc https://orcid.org/0000-0002-2767-772X

References

1. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus non-surgical therapy for lumbar spinal stenosis. *N Engl J Med*. 2008;358(8):794-810.
2. Lurie JD, Tosteson TD, Tosteson A, et al. Long-term outcomes of lumbar spinal stenosis: eight-year results of the Spine Patient Outcomes Research Trial (SPORT). *Spine (Phila Pa 1976)*. 2015;40(2):63-76.
3. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus non-surgical treatment for lumbar degenerative spondylolisthesis. *N Engl J Med*. 2007;356(22):2257-2270.
4. Bailey CS, Rasoulinejad P, Taylor D, et al. Surgery versus conservative care for persistent sciatica lasting 4 to 12 months. *N Engl J Med*. 2020;382(12):1093-1102.
5. McGirt MJ, Bydon M, Archer KR, et al. An analysis from the Quality Outcomes Database, Part 1. Disability, quality of life, and pain outcomes following lumbar spine surgery: predicting likely individual patient outcomes for shared decision-making. *J Neurosurg Spine*. 2017;27(4):357-369.
6. Nasser R, Yadla S, Maltenfort MG, et al. Complications in spine surgery. *J Neurosurg Spine*. 2010;13(2):144-157.
7. Yeramaneni S, Robinson C, Hostin R. Impact of spine surgery complications on costs associated with management of adult spinal deformity. *Curr Rev Musculoskelet Med*. 2016;9(3):327-332.
8. Kalakoti P, Hendrickson NR, Bedard NA, Pugely AJ. Opioid utilization following lumbar arthrodesis: trends and factors associated with long-term use. *Spine (Phila Pa 1976)*. 2018;43(17):1208-1216.
9. Austevoll IM, Gjestad R, Grotle M, et al. Follow-up score, change score or percentage change score for determining clinical important outcome following surgery? An observational study from the Norwegian Registry for Spine Surgery evaluating patient reported outcome measures in lumbar spinal stenosis and lumbar degenerative spondylolisthesis. *BMC Musculoskelet Disord*. 2019;20(1):31.
10. Katz JN, Stucki G, Lipson SJ, Fossel AH, Grobler LJ, Weinstein JN. Predictors of surgical outcome in degenerative lumbar spinal stenosis. *Spine*. 1999;24(21):2229.
11. Hägg O, Fritzell P, Ekselius L, Nordwall A. Predictors of outcome in fusion surgery for chronic low back pain. A report from the Swedish Lumbar Spine study. *Eur Spine J*. 2003;12(1):22-33.
12. Kohlboeck G, Greimel KV, Piotrowski WP, et al. Prognosis of multifactorial outcome in lumbar discectomy: a prospective longitudinal study investigating patients with disc prolapse. *Clin J Pain*. 2004;20(6):455-461.
13. Trief PM, Ploutz-Snyder R, Fredrickson BE. Emotional health predicts pain and function after fusion: a prospective multicenter study. *Spine*. 2006;31(7):823-830.
14. Slover J, Abdu WA, Hanscom B, Weinstein JN. The impact of comorbidities on the change in Short-Form 36 and Oswestry scores following lumbar spine surgery. *Spine*. 2006;31(17):1974-1980.
15. Braybrooke J, Ahn H, Gallant A, et al. The impact of surgical wait time on patient-based outcomes in posterior lumbar spinal surgery. *Eur Spine J*. 2007;16(11):1832-1839.
16. Mannion AF, Elfering A, Staerkle R, et al. Predictors of multi-dimensional outcome after spinal surgery. *Eur Spine J*. 2007;16(6):777-786.
17. Park P, Upadhyaya C, Garton HJL, Foley KT. The impact of minimally invasive spine surgery on perioperative complications in overweight or obese patients. *Neurosurg*. 2008;62(3):693-699.
18. Garcia RM, Messerschmitt PJ, Furey CG, Bohlman HH, Cassinelli EH. Weight loss in overweight and obese patients following successful lumbar decompression. *JBJS*. 2008;90(4):742-747.

19. Vaidya R, Carp J, Bartol S, Ouellette N, Lee S, Sethi A. Lumbar spine fusion in obese and morbidly obese patients. *Spine*. 2009; 34(5):495-500.

20. Chen S, Anderson MV, Cheng WK, Wongworawat MD. Diabetes associated with increased surgical site infections in spinal arthrodesis. *Clin Orthop Relat Res*. 2009;467(7):1670-1673.

21. Abbott AD, Tyni-Lenné R, Hedlund R. Leg pain and psychological variables predict outcome 2–3 years after lumbar fusion surgery. *Eur Spine J*. 2011;20(10):1626-1634.

22. Senker W, Meznik C, Avian A, Berghold A. Perioperative morbidity and complications in minimal access surgery techniques in obese patients with degenerative lumbar disease. *Eur Spine J*. 2011;20(7):1182-1187.

23. Chaichana KL, Mukherjee D, Adogwa O, Cheng JS, McGirt MJ. Correlation of preoperative depression and somatic perception scales with postoperative disability and quality of life after lumbar discectomy. *J Neurosurg Spine*. 2011;14(2):261-267.

24. Sinikallio S, Aalto T, Airaksinen O, Lehto SM, Kröger H, Viinamäki H. Depression is associated with a poorer outcome of lumbar spinal stenosis surgery: a two-year prospective follow-up study. *Spine*. 2011;36(8):677-682.

25. Kalanithi PA, Arrigo R, Boakye M. Morbid obesity increases cost and complication rates in spinal arthrodesis. *Spine*. 2012;37(11): 982-988.

26. Sørlie A, Moholdt V, Kvistad KA, et al. Modic type I changes and recovery of back pain after lumbar microdiscectomy. *European Spine J.*, 2012;21(11): 2252-2258.

27. Pakarinen M, Vanhanen S, Sinikallio S, et al. Depressive burden is associated with a poorer surgical outcome among lumbar spinal stenosis patients: a 5-year follow-up study. *Spine J*. 2014;14(10): 2392-2396.

28. Hellum C, Johnsen LG, Gjertsen Ø, et al. Predictors of outcome after surgery with disc prosthesis and rehabilitation in patients with chronic low back pain and degenerative disc: 2-year follow-up. *Eur Spine J*. 2012;21(4):681-690.

29. Gaudelli C, Thomas K. Obesity and early reoperation rate after elective lumbar spine surgery: a population-based study. *Evid Based Spine Care J*. 2012;3(02):11-16.

30. Mehta AI, Babu R, Karikari IO, et al. 2012 Young investigator award winner: the distribution of body mass as a significant risk factor for lumbar spinal fusion postoperative infections. *Spine*. 2012;37(19):1652-1656.

31. Sharma A, Muir R, Johnston R, Carter E, Bowden G, Wilson-MacDonald J. Diabetes is predictive of longer hospital stay and increased rate of Clavien complications in spinal surgery in the UK. *Ann R Coll Surg Engl*. 2013;95(4):275-279.

32. Takahashi S, Suzuki A, Toyoda H, et al. Characteristics of diabetes associated with poor improvements in clinical outcomes after lumbar spine surgery. *Spine*. 2013;38(6):516-522.

33. Bekelis K, Desai A, Bakhoum SF, Missios S. A predictive model of complications after spine surgery: the National Surgical Quality Improvement Program (NSQIP) 2005–2010. *Spine J*. 2014; 14(7):1247-1255.

34. Lee D, Armaghani S, Archer KR, et al. Preoperative opioid use as a predictor of adverse postoperative self-reported outcomes in patients undergoing spine surgery. *JBJS*. 2014;96(11):e89.

35. Kim JS, Arvind V, Oermann EK, et al. Predicting surgical complications in patients undergoing elective adult spinal deformity procedures using machine learning. *Spine Deform*. 2018;6(6): 762-770.

36. Coronado RA, George SZ, Devin CJ, Wegener ST, Archer KR. Pain sensitivity and pain catastrophizing are associated with persistent pain and disability after lumbar spine surgery. *Arch Phys Med Rehabil*. 2015;96(10):1763-1770.

37. McGirt MJ, Sivaganesan A, Asher AL, Devin CJ. 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*. 2015;39(6):E13.

38. Anderson JT, Haas AR, Percy R, Woods ST, Ahn UM, Ahn NU. Chronic opioid therapy after lumbar fusion surgery for degenerative disc disease in a workers' compensation setting. *Spine (Phila Pa 1976)*. 2015;40(22):1775-1784.

39. Chotai S, Sivaganesan A, Parker SL, McGirt MJ, Devin CJ. Patient-specific factors associated with dissatisfaction after elective surgery for degenerative spine diseases. *Neurosurg*. 2015; 77(2):157-163.

40. Schöller K, Steingrüber T, Stein M, et al. Microsurgical unilateral laminotomy for decompression of lumbar spinal stenosis: long-term results and predictive factors. *Acta Neurochir*. 2016;158(6): 1103-1113.

41. Archer KR, Devin CJ, Vanston SW, et al. Cognitive-behavioral–based physical therapy for patients with chronic pain undergoing lumbar spine surgery: a randomized controlled trial. *J Pain*. 2016; 17(1):76-89.

42. Asher AL, Devin CJ, McCutcheon B, et al. Patient characteristics of smokers undergoing lumbar spine surgery: an analysis from the Quality Outcomes Database. *J Neurosurg Spine*. 2017;27(6): 661-669.

43. Mummaneni PV, Bisson EF, Kerezoudis P, et al. Minimally invasive versus open fusion for grade I degenerative lumbar spondylolisthesis: analysis of the Quality Outcomes Database. *Neurosurg Focus*. 2017;43(2):E11.

44. Crawford CH, Carreon LY, Bydon M, Asher AL, Glassman SD. Impact of preoperative diagnosis on patient satisfaction following lumbar spine surgery. *J Neurosurg Spine*. 2017;26(6):709-715.

45. Suri P, Pearson AM, Zhao W, et al. Pain recurrence after discectomy for symptomatic lumbar disc herniation. *Spine*. 2017; 42(10):755.

46. Sharma M, Ugiliweneza B, Aljuboori Z, Nuño MA, Drazin D, Boakye M. Factors predicting opioid dependence in patients undergoing surgery for degenerative spondylolisthesis: analysis from the MarketScan databases. *J Neurosurg Spine*. 2018;29(3): 271-278.

47. Dunn LK, Durieux ME, Fernández LG, et al. Influence of catastrophizing, anxiety, and depression on in-hospital opioid consumption, pain, and quality of recovery after adult spine surgery. *J Neurosurg Spine*. 2018;28(1):119-126.

48. Chan AK, Bisson EF, Bydon M, et al. Laminectomy alone versus fusion for grade 1 lumbar spondylolisthesis in 426 patients from the prospective Quality Outcomes Database. *J Neurosurg Spine*. 2018;30(2):234-241.

49. O'Donnell JA, Anderson JT, Haas AR, et al. Preoperative opioid use is a predictor of poor return to work in workers' compensation patients after lumbar diskectomy. *Spine*. 2018;43(8):594-602.

50. Khor S, Lavallee D, Cizik AM, et al. Development and validation of a prediction model for pain and functional outcomes after lumbar spine surgery. *JAMA Surg*. 2018;153(7):634-642.

51. Dobran M, Nasi D, Paracino R, et al. Analysis of risk factors and postoperative predictors for recurrent lumbar disc herniation. *Surg Neurol Int*. 2019;10:36.

52. Staub LP, Aghayev E, Skrivankova V, Lord SJ, Haschtmann D, Mannion AF. Development and temporal validation of a prognostic model for 1-year clinical outcome after decompression surgery for lumbar disc herniation. *Eur Spine J*. 2020;29(7):1742-1751.

53. Mauro D, Nasi D, Paracino R, et al. The relationship between preoperative predictive factors for clinical outcome in patients operated for lumbar spinal stenosis by decompressive laminectomy. *Surg Neurol Int*. 2020;11:27.

54. Rudolfsen JH, Solberg TK, Ingebrigtsen T, Olsen JA. Associations between utilization rates and patients' health: a study of spine surgery and patient-reported outcomes (EQ-5D and ODI). *BMC Health Serv Res*. 2020;20(1):1-8.

55. Asher AL, Devin CJ, Archer KR, et al. An analysis from the quality outcomes database, part 2. Predictive model for return to work after elective surgery for lumbar degenerative disease. *J Neurosurg Spine*. 2017;27(4):370-381.

56. Azimi P, Mohammadi HR, Benzel EC, Shahzadi S, Azhari S. Use of artificial neural networks to decision making in patients with lumbar spinal canal stenosis. *J Neurosurg Sci*. 2017;61(6): 603-611.

57. Azimi P, Mohammadi HR, Benzel EC, Shahzadi S, Azhari S. Use of artificial neural networks to predict recurrent lumbar disk herniation. *J Spinal Disord Tech*. 2015;28(3):E161-165.

58. André A. The information technology revolution in health care. In: Arthur A, ed. *Digital Medicine*. Springer International Publishing; 2019:1-7.

59. Dillavou ED, Sun Z, Peiris P, Huang E, Benrashid E, Chang B. IP145. Deep learning-based risk model for best management of closed surgical incisions following vascular surgery. *J Vasc Surg*. 2019;69(6):e149-e150.

60. Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: a scoping review. *PLoS One*. 2019;14(2):e0212356.

61. Mandal I. Developing new machine learning ensembles for quality spine diagnosis. *Knowl-Based Syst*. 2015;73(1):298-310.

62. Iderberg H, Willers C, Borgström F, et al. Predicting clinical outcome and length of sick leave after surgery for lumbar spinal stenosis in Sweden: a multi-register evaluation. *Eur Spine J*. 2019;28(6):1423-1432.

63. Karhade AV, Ogink PT, Thio QCBS, et al. Development of machine learning algorithms for prediction of prolonged opioid prescription after surgery for lumbar disc herniation. *Spine J*. 2019;19(11):1764-1771.

64. Siccoli A, Marlies P, Schröder ML, Staartjes VE. Machine learning–based preoperative predictive analytics for lumbar spinal stenosis. *Neurosurg Focus*. 2019;46(5):E5.

65. Kalagara S, Eltorai AEM, Durand WM, DePasse JM, Daniels AH. Machine learning modeling for predicting hospital readmission following lumbar laminectomy. *J Neurosurg Spine*. 2018;30(3): 344-352.

66. Han SS, Azad TD, Suarez PA, Ratliff JK. A machine learning approach for predictive models of adverse events following spine surgery. *Spine J*. 2019;19(11):1772-1781.

67. Azimi P, Benzel EC, Shahzadi S, Azhari S, Mohammadi HR. Use of artificial neural networks to predict surgical satisfaction in patients with lumbar spinal canal stenosis. *J Neurosurg Spine*. 2014;20(3):300-305.

68. Azimi P, Benzel EC, Shahzadi S, Azhari S, Zali AR. Prediction of successful surgery outcome in lumbar disc herniation based on artificial neural networks. *Global Spine J*. 2014;4(1_suppl): s-0034.

69. Azimi P, Mohammadi HR, Benzel EC, Shahzadi S, Azhari S. Use of artificial neural networks to predict recurrent lumbar disk herniation. *Clin Spine Surg*. 2015;28(3):E161-E165.

70. Ratliff JK, Balise R, Veeravagu A, et al. Predicting occurrence of spine surgery complications using "big data" modeling of an administrative claims database. *JBJS*. 2016;98(10): 824-834.

71. Oh T, Scheer JK, Smith JS, et al. Potential of predictive computer models for preoperative patient selection to enhance overall quality-adjusted life years gained at 2-year follow-up: a simulation in 234 patients with adult spinal deformity. *Neurosurg Focus*. 2017;43(6):E2.

72. Scheer JK, Smith JS, Schwab F, et al. Development of a preoperative predictive model for major complications following adult spinal deformity surgery. *J Neurosurg Spine*. 2017;26(6): 736-743.

73. Staartjes VE, Marlies P, Vandertop WP, Schröder ML. Deep learning-based preoperative predictive analytics for patient-reported outcomes following lumbar discectomy: feasibility of center-specific modeling. *Spine J*. 2019;19(5):853-861.

74. Karhade AV, Ogink P, Thio Q, et al. Development of machine learning algorithms for prediction of discharge disposition after elective inpatient surgery for lumbar degenerative disc disorders. *Neurosurg Focus*. 2018;45(5):E6.

75. Kuo C-Y, Yu L-C, Chen H-C, Chan C-L. Comparison of models for the prediction of medical costs of spinal fusion in Taiwan Diagnosis-Related Groups by machine learning algorithms. *Healthc Inform Res*. 2018;24(1):29-37.

76. Goyal A, Ngufor C, Kerezoudis P, McCutcheon B, Storlie C, Bydon M. Can machine learning algorithms accurately predict discharge to nonhome facility and early unplanned readmissions following spinal fusion? Analysis of a national surgical registry. *J Neurosurg Spine*. 2019;1-11.

77. Shah AA, Karhade AV, Bono CM, Harris MB, Nelson SB, Schwab JH. Development of a machine learning algorithm for prediction of failure of nonoperative management in spinal epidural abscess. *Spine J*. 2019;19(10):1657-1665.

78. Karhade AV, Shah AA, Bono CM, et al. Development of machine learning algorithms for prediction of mortality in spinal epidural abscess. *Spine J*. 2019;19(12):1950-1959.

79. Hopkins BS, Yamaguchi JT, Garcia R, et al. Using machine learning to predict 30-day readmissions after posterior lumbar fusion: an NSQIP study involving 23,264 patients. *J Neurosurg Spine*. 2019;1(aop):1-8.

80. Nelson A, Herron D, Rees G, Nachev P. Predicting scheduled hospital attendance with artificial intelligence. *NPJ Digit Med*. 2019;2(1):1-7.

81. Hopkins BS, Mazmudar A, Driscoll C, et al. Using artificial intelligence (AI) to predict postoperative surgical site infection: a retrospective cohort of 4046 posterior spinal fusions. *Clin Neurol Neurosurg*. 2020;192:105718.

82. Yoon J, Drumright LN, Van Der Schaar M. *Anonymization through data synthesis using Generative Adversarial Networks (ADS-GAN)*. IEEE Journal of Biomedical and Health Informatics; 2020.

83. Baowaly MK, Lin C-C, Liu C-L, Chen K-T. Synthesizing electronic health records using improved generative adversarial networks. *J Am Med Inform Assoc*. 2019;26(3):228-241.

84. Buczak AL, Babin S, Moniz L. Data-driven approach for creating synthetic electronic medical records. *BMC Med Inform Decis Mak*. 2010;10(1):59.

85. Yale A, Dash S, Dutta R. Privacy preserving synthetic health data. *F1000Res*. 2019;8:724.

86. Walonoski J, Kramer M, Nichols J, et al. Synthea: an approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *J Am Med Inform Assoc*. 2017;25(3):230-238.

87. He H, Bai Y, Garcia EA, Li S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, Hong Kong, China, 1-8 June 2008. IEEE; 2008.

88. Teutonico D, Musuamba F, Maas HJ, et al. Generating virtual patients by multivariate and discrete re-sampling techniques. *Pharm Res*. 2015;32(10):3228-3237.

89. McLachlan S, Dube K, Gallagher T.Using the caremap with health incidents statistics for generating the realistic synthetic electronic healthcare record. In: *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, Chicago, IL, 4-7 October 2016. IEEE; 2016.

90. Kim JS, Merrill RK, Arvind V, et al. Examining the ability of artificial neural networks machine learning models to accurately predict complications following posterior lumbar spine fusion. *Spine (Phila Pa 1976)*. 2018;43(12):853-860.

91. Pollack AH, Simon TD, Snyder J, Pratt W. Creating synthetic patient data to support the design and evaluation of novel health information technology. *J Biomed Inform*. 2019;95: 103201.

92. Lønne G, Fritzell P, Hägg O, et al. Lumbar spinal stenosis: comparison of surgical practice variation and clinical outcome in three national spine registries. *Spine J*. 2019; 19(1):41-49.

93. Hestness J, Narang S, Ardalani N, et al. Deep learning scaling is predictable, empirically. arXiv. *arXiv preprint arXiv:171200409*. 2017.

94. Malik AT, Khan SN. Predictive modeling in spine surgery. *Ann Transl Med*. 2019;7(suppl 5):S173.

95. Hills JM, Pennings JS, Archer KR, et al. Preoperative opioids and 1-year patient-reported outcomes after spine surgery. *Spine (Phila Pa 1976)*. 2019;44(12):887-895.

96. Oleisky ER, Pennings JS, Hills J, et al. Comparing different chronic preoperative opioid use definitions on outcomes after spine surgery. *Spine J*. 2019;19(6):984-994.