Development and validation of a multimodal feature fusion prognostic model for lumbar degenerative disease based on machine learning: a study protocol

Zhipeng Wang 1,2, Xiyun Zhao 1,2, Yuanzhen Li 2, Hongwei Zhang 2, Daping Qin 1, Xin Qi 1, Yixin Chen 1, Xiaogang Zhang 1,2

ABSTRACT
Introduction Lumbar degenerative disease (LDD) is one of the most common reasons for patients to present with low back pain. Proper evaluation and treatment of patients with LDD are important, which clinicians perform using a variety of predictors for guidance in choosing the most appropriate treatment. Because evidence on which treatment is best for LDD is limited, the purpose of this study is to establish a clinical prediction model based on machine learning (ML) to accurately predict outcomes of patients with LDDs in the early stages by their clinical characteristics and imaging changes.

Methods and analysis In this study, we develop and validate a clinical prognostic model to determine whether patients will experience complications within 6 months after percutaneous endoscopic lumbar discectomy (PELD). Baseline data will be collected from patients’ electronic medical records. As of now, we have recruited a total of 580 participants (n=400 for development, n=180 for validation). The study’s primary outcome will be the incidence of complications within 6 months after PELD. We will use an ML algorithm and a multiple logistic regression analysis model to screen factors affecting surgical efficacy. We will evaluate the calibration and differential performance of the model by the area under the curve. Sensitivity (Sen), specificity, positive predictive value and negative predictive value will be reported in the validation data set, with a target of 80% Sen. The results of this study could better illustrate the performance of the clinical prediction model, ultimately helping both clinicians and patients.

Ethics and dissemination Ethical approval was obtained from the medical ethics committee of the Affiliated Hospital of Gansu University of Traditional Chinese Medicine (Lanzhou, China; No. 2022-57). Findings and related data will be disseminated in peer-reviewed journals, at conferences, and through open scientific frameworks.


INTRODUCTION
Low back pain, as an important cause of quality-of-life decline in middle-aged and elderly people, has brought a heavy burden to society.1 Lumbar disc degeneration and herniation are the primary causes of low back pain, primarily involving a series of pathophysiological changes such as disc degeneration, intervertebral height reduction, disc endplate degeneration and vertebral body degeneration.2 At present, empirical diagnosis of lumbar degeneration is based on imaging data: herniation site, intervertebral disc (IVD) signal changes, vertebral body signal and morphological changes, and spinal-sequence stability. These factors help determine the necessity of surgery and, if surgery is necessary, inform the surgical plan for the patient.3 However, as this diagnostic process is extremely dependent on the accumulated experience of doctors, diagnostic and surgical plans created by different
Physicians can differ significantly. Therefore, a decision aid system based on artificial intelligence (AI) with automatic analytic and diagnostic capabilities can help clinicians diagnose lumbar degenerative disease (LDD) and draw up surgical plans, improving clinician efficiency and the quality of the patient’s medical treatment.

Percutaneous endoscopic lumbar discectomy (PELD) has been widely used, having achieved good clinical efficacy in the treatment of LDD. However, many risk factors lead to poor postoperative efficacy of this intervention. Most published studies use logistic multivariate analysis to establish prediction models. In traditional machine learning (ML) models, data engineers extract features from raw data and develop corresponding feature extractors. The ML model learns by taking features as inputs and the corresponding results of features as outputs. After a certain amount of training, the ML model can predict the results of new data features. ML algorithms have been used by many scholars to diagnose and treat clinical diseases. It has been reported that by analysing data on gait and inertia in the elderly, ML algorithms can be used to predict fall risk. In addition, using ML to differentiate patients with low back pain from healthy individuals has been reported to help clinicians diagnose low back pain. Finally, some scholars have used ML algorithms to predict biological stress on the spine to judge the distribution of spinal biological stress across different populations.

ML is a branch of AI that uses various learning algorithms to generate predictive models from input data. Typical algorithms include K-nearest neighbour, linear regression, logistic regression (LR), support vector machine (SVM), decision tree (DT), random forest, Bayes point machine (BPM), gradient boosting machine and neural network (NN). ML uses extensive mathematical operations to better define complex relationships between predictors and outcomes. At present, all ML algorithms applied to the construction of prediction models in studies are classified as supervised learning.

The ultimate goal is to enable the computer to accurately predict clinical prognosis based on the experience gained in the learning process. Therefore, the purpose of this study is to conduct an extensive validation of the clinical prediction model we have designed to predict the prognoses of patients with LDD. We will collect clinical data from patients at the Intervertebral Disc Center (IVDC) of the Affiliated Hospital of Gansu University of Traditional Chinese Medicine (TCM).

METHODS AND ANALYSES

Study design
The model for predicting prognosis of LDD study is a single-arm proof-of-concept trial involving a prospective cohort of 580 patients. Patients will be admitted to a single centre specialising in the treatment of LDD and will be followed up for 6 months after treatment. This study has been registered with the Chinese Clinical Trial Register (www.chictr.org.cn) and complies with the Standard Protocol Items: Recommendations for Interventional Trials checklist.

This study was approved by the clinical research ethics committee of the Affiliated Hospital of Gansu University of TCM (No. 2022-57). From October 2022 to February 2024, 580 patients with lumbar disc herniation or lumbar spinal stenosis (LSS) have received or will receive single-level PELD at the IVDC of the Affiliated Hospital of Gansu University of TCM. Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research. For trials flow diagram of the study, see figure 1.

Patient and public involvement
Patients were not directly involved in developing research questions, study design, intervention designs, outcome measures, recruitment and conducting of the study.

Source of data
In this study, we aim to develop and validate a clinical prognostic prediction model based on prospective studies. We will continue to enrol all patients at the IVDC of the Affiliated Hospital of Gansu University of TCM who meet the screening criteria to develop the predictive model (October 2022–February 2023) and the validation model (March 2023–December 2024).

Participants
Inclusion and exclusion criteria
Inclusion criteria
- A definite diagnosis of lumbar disc herniation or LSS.
- Single-segment PELD.
- Age>18 years, regardless of nationality, gender or marital status.
- Willingness to undergo the procedure and to sign the informed consent form.
- Ability to cooperate in completing the relevant examinations and providing reliable medical history data.

Exclusion criteria
- Death of patient (not related to PELD) before the follow-up.
- Fracture, ankylosing spondylitis, scoliosis, tumour, tuberculosis, infections and other diseases.
- Complication with cardiovascular and cerebrovascular diseases, lower extremity vascular diseases and other diseases affecting clinical evaluation.
- Severe mental illness rendering the patient unable to communicate effectively.
- Pregnancy.

Surgical methods
Included patients will have undergone PELD via a foraminal approach using the joimax system (joimax GmbH, Karlsruhe, Germany).

In this procedure, the patient lies in the prone position, with hip and knee flexion and an empty abdominal pad. The midline of the lumbar spinous process and the
horizontal line through the upper edge of the IVD are marked on the skin from the anteroposterior perspective, and the lateral line through the posterior upper edge of the lower vertebral body is marked along the oblique direction of the intervertebral space from the lateral perspective. The intersection point of this lateral line and the horizontal line through the upper edge of the IVD is the puncture point. Local anaesthesia is performed at the puncture site and puncture path with 1% lidocaine. Under anteroposterior and lateral fluoroscopy, an 8-gauge needle is inserted into the anterior and inferior edge of the superior articular process of the lower vertebra in the direction of the above-marked line, and 0.5% lidocaine is injected around the articular process. The guide wire is inserted through an 18-gauge needle, the needle is pulled out, and an 8 mm long skin incision centred on the guide wire is made. An expansion guide rod is inserted along the guide wire, and an expansion catheter is inserted step by step along the guide rod to provide wider surgical access. Through a protective cannula, the trephine is inserted along the guide rod, and part of the bone at the lateral edge of the superior articular process is removed. The ring saw is then removed and repositioned in the working channel along the guide rod, and its position is confirmed via anteroposterior and lateral fluoroscopy. The surgical field is continuously irrigated with 3L of normal saline. Under endoscopic supervision, the nucleus pulposus tissue is removed using an appropriately sized and angled nucleus pulposus clamp, the nerve root is exposed and released, bipolar radiofrequency is used to fully stop bleeding and IVD decompression and annuloplasty are performed. The pulse of the dural sac and the nerve root’s range of motion are examined under a microscope, and the skin incision is sutured after the working channel is removed. Patients can wear a lumbar brace to get out of bed 6 hours after surgery. Sedentary behaviour and excessive waist flexion are to be avoided for 3 months after surgery, and excessive physical labour and strenuous exercise are prohibited.

Outcome
The primary outcome of this study is the incidence of postoperative complications within 6 months after surgery. Complications may include epidural haematoma, infection and recurrence.

Sample size
In predictive modelling, large sample sizes reduce bias and variance, allowing for prospective predictions of new
observations. Previous studies have reported a minimum framework for each predictor of 10 participants in predictive models. Therefore, for this study, we have calculated a minimum required sample of 520 patients. With a possible 10% patient loss, the total sample of participants was set at n=580.

Clinical factors
Each patient has undergone or will undergo a standard clinical assessment on admission. Clinical variables to be recorded are used for model development, including age, sex, height, weight, body mass index, basal body temperature, blood pressure (BP), smoking, alcohol consumption, hypertension, history of diabetes, history of hypertension, history of coronary heart disease, chronic lung disease, atrial fibrillation, chronic renal insufficiency, liver insufficiency, previous surgery and medications taken. Preoperative and postoperative laboratory indicators are as follows: platelet count, total protein, albumin and coagulation function.

We will continue to comprehensively collect patient with LDD clinical data to evaluate risk factors for complications after PELD. Recorded surgical details include surgical approach, operation time, intraoperative blood loss, use of hemostatic materials, amount of intraoperative blood transfusion and amount and type of drugs used. After the operation, key data such as drainage volume, body temperature, BP, blood glucose, length of hospital stay, bedtime and complications are also recorded.

The following imaging features are recorded: grade of IVD degeneration, disc height, height of foramina, classification of end-plate osteochondritis, type of disc herniation, degree of spinal-canal stenosis, high-intensity zone, end-plate form, Cobb angle, pelvic incidence, pelvic tilt and sacral slope.

Blinding
Due to the special nature of surgical intervention, neither patients nor medical staff can be blinded to the interventional measures. However, all investigators, assessors and statisticians will be blinded to the allocation to minimise potential bias.

Statistical analysis methods
Using Python V.3.8 (Guido van Rossum; https://www.python.org/downloads/), by calling the language environment implementing the ML model of data acquisition, data preprocessing, training model, model assessment and prediction. We will use leave-one-out cross-validation to divide the dataset into three parts: training set (80%), test set (10%) and cross-validation set (10%). The training set will be used for the training of the algorithm, from which we can obtain many different models, and then use the cross-validation set to test the generalisation ability of each model and select the best model. Finally, the test set was used to test the generalisation ability of the optimal model, and we will analyse the influences of independent variables on outcome variables to find the most important factors affecting the latter. Bias will be used to measure the learning ability of a single model, and variance to measure the stability of the same model on different data sets.

DT, SVM, NN, BPM and LR algorithms will be tested as prediction models for complication rates. In all five algorithms, regularisation terms are added to prevent overfitting. For each ML algorithm, 10-fold cross-validation will be repeated three times on the training set (80%), to train the algorithms to recognise patterns related to complications that may occur within 6 months after PELD and to subsequently assess their predictive performance based on the following performance characteristics: area under the receiver operating characteristic (ROC) curve, calibration (calibration slope, calibration intercept) and Brier score will be calculated. The model's predicted probability is plotted against the actual observed probability to calculate calibration of a model. The permutation test will be used to compare the area under the area under the ROC curve (ROC AUC) between the five models, and the test level α value will be 0.05. We will report sensitivity (Sen), specificity, positive predictive value and negative predictive value in the validation dataset, with a target of 80% Sen.

DISCUSSION
This protocol describes in detail how we will validate existing clinical prediction models for predicting outcomes in patients with LDD. This study will be conducted on a diverse sample of patients from the IVDC of the Affiliated Hospital of the Gansu University of TCM, and the results will better illustrate the performance of existing models. This will be the first confirmatory study to evaluate the performance of the original prediction model designed to predict the likelihood of recovery in patients with LDD at various follow-up time points, and it will be one of only a few studies that validate a predictive model designed to inform the prognoses of patients with LDD.

ML holds great promise for improving the efficiency and quality of clinical research, but substantial barriers remain; surmounting them will require addressing significant gaps in evidence. Previous studies have confirmed the feasibility of our research. Zemp et al used five ML algorithms to analyse the influence of human sitting posture on low back pain based on force and acceleration sensors and studied seven different prescribed sitting positions (n=1148). The results of the five algorithms were compared using the leave-one-out cross-validation method. Meanwhile, the relationship between sitting position and low back pain was analysed in depth by accurate assessment of chair use. Azimi et al reported using NNS to evaluate decision-making in LSS using artificial-NN (ANN) and LR model prediction and analysing 346 patients. Measurement indicators included the Visual Analogue Scale of pain, the Japanese Orthopaedic Association Score, the Neurogenic Claudication Outcome
Score and the Oswestry Disability Index. Compared with the LR model, the ANN model showed better accuracy (Acc; 97.8%) and a greater ROC AUC (89.0%).

Kim et al. proposed an ANN model to accurately predict incision complications, venous thromboembolism and mortality after posterior lumbar fusion surgery. The predictive variables used included gender, age, race, diabetes, smoking, hormones and coagulopathy. The ANN-based ML model had greater Sen and Acc than other AI models in identifying the factors for lumbar fusion complications. Staartjes et al proposed to establish a prediction model based on pathology and LR for preoperative prediction of patient prognosis after lumbar discectomy. The deep-learning model had a predictive Acc of 85% and could therefore potentially inform patients of symptom improvement before surgery. It can be seen that various ML algorithms are feasible in the diagnosis and prognosis of spinal diseases, and the feasibility of the current study in particular is extremely high.

This study may conclude that compared with the traditional multivariate LR analysis model, a prediction model based on ML algorithms has significant advantages in disease prognosis and multivariate screening of risk prediction, which could be used to construct models predicting the postoperative efficacy in LDDs. Our model can provide a theoretical basis for clinical decision-making, and patients can use it as a predictive tool to evaluate their own prognoses, thereby helping improve their self-management and compliance before and after surgery.

Ethics and dissemination

This study will be conducted according to the principles of the Declaration of Helsinki, adhering to the transparent reporting of multivariate predictive models of individual outcomes or prognostic models of diagnostic claim guidelines. The data obtained from this study will be obtained in the form of journal articles published in peer-reviewed journals, and the input data will be made available in a public repository (http://www.medresman.org.cn/). We intend to publish at least two core articles for peer review, written by members of the core research team. To ensure the maximum impact of our findings, we will actively communicate with academic, community and public health audiences.

Contributors XGZ and ZPW have led on design and were overseeing data analysis plans. Data management was performed by ZPW, XYL, YZL, DPQ, XQ and YXC. Data quality checks were performed by ZPW and HZW. XYL and DPQ were the study statisticians. ZPW and XYL drafted the manuscript. The final dataset was curated by XQ, YXC and ZPW, access will be at the discretion of study investigators.

Funding

This work was supported by Higher Education Innovation Fund Project of Gansu Provincial Education Department, grant number 2022B-109; Lanzhou Talent Innovation and Entrepreneurship Project, grant number 2022-3-25; Construction Project of National Famous Old Chinese Medicine Expert Inheritance Studio of Zhang Xiaogang (Chinese Medicine Education Letter (2022) No. 75); Natural Science Foundation of Gansu Province, grant number 23JRA1192.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs

Zhipeng Wang http://orcid.org/0000-0003-1853-8488 Xiaogang Zhang http://orcid.org/0000-0002-1953-848X

REFERENCES

16 Shmueli G. To explain or to predict Statist Sci 2010;25:289–310.


