medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Integrating Multidimensional Data Analytics for Precision Diagnosis of Chronic Low Back Pain

Sam Vickery[1]*, Frederick Junker[1], Rebekka Döding[1], Daniel L Belavy[1], Maia Angelova[2,3], Chandan Karmakar[3], Louis Alexander Becker[4,5], Nima Taheri[4,5], Matthias Pumberger[4], Sandra Reitmaier[5], Hendrik Schmidt[5]

[1] Division of Physiotherapy / Department of Applied Health Sciences | Hochschule für Gesundheit (University of Applied Sciences) | Bochum | Germany

[2] Aston Digital Futures Institute | Aston University| Birmingham | United Kingdom

[3] School of Information Technology | Deakin University | Geelong | Australia

[4] Center for Musculoskeletal Surgery, Charité – Universitätsmedizin Berlin, Berlin, Germany

[5] Julius Wolff Institut, Berlin Institute of Health - Charité at Universitätsmedizin Berlin | Berlin | Germany

Abstract

Low back pain (LBP) is a leading cause of disability worldwide, with up to 25% of cases become chronic (cLBP). Optimal diagnostic tools for cLBP remains unclear. Here we leveraged a comprehensive multi-dimensional data-set and machine learning-based feature importance selection to identify the most effective diagnostic tools for cLBP patient stratification. The dataset included questionnaire data, clinical and functional assessments, and spino-pelvic magnetic resonance imaging (MRI), encompassing a total of 144 parameters from 1,161 adults with (n=512) and without cLBP (n=649). Boruta and random forest were utilised for variable importance selection and cLBP classification respectively. Boruta feature selection led to pronounced variable reduction (median of all 15 datasets: 63.3%), while performing

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

comparable to using all variables across all modality datasets. Multi-modality models performed better than single modality models. Boruta selected key variables from questionnaire, clinical, and MRI data were the most effective in distinguishing cLBP patients from controls with an AUC (area under the receiver operating characteristic curve) of 0.699 (95% confidence interval [CI], 0.669 – 0.729). The most robust features (n=9) across the whole dataset identified were psychosocial factors, neck and hip mobility, as well as lower lumbar disc herniation and degeneration. These critical variables (AUC = 0.664, 95% CI = 0.514 – 0.814) outperformed all parameters (AUC = 0.602, 95% CI = 0.538 – 0.666) in an unseen holdout dataset, demonstrating superior patient delineation. Paving the way for targeted diagnosis and personalized treatment strategies, ultimately enhancing clinical outcomes for cLBP patients.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Key words

Chronic low back pain; classification; data-driven; feature selection; multi-modality; psychosocial; MRI

CRediT Table

Term	Who?	Definition
Conceptualization	SV, HS, DLB	Ideas; formulation or evolution of overarching research goals and aims
Methodology	SV, HS, DLB, MA, CK	Development or design of methodology; creation of models
Software	SV, FJ	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components
Validation	SV	Verification, whether as a part of the activity or separate, of the overall replication/ reproducibility of results/experiments and other research outputs
Formal analysis	SV, FJ	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data
Investigation	SV, LAB, NT, MP, SR	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection
Resources	HS, DLB	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools
Data Curation	SV, FJ	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse
Writing - Original Draft	SV, DLB, FJ, RD, HS,	Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)
Writing - Review & Editing	DLB, HS, RD, FJ, SV, NT, LAB, MP, SR, MA, CK	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre-or postpublication stages
Visualization	SV, FJ	Preparation, creation and/or presentation of the published work, specifically visualization/ data presentation
Supervision	HS, DLB, MA, CK	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team
Project administration	HS, SV	Management and coordination responsibility for the research activity planning and execution
Funding acquisition	HS, DLB	Acquisition of the financial support for the project leading to this publication

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Glossary

AUC - area under the receiver operating characteristic curve

BMI – body mass index

C – Clinical examination

CI – confidence interval

cLBP – Chronic low back pain

FABQ - Fear-Avoidance Belief Questionnaire

IPAQ - International Physical Activity Questionnaire

IVD – intervertebral disc

M – MRI

MET – metabolic equivalents

MRI - magnetic resonance imaging

Q – Questionnniare

RMDQ - Disability Questionnaire: Roland and Morris

S - Back shape and function

SF-36 - Short-form 36 Health Status Questionnaire

SRBAI - Self-Report Behavioural Automaticity Index

TSK-GV - Tampa Scale for Kinesiophobia

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

Chronic low back pain (cLBP) represents a significant social and economic burden to society with increasing numbers of patients requiring surgical or non-surgical treatment. Several studies showed that 70–85% of the global population suffers cLBP at some point in their life [5], of which 4–25% become chronic [34]. Both surgical (e.g., spinal fusion) and non-surgical (e.g., rehabilitation and pharmacological) treatments for cLBP have risen tremendously in recent years compared to other major musculoskeletal conditions [12,69]. However, the treatment outcomes of both are still inconsistent, which is reflected in a high rate of treatment-refractory cLBP patients [17].

Addressing chronic pain as a multidimensional phenomenon is a major challenge in practice. This challenge is exacerbated [41,51] by the number of potential factors that may differentiate pain from pain-free and a lack of consensus as to what key parameters are. Studies have shown that permanent changes in central nervous system sensitisation contribute to the persistence of pain [47,59], as opposed to acute pain, which typically resolves as tissue heals Previous meta-analyses [55] underscored that experience of cLBP is related to physical, psychological and social elements, highlighting the need for a multifaceted approach to pain assessment.

Adopting a broader perspective on chronic pain, and assessing multidimensional data, promises to advance pain management. In particular, elucidating relevant or irrelevant assessment factors may strengthen the reliability of a diagnosis and thus make treatment more targeted. However, despite the recognition of physical, psychological, and social domains contributing to pain, as highlighted in recent work [55], examining these domain in the same population is currently lacking.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Machine learning algorithms provide models for identifying distinct subgroups that elucidate the occurrence and characteristics of a disease [29,31]. A previous systematic review by our team using machine learning applications in LBP [53] highlighted that a narrow range of mechanistic domains have been assessed, and sample sizes in these studies were consistently small, ranging up to only 171 participants. Consequently, probing limited data and modalities limits the robustness and applicability of such models. Through gathering many data points across multiple modalities one can ascertain which variables and modalities are the most informative at distinguishing cLBP patients from asymptomatic controls. Reducing the number of variables to those that are most informative has been previously employed in predictive and classification modelling to improve accuracy [37,39]. This approach can be applied as the main outcome and not only in model preprocessing, in order to obtain a data-driven decision on the most important variables in multi-dimensional clinical data. Such a systematic data-informed investigation of back pain diagnosis in a large multi-modality sample is lacking to help inform future studies in selecting which data to acquire and for clinicians in which tests to conduct. Therefore, the aim of this study was to identify and compare the most informative domains and variables in delineating patients with and without cLBP utilising a large multi-modality dataset.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Materials and methods

Study design

The prospective cross-sectional study draws its data from the ongoing "Berliner Rückenstudie" ("Berlin Back Study"; https://spine.charite.de/en/spine_study/; running time: 01/01/2022 to 31/12/2025), which was registered at the German Clinical Trial Register (DRKS-ID: DRKS00027907). Recruitment procedures vary from local promotion (i.e., postal flyers, notice boards, internet approaches, and social media) at the Charité-Universitaetsmedizin Berlin, in the general public (i.e., newspapers, magazines, podcasts) to cooperation with local companies, administrative authorities, and word-of-mouth. The protocol is in accordance with the Helsinki Declaration of ethical principles [67] and has been approved by the Ethics Committee of the Charité – Universitätsmedizin Berlin (registry numbers: EA4/011/10, EA1/162/13). Written informed consent was obtained from all participants. The STROBE guideline [63] (Supplemental Table 1) and TRIPOD statement [10] (Supplemental Table 2) for prediction model development were used to report this study. Data collection started on 1st January 2022 and cut-off for inclusion in the current analysis was 5th April 2024. Data collection occurred in a research centre within a university-hospital.

Participants

Study participants were recruited through a telephone interview and excluded if they met any exclusion criteria, as well as some excluded at the testing site (Supplementary Table 3). A total of 1273 participants were included in the study at cut-off point. These participants were initially guided through self-administered questionnaires by a study coordinator. Then they continued

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

to a clinical examination by a trained medical doctor, which included physical examinations, questions, as well as a back shape and function test. The examinations and questionnaires took a total of 90 minutes to complete. Additionally, participants were offered a magnetic resonance imaging (MRI) within 14 days of the spino-pelvic region. During the clinical assessment the participants were classified by the clinician as asymptomatic (no back pain), symptomatic (cLBP), or previously suffering from cLBP. To ensure a more robust cLBP patient classification previous symptomatic subjects were removed from the sample. Furthermore, participants who revoked their inclusion in the study and those who were missing demographic data; age, sex, body mass index (BMI), and patient status were removed. This resulted in a study sample of 1161 subjects that included 649 asymptomatic (19 – 72 years old, mean age = 40.7 ± 12.6, females = 353) and 512 cLBP (19 – 65 years old, mean age = 43.5 ± 11.7, females = 306) participants. This sample was sub-divided into four modalities; questionnaires (Q), clinical physical assessment (C), back shape and function (S), and MRI (M). Only participants with all data within a particular dataset modality were considered for modelling and no interpolation was conducted.

Quantitative variables and data collection

Patients had to meet all of the following inclusion criteria: written informed consent to participate in the study, asymptomatic (no back pain) or symptomatic (cLBP) caucasian women and men aged 18–67 years, pain duration ≥12 weeks daily (cLBP only), pain localization in the lumbopelvic region (cLBP only). A telephone interview was conducted during recruiting and subjects were excluded if they met any exclusion criteria. However, some subjects came to testing that should have been excluded during the telephone interview, and were then excluded at the testing site (Supplementary Table 3). No minimal threshold for LBP intensity was

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

defined. A list of all variables including number of missing values is shown in Supplementary Table 4.

Questionnaires

The localization, type, course, possible radiation, intensity, quality, duration, and any factors that may relieve or exacerbate the pain, as well as possible triggers or the patient's own explanations regarding the cause of the pain has been asked. The patient's medical history has been recorded, which includes any previous diseases and surgeries, and a detailed pain and general medication history as well as allergies, intolerances, and vaccination status among others. A family and social history (anamnesis) was taken (professional activity, family situation, diseases in the family, stressful situations, etc.). In addition, any past or present use of addictive substances has been asked (alcohol, nicotine, etc.).

The following questionnaires were completed within 30 minutes:

- Pain intensity and duration and pain-related disability: von Korff et al. [64]
- Disability Questionnaire: Roland and Morris (RMDQ) [46]
- Short-form 36 Health Status Questionnaire: SF-36 [66]. The following four domains were considered: general mental health (psychological distress and well-being), limitations in usual role activities because of emotional problems, vitality (energy and fatigue), and general health perceptions.
- International Physical Activity Questionnaire (IPAQ) [11]
- Self-Report Behavioural Automaticity Index (SRBAI) [15]
- Behavioural Regulation in Sport Questionnaire (BRSQ) [30]
- Tampa Scale for Kinesiophobia (TSK-GV) [48].
- Fear-Avoidance Belief Questionnaire (FABQ) [65].

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

The participants primarily answered the questionnaires in digital form using a survey program specially developed for the study. The data were collected under similar conditions (e.g., same room, same computer) for all subjects at the study centre.

Demographic data

During the clinical assessment, age, sex, body height, body weight, hip diameter, and waist diameter of the subjects were recorded. BMI was chosen instead of waist hip ratio to measure physical body size and health, as the BMI variable contained less missing values compared to waist hip ratio (Supplementary Table 4).

Clinical examination

The clinical examination included the evaluation of organ functions (inspection, palpation, percussion, and auscultation), the general impression, and the vital parameters of the patient (temperature, heart rate, blood pressure, etc.). It was performed by an experienced orthopaedic consultant. The neurological status was assessed by the examination of the coordination, reflexes, sensitivity, and motor function. The evaluation of the functional parameters, that is, the assessment of posture, shape, orientation, and movement of the lumbar spine and pelvis, was based on current clinical standards (e.g., Ott and Schober test, 3-step hyperextension test, passive lumbar extension test, etc.) and self-assessment by the persons investigated. Data were documented according to their dimension using distances in cm, degrees of angle, and number of repetitions per defined time interval or bivalent whether pain provocation occurred. Self-assessment of functional restrictions of the back was recorded according to a scale from 1 (best) – 10 (worst).

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Back shape and function

All study participants received measurements of the back shape in the sagittal and frontal planes during upright standing and sitting using the Idiag M360 (Idiag AG, Fehraltorf, Switzerland). The device measures segmental angles of the thoracic and lumbar spine. In both postures, study participants were measured upright, in flexed, extended, and in left and right lateral bending (3 repetitions, ~10 sec each). Maximum upper body flexion, extension, as well as left and right lateral bending were performed with extended knees. During extension the arms were crossed in front of the body. The order of performed tasks was randomised. The measurements were performed by trained medical students. The validity and reliability were demonstrated in previous studies [7,13,18,60].

Spino-pelvic MRI

MRIs were conducted using a 1.5 MRI scanner. Following sequences were evaluated: 1) Sag T1 (4 mm slices), 2) Sag T2 (4 mm slices), 3) Cor STIR-T2 (4 mm slices) and 4) Axial T2 (3 mm slices). MRIs were evaluated for intervertebral disc degeneration (Pfirrmann classification [42]), disc herniation (Kramer classification [25]), facet joint arthrosis (Fujiwara classification [14]), osteochondrosis intervertebralis [36], spondylolisthesis (Meyerding classification [35] and spinal canal stenosis (Schizas classification [50]) at each level of the lumbar spine. The spino-pelvic MRI evaluation was performed blinded by two spine surgeons and a radiologist, all of whom have many years of experience in the evaluation of spinal pathologies. The inter-rater reliability was good-to-excellent for all measurement parameters.

Data storage

All data files electronically recorded during the study period were stored on a database server folder (SharePoint folder) hosted by Charite-Universitaetsklinikum. A data back-up for the

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

database is run daily. Local study team members signed a non-disclosure agreement. They have access to the database using a personal password. They are authorized for entries depending on their function based on a role concept (investigator, statistician, monitor, administrator etc.) that regulates permission for each user. A multilevel data validation plan was developed to guarantee the correctness and consistency of the data. Data were entered only after a check for completeness and plausibility. Furthermore, data were cross-checked for plausibility with previously entered data for each participant. Questionnaires filled out on paper are stored in a lockable cabinet at the university.

Potential sources of bias and minimisation

To generally reduce the possible location and assessor bias, measurements (clinical physical and questionnaire assessment and back shape and function) were administered by a few trained clinicians in the same room with the same lighting. The self-administered questionnaires (von Korff, RMDQ, SF-36, IPAQ, SRBAI, BRSQ, TSK-GV, and FABQ) were completed by the subjects under supervision by our trained study coordinator who provided explanations for unclear questions and mitigated possible lack of motivation to complete the questions by assuring the subjects of the importance to complete the questionnaires. Furthermore, generic questionnaires (SF-36 or IPAQ) were placed before specific ones (SRBAI, BRSQ) to minimize bias from order effects.

To minimise the bias in our classification and feature selection, model as well as the modality comparison, variables directly assessing back pain and questions heavily biased to pain patients were removed. Such variables assessed back pain during particular movements or upon physical manipulation. Furthermore, questions only back pain patients were asked for example, pain intensity, duration, and disability, as well as pain and health biased self-questionnaires (von

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Korff, RMDQ, TSK-GV, FABQ, therapies, and the SF-36 sub-categories regarding physical function, physical role function, physical pain, health perception, and vitality), and clinician administered questions regarding pain medication intake, previous spinal disorder diagnosis, and participants' subjective physical health assessment were removed to reduce model bias.

Outcome

The outcome target for our classification model is cLBP patients. All participants were assessed by a clinician and diagnosed as cLBP patient, asymptomatic control, or suffered from LBP in the past but not at present. We used the clinician diagnosis of either current cLBP patient or asymptomatic control as our two-class target outcome. The participants with LBP in the past were removed to enable better distinguishable groups for binary classification and feature importance selection.

Data handling, preprocessing, cleaning and missing data

The variables can be distinguished into four modalities; questionnaires, clinical assessment, back shape and function and motion, and spino-pelvic MRI. Each modality was combined with demographic data (age, sex, and BMI) and then joined with all combinations of the four modalities. This resulted in 15 dataset containing either single-modality, dual-modalities, or multi-modalities data (Fig. 1A). An overview of all variables removed including the reason can be found in Supplementary Table 5 and a list of all variables (144) used for modelling is presented in Supplementary Table 6.

Total and sub-scores for the self-administered questionnaires were used after removal of pain biased questionnaires (see *Potential sources of bias and minimisation*). This includes the SF-36 [66], the IPAQ [11], the SRBAI [15], and the BRSQ [30]. The SF-36 was used to collect statements

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

related to the health domains 'emotional role limitation, (three items) 'social functioning' (two items), and 'mental health' (five items). A scoring algorithm was used to convert the raw scores into these three domains. The scores were transformed to range from zero (worst possible health) and 100 (best possible health). The SRBAI and BRSQ collected ratings for multiple statements on a numerical scale from 1 (strongly disagree) to 6 (strongly agree). For the SRBAI, the total score was calculated by summing the numerical values across all 4 statements. In contrast, sub-scores were created for the BRSQ that related to 'intrinsic motivation', 'integrated regulation' and 'external regulation'. These scores were calculated by summing the numerical values across two statements per sub-score. The IPAQ, recorded the average time spent per day over the past 7 days while 'sitting', 'walking', doing 'moderate activities' (e.g., heavy lifting, digging, aerobics, or fast bicycling), and 'vigorous activities' (e.g. carrying light loads, bicycling at a regular pace, or doubles tennis). To estimate the energy requirements for each activity type, the average time spent per day in minutes was multiplied by MET-score (metabolic equivalents) of 1.5, 3.3, 4, or 8 for sitting, walking, moderate and vigorous activities, respectively. This resulted in MET-minutes scores, describing the amount of energy in kilocalories required for a 60 kilogram person. Finally, the MET-minutes scores were summed across 'sitting', 'walking', 'moderate activities' and 'vigorous activities', resulting in a total MET-minutes score per subject.

Modelling preprocessing was conducted by checking variables within each modality for very low variance and collinearity. Features with close to zero (variance < 1) variance were removed. Furthermore, features presenting a Spearman correlation greater than 0.9 were also removed to reduce collinearity between features. The decision on which of the correlated features to removed, was the feature showing less correlation to the target, cLBP patient status. Following these cleaning steps, single data modalities were joined with demographic data and then for the dual- and multi-modalities datasets join with other datasets. Subjects having any missing values

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

were removed from our analyses and therefore no imputation was conducted. Following cleaning and preprocessing modality dataset presented different number of subjects and features with a similar age and sex distribution across cLBP patients and asymptomatic controls (Fig. 1A).

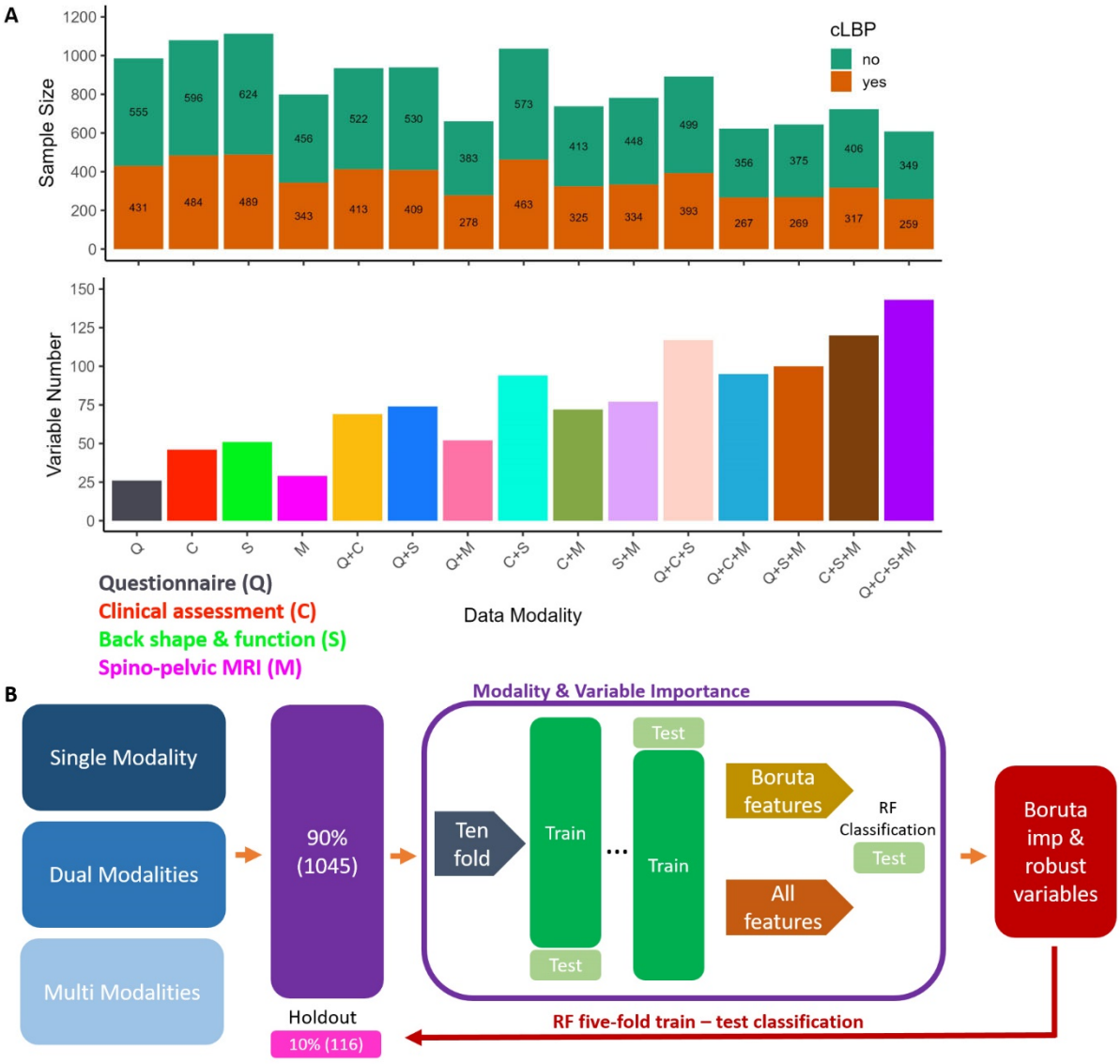

Figure 1. Modality dataset distributions and machine leaning workflow. A – Top shows the chronic low back pain (cLBP) sample size distribution across all 15 dataset modalities. Bottom presents the number of variables used for cLBP classification and variable importance selection across the 15 dataset modalities. B – Represents the machine learning workflow implemented to compare the different modalities and determine the most important variables for cLBP patient delineation using a random forest binary classification algorithm for training and testing.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Univariate statistics

The univariate statistics were carried out separately for continuous, ordinal and nominal data to compare patients suffering from cLBP against asymptomatic controls using R (version 4.3.1; www.r-project.org). As most continuous variables did not follow a normal distribution according to the Anderson-Darling test [4], we implemented the non-parametric Wilcoxon-Mann-Whitney test to determine significant difference between cLBP patients and asymptomatic controls for ordinal and continuous data. Hence, u-values, z-values, r-value (effect sizes), as well as the p-value are reported from. Nominal data were compared using the Chi-Square test and reported with Chi^2 values, Cohen's ω-values (effect size), and p-values. Statistical significance was determined at $p \leq 0.05$ following family wise error (FWE) [21] correction for multiple comparisons within each modality (demographics, questionnaires, clinical examinations, superficial spine morphology and motion, and spino-pelvic MRI).

Machine learning

Boruta feature selection

We used the Boruta method [26] for importance feature selection. Boruta utilises random forest (RF) classification algorithm [8] with both the real variables and set of 'dummy' or 'shadow' variables, that are created by shuffling the feature values. This creates random variables (dummy features) that have the same distribution as the original features, although represent the classification accuracy of this feature randomly sampled. As these dummy features represent random noise, they have had their possible correlation to the target (cLBP) removed. All real and dummy features are used to classify cLBP and feature importance is calculated. Feature importance is calculated as the Z-score of the mean decrease in classification accuracy following the removal of this feature from the model. The importance of the dummy features can be used as a reference to test the feature importance of the real features. Through an iterative

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

process, real features that have significantly greater importance than the maximum dummy variable importance are marked as important. The features that have a significantly lower importance than the maximum dummy features are deemed unimportant and removed for the next iteration of selection. Iterations are repeated until the importance is assigned to all features or a user defined iteration number is reached. We used a max of 2000 iterations with a random forest containing 1000 trees. As there are occasionally a few features not definitively identified as important or removed after 2000 iterations by Boruta, we only selected the features that have been confirmed as important in our subsequent modelling and analyses.

Random forest classification algorithm

We utilised RF [8] implemented using the ranger package [68] in R (version 4.3.1, www.r-project.org) to classify cLBP patients and pain-free controls. As RF provides feature importance measures and can handle categorical, ordinal, and continuous data it represents the ideal choice to deal with the different data types present in the Berlin Back dataset. Ten-fold cross validation was conducted during model training and hyper-parameter tuning utilising 1000 tree RF. We conducted hyper-parameter tuning using a tuning parameter search grid, which contained number of variables to be sampled at each split (mtry) of 1 – square root of the number of features, and a minimum node size of 5 and 10. Therefore, using a gini split rule a grid search was conducted with all combination of hyper-parameters to determine the best. A ten-fold train-test loop (Fig. 1B) was conducted to determine the cLBP classification performance within each of the 15 datasets utilising all and Boruta selected variables independently. Model performance was calculated as follows; accuracy = (TP + TN) / (TP + FP + TN + FN), sensitivity = TP / (TP + FN), specificity = TN / (TN + FP), and AUC (area under the receiver operating characteristic (ROC) curve). Whereby, TP – true positive, TN – true negative, FP – false positive, and FN –

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

false negative. The ROC curve represents the classification performance measured by sensitivity and specificity over a range [0, 1] of classification thresholds.

Robust variable selection workflow

Selecting important variables and comparing the 15 different dataset modalities was conducted on 90% (n = 1045, 19 – 72 y/o, mean age = 41.71 ± 12.28, cLBP = 469) of the preprocessed Berlin Back study dataset, with a 10% (n = 116, 20 – 64 y/o, mean age = 43.81 ± 12.12, cLBP = 43) hold-out set used to evaluate the most robust and important features (Fig. 1B). The hold-out sample comprised of participants that had data for all variables. Each 15 modality datasets went through the iterative ten-fold train-test loop, where Boruta feature selection was conducted on the training set. Followed by RF training using all and Boruta selected variables implemented on the training dataset. Model performance was calculated using the test set. Across the ten loops, the percentage a variable was selected and average importance value was calculated across all 15 datasets. Finally, the variables that were selection within every train-test loop and within every possible dataset were provided as the most robust and important variables for cLBP classification. The demographic features can be selected a maximum of 15 times, while the modality-specific variables a maximum of 8 times across the 15 dataset modalities. These variables were then used in a five-fold train test loop on the hold-out dataset and compared to a model trained using all variables.

Data and code availability

All results in this study are provided in the (Supplementary) tables. The Berlin Back study is currently ongoing (end date 31/12/2025) and therefore the raw data used in this manuscript cannot be provided. The raw data will be openly released from the Berlin Back Study as per

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

agreement with the funding agency following the completion of the data acquisition. A link to the raw data will be provided on the Github repository where the analysis code is located (https://github.com/viko18/BerlinBack_FeatImp/) when it is made available.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Results

Chronic low back pain classification

The best single modality for cLBP classification was Boruta selected features MRI (Fig. 2A) with a mean AUC of 0.645 with a 95% confidence interval (CI) of 0.618 – 0.672 and accuracy of 0.657 (95% CI, 0.636 – 0.678). The Boruta selected questionnaire dataset modality produced only minimally worse classification performance (mean AUC = 0.631, 95% CI = 0.610 – 0.652) than Boruta reduced and all MRI variables (mean AUC = 0.637, 95% CI = 0.599 – 0.675) using the least amount of variables (mean = 8) across all modality datasets (Fig. 2A-C). Dual and multi modalities generally performed better than single modality models in cLBP classification. Questionnaire, clinical physical assessment, and MRI (Q + C + M, Fig. 2C) modality with Boruta selected features represents the best performing modality model with a mean AUC of 0.699 (95% CI, 0.669 – 0.729) and a mean accuracy of 0.709 (95% CI, 0.679 – 0.739). Moreover, this model showed the highest sensitivity (mean = 0.622, 95% CI – 0.568 – 0.676). The modality model showing the highest specificity was using all features and all modalities (Q + C + S + M, mean = 0.840, 95% CI = 0.799 – 0.881). The three best dual modalities models, C + M (mean AUC = 0.679, 95% CI = 0.633 – 0.725), Q + M (mean AUC = 0.674, 95% CI = 0.623 – 0.725), and Q + C (mean AUC = 0.674, 95% CI = 0.639 – 0.709) all with Boruta selected features (Fig. 2B), performed only slightly worse than the best model (Boruta – Q + C + M). Overall back shape and function dataset using Boruta selected features produces the worst classification performance (Fig. 2A, AUC = 0.569, 95% CI = 0.538 – 0.60). Additionally, the dual modalities continuing back shape and function data always performed worse than those without (Fig. 2B), when using both all and Boruta selected features. Classification performance metrics across all models is provided in Supplementary Table 7.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

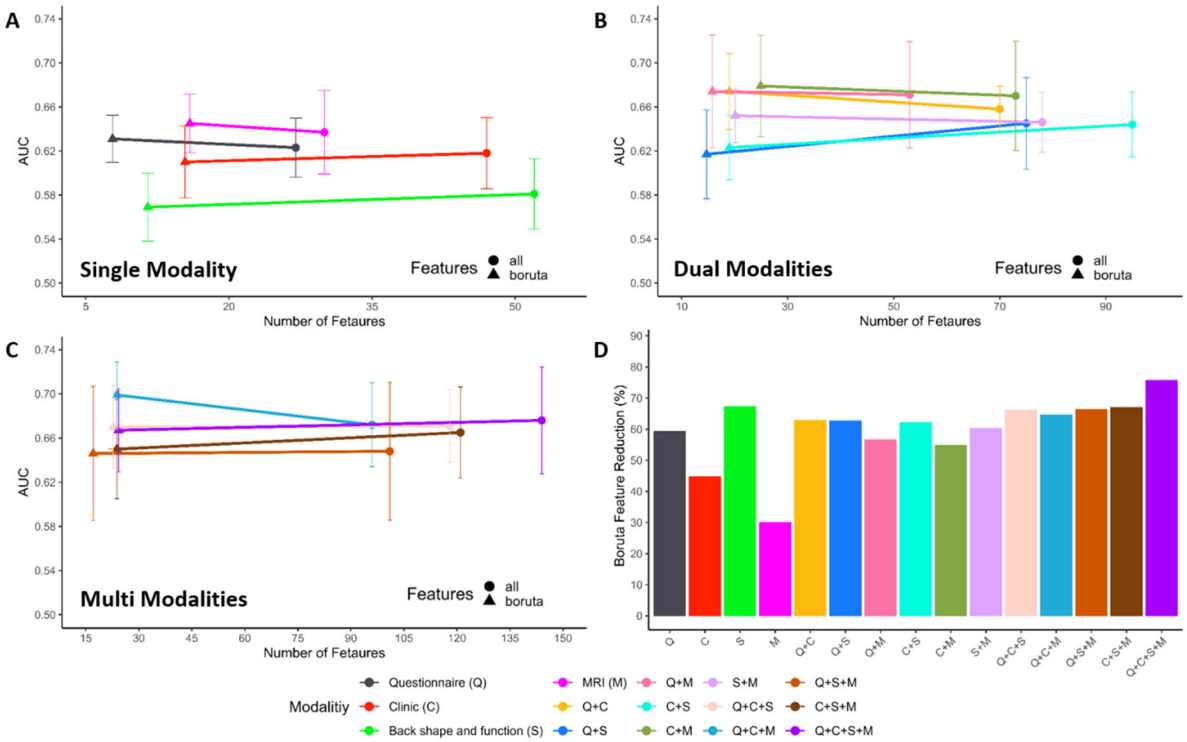

Figure 2. Boruta feature reduction performance. A – C shows RF classification model performance (AUC) following the reduction of features using Boruta and all variables in the single, dual, and multi data modalities respectively. This shows the change in performance follow feature reduction. Error bars represent 95% CI in AUC over 10-fold train-test splits. D – Shows the amount of feature reduction by using Boruta as a percentage of the total number of features within each modality dataset.

Boruta feature importance

Boruta feature importance selection resulted in a median of 62.7% reduction in the number of variables across all 15 datasets (Fig. 2D). This large reduction in variables from Boruta performed comparable (eight slightly worse and seven better) compared to using all features, with high overlap in confidence intervals. Furthermore, three of the top five performing modality models where those employing Boruta selected features (Supplementary Table 7). The greatest performance improvement following Boruta feature selection was found in the Q + C + M datasets with an AUC increase of 0.270 and an average reduction of 72.2 features (Fig. 2C). The smallest feature reduction was shown in MRI (30%) with an AUC increase of 0.008,

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

while the largest feature reduction was found in the whole dataset (Q + C + S + M) with a reduction of 75.7% of the features and an AUC decrease of 0.009.

Boruta selected variables represented the best patient delineation performance in single, dual, and multi-modality datasets (Fig. 2A-C). MRI was found to be the best single modality model using Boruta selected features (Fig. 2A). The most important variables (Supplementary Table 11) were intervertebral disc (IVD) herniation L4 – L5, spinal canal width L2, IVD degeneration L3 – L4 and L4 – L5, and spinal canal width L1, showing a mean importance across the ten iterations of 19.49, 9.94, 9.46, 9.14, and 8.13 respectively. The best dual modalities cLBP patient stratification model, Boruta selected C + M (Fig. 2B), showed the second highest AUC (mean = 0.679). The MRI variable IVD herniation L4 – L5 was found to be the most important (mean = 13.72), with clinical mobility assessments of the hip, cervical spine, and the whole body showing high importance (Supplementary Table 15). The most important and robust variables of the best performing model (Boruta – Q + C + M, Fig. 2C) contained assessments from all three modalities (Supplementary Table 18). The Short-form 36 Health Status Questionnaires (SF-36) psychological well-being, SF-36 social function, and hip pain presented a mean importance of 13.36, 13.07, and 9.71 respectively. The clinical assessment, cervical axial rotation (left) and MRI variable IVD herniation L4 – L5 each showed an importance of 9.47 and 9.12. These represent the top five most important variables and the rest are provided in Table 3. The questionnaire Boruta reduced model provides a sparse model with decent performance, meaning the modality performs comparably well with a small amount of data. On average, the questionnaire Boruta model used 8 variables with a mean AUC reduction of 0.068 compared to the best model (Q + C + M). The most robust and important features (Supplementary Table 8) were SF-36 social function (mean = 21.42), SF-36 psychological well-being (mean = 20.02), hip pain (mean =15.0), smoking in pack years (mean = 11.8), and family

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

history of back pain (mean = 7.27). All Boruta selected important variables across the additional datasets are provided in Supplementary Table 8 - 22.

Most robust and important variables

To select the most robust and important variables for cLBP patient classification, the percentage of selection and average importance score was calculated across the 15 datasets (Supplementary Table 23). The features that were selected at every opportunity (100%) across the multiple iterations and datasets were defined as most important and solely implemented in cLBP classification compared to all variables. This resulted in nine robust and important variables (Fig. 3A). These variables represented psychosocial factors, IVD herniation and degeneration of the lower lumbar spine, presence of hip pain, as well as mobility of the neck and general mobility of the whole body. The performance of these nine variables were compared to all variables utilising the hold-out dataset (Fig. 1) in a five-fold train-test workflow to provide an unbiased comparison. The best nine variables showed better mean accuracy (Fig. 3B, Boruta = 0.724, 95% CI = 0.593 – 0.855, All = 0.680, 95% CI = 0.625 – 0.736), AUC (Boruta = 0.664, 95% CI = 0.514 – 0.814, All = 0.602, 95% CI = 0.538 – 0.666), and sensitivity (Boruta = 0.442, 95% CI = 0.176 – 0.708, All = 0.298, 0.139 – 0.457), while all variables provided better specificity (Boruta = 0.892, 95% CI = 0.795 – 0.989, All = 0.902, 95% CI = 0.803 – 1.0). Moreover, all these nine variables showed significant univariate statistically significant differences between cLBP patients and asymptomatic controls in the questionnaire, clinical, and MRI datasets (Fig. 3A). Utilising a Wilcoxon-Mann-Whitney test revealed reduced scores of the SF36 for social function ($u = 183751$, $z = -7.08$, $p < 0.001$, effect size $r = -.225$) and psychological well-being ($u = 178506.5$, $z = -7.74$, $p < 0.001$, $r = -0.247$) in people suffering from cLBP. In addition, the occurrence of hip pain was also altered comparing people with and without cLBP (χ^2 (3, N=986) = 40.17, $p = 0.004$, $\omega = 0.202$). Comparing clinical examinations

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

further revealed reduced cervical axial rotation to the left (u = 221595, z = -7.98, p < 0.001, r = -0.243), reduced sit to stand 30 second repetition (u = 226818.5, z = -6.83, p < 0.001, r = -0.208), as well as altered general mobility ($\chi 2$ (2, N=1080) = 46.81, p = 0.006, ω = 0.208) in cLBP patients. Regarding MRI investigations, increased IVD degeneration at L2-L3 (u = 147168, z = 3.38, p = 0.011, r = 0.120) and L4-L5 (u = 147691, z = 3.43, p = 0.010, r = 0.121), as well as increased disc herniation at L4-L5 (u = 153470.5, z = 6.02, p < 0.001, r = 0.213) were found in people suffering from cLBP compared to asymptomatic controls. Univariate statistical results for all variable can be found in Supplementary Table 24 – 31.

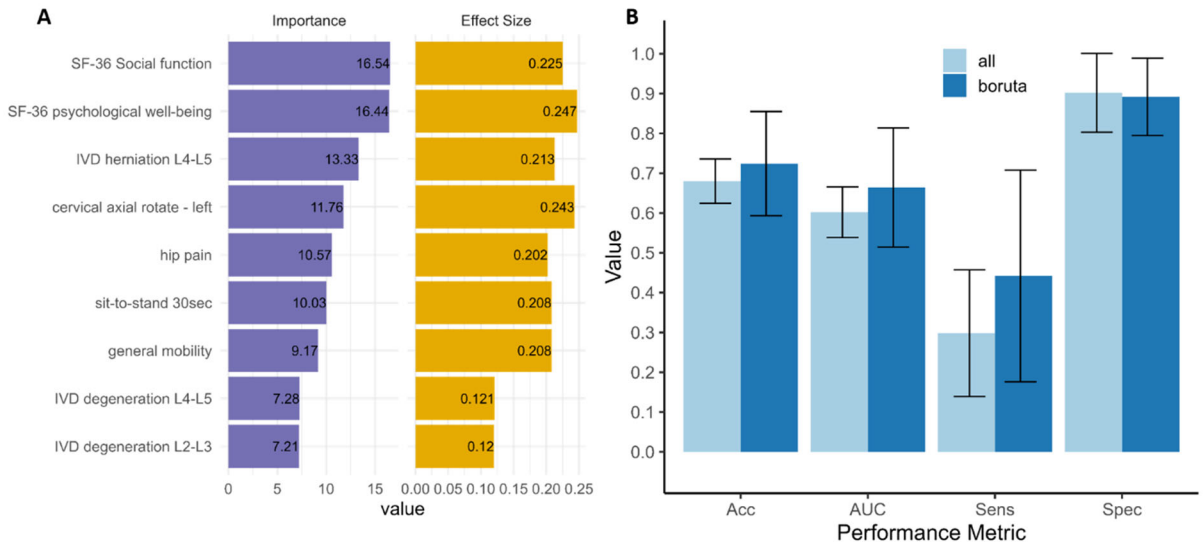

Figure 3. Most robust and important variables chronic low back pain classification performance. A – Presents a bar plot of the nine most robust variables in order of average Boruta importance score (left). The right bar plot shows the absolute effect size (Cohen's r or ω depending on data type) comparing controls and cLBP patients of the nine robust variables. B – Column plot showing RF classification performance as mean of five-fold train-test iterations in hold-out set using Boruta selected and all features. Column plot error bars represent 95% CI. IVD – intervertebral disc, SF-36 – short form 36 health status questionnaire, Acc – accuracy, AUC – Area under the receiver operating characteristic curve, Sens – sensitivity, Spec – specificity.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Discussion

This study employes a large multi-modal dataset and a machine learning workflow to demonstrate the importance of using data from different domains in cLBP patient delineation. Increasing the number of modalities generally lead to a model performance improvement although it seems the inclusion of back shape and motion data resulted in little to no performance improvement. Utilising Boruta in our iterative selection workflow resulted in considerable variable reduction across all datasets (median = 62.7%), while model performance remained comparable. This may reflect many variables showing little difference between patients and controls, or the underlying machanisms are more robustly captured by a small sub-set of variables. Both the best performing modality model (Q + C + M) and the most robust variables (Fig. 3), show the importance of measuring psychosocial factors, cervical axial rotation, general mobility, hip flexion, and lower lumbar spine disc degeneration and herniation ratings in cLBP patients.

The questionnaires probing the psychosocial factors, social function and psychological well-being, showed highest importance among all variables (Fig. 3A) and were the most important variables in the best sparse model (Boruta reduced questionnaire, Fig. 2A). Moreover, the quesionnaires represent the most cost effective modality to the examiner. Highlighting the clinical importance of psychosocial factors in cLBP diagnosis and treatment.

Social functioning describes the ability of a person to engage in social activities, which we have shown to be an important marker for delineating cLBP patients from asymptomatic controls. Using cross-sectional data from 180 chronic low back pain patients, Ge and colleagues [16] showed that these patients reported more limitations in performing (major life tasks and) social activities as compared to subjects without cLBP, even after adjusting for influencing factors,

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

such as socio-demographics, lifestyle and number of diseases. Furthermore, Tagliaferri et al. [57] were able to separate 4156 chronic back pain patients from the UK Biobank dataset into five sub-groups based on their scores of social isolation and depressive symptoms. Interestingly, increased social isolation was only a feature of three sub-groups, encompassing 26% of all back pain patients (n = 1085), while the remaining subgroups showed either no changes (4.1%; n = 776) or a reduction in social isolation scores (12%; n = 2296). This prevalence in patients with LBP may indicate that reduced social functioning was identified by some studies, while others did not find similar changes as compared to asymptomatic controls [22]. However, as levels of social function (here: social participation) were found to be correlated with self-perceived physical health status [58], a direct impact of social functioning on personal functional impairments remains feasible.

In addition to social function, the psychological health or well-being was shown to be an important variable in cLBP patient delineation. Using longitudinal data from the SwePain cohort, including 9361 participants with and without chronic pain, psychological well-being scores at baseline were able to predict pain intensity after 2 years [28]. Within this study, positive well-being was predictive of lower pain severity in participants without and with chronic pain. Similar conclusions were drawn from the comparison of back pain patients with different levels of mental distress, in which patients with higher mental distress showed, among other things, reduced psychological well-being, reduced social function, and higher severe pain than patients with lower mental distress [20]. Furthermore, patients suffering from chronic pain exhibit significantly lower quality of life scores across all sub-domains, including psychological well-being [19]. Alterations in quality of life are stronger associated with changes in social functioning and psychological well-being (via pain catastrophizing) than pain intensity it self [27], indicating the high importance of psycho-social aspects for daily living with painful conditions such as back pain.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Our findings suggest that spinal herniation and degeneration observed on MRI may contribute to pain mechanisms in the bio-psychosocial model of cLBP, consistent with previous meta-analyses [9] indicating a higher prevalence of disc herniation and degeneration on MRI in adults with cLBP compared with asymptomatic individuals. However, it is clear that MRI findings alone do not fully explain pain presence in cLBP, highlighting the need for caution when interpreting MRI results at the individual patient level. Interestingly, cervical spine rotation but not lumbar back motion assessments were shown to be robust important examinations for cLBP delineation. This contradictory finding is likely a result of two factors. First, the poor cervical rotation may be the result of neck pain that has high comorbidity with cLBP [62] and can lead to decreased axial rotation [44]. Second, poor psychological health has been associated with neck pain [32], which we found to also present high importance in patient stratification and relates to the bio-physical-psychosocial interplay present in cLBP patients [55]. A systematic review on hip mobility in LBP patients [6] showed small to no changes in hip flexion compared to controls. As all studies had less than 110 subjects, they were likely under powered to uncover the decreased mobility we show here and may represent a diagnostic test for LBP. Previous studies have shown that clinical kinematic data can effectively stratify cLBP patients into high, low and intermediate risk groups [1], suggesting that pain correlates with reduced physical function. Persistent nociceptive input from aggravated spinal joints/muscles may lead to reduced motor output and spinal cord excitability [38], potentially resulting in a reduced ability to recruit specific muscles and necessitating compensatory movement strategies. Our findings underscore that detailed movement analysis could serve as a diagnostic biomarker for LBP, potentially rivalling medical imaging in diagnostic accuracy and improving patient care by identifying sub-populations likely to respond well to specific therapies or at risk of adverse outcomes.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Classification of cLBP patients has often been conducted on relatively small sample sizes (< 200) as well as utilising a single data domain [53]. Performance of such models are subject to overfitting due to their small samples and would likely perform poorly at out-of-sample classification in unseen external datasets [43]. Furthermore, several studies have established classification models with high accuracy (> 0.8) at determining particular LBP symptoms [2,3,24,33,45,49,61,70], although these lack clinical applicability in understanding the most appropriate variables and modalities in classification of cLBP as well as the classification of the disorder in general. Classification models created using large datasets (n > 1000) either contained psychosocial and demographic variables, without imaging and physical features [40,52]. On the other hand, Jin-Heekun and colleagues [23] employed only physical features without considering important psychosocial factors, which have shown to be important in previous research [54–57] as well as in our current study (Fig. 3). Our best model (Boruta – Q + C + M) performed slightly worse than Parsaeian et al. [40] (AUC 0.693 – 0.75) and compared to Shim et al. [52] (AUC, 0.693 – 0.716), we utilised more plentiful data points per subject in a significantly smaller sample size (approximately 34x and 6x smaller respectively) to address the clinically relevant question of what modalities and variables a best suited for cLBP classification.

The multimodal nature of the data utilised and the amount of subjects that have participated in physical, imaging, and questionnaire measurements are major strengths in our current study. Utilising the "Berlin Back Study" dataset that contains more than 500 subjects in the different domains highlighted as lacking in a recent review by our group [53], which enables us to robustly investigate the importance of different modalities as well as specific variables in cLBP patient stratification. The large sample size enabled us to minimise model overfitting and data leakage through cross-validation and hold-out testing samples of good size and variable distributions. A more accurate representation of a models performance is provided by out-of-sample testing that uses a new sample population containing comparable variables. This provides a test set

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

with minimised sampling and dataset bias intrinsic within a single, even large, dataset and greatly improves the generalisability and applicability of the findings. This represents a limitation in our modelling, but as to our knowledge, there are no available open dataset containing comparable multi-modality data for out of sample testing. Furthermore, our study provides a cross-sectional investigation of cLBP, whereas a prospective design would be better suited to examine the causality of the disorder. Two different types of interviews were conducted: face-to-face interviews (clinical examination) and electronic interviews (questionnaires). The main difference is that body language, facial expressions and other non-verbal social cues are obvious to the interviewer in face-to-face interviews, whereas these aspects are absent in electronic surveys. As both surveys have advantages and disadvantages, the answers of the study participants were weighted equally in this study. A further limitation is that the high number of questions could lead to a reduction in the participants' attention and concentration.

Conclusion

Our study underscores the transformative potential of utilizing multidimensional data across various modalities for cLBP patient stratification. By integrating physical (mobility), biological (MRI), and psychological data, we pave the way for more precise, targeted, and individualized diagnostic and treatment strategies. This holistic approach not only enhances the accuracy of patient stratification but also significantly improves clinical outcomes, offering a robust framework for advancing cLBP management.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Disclosures

Funding: This study is part of the Research Unit FOR 5177 funded by the German Research Foundation (DFG), Hendrik Schmidt: SCHM 2572/11-1, SCHM 2572/12-1, SCHM 2572/13-1; Sandra Reitmaeier: RE 4292/3-1, Matthias Pumberger: PU762/1-1. The analyses and contribution from the Hochschule für Gesundheit were funded, in part, by grant number 50WK2273A (to DLB) from the German AeroSpace Center (DLR).

Conflicts of interest: All authors declare no conflict of interests.

Acknowledgments

We would like to thank all patients and healthy participants for their selfless participation in this study and the participating companies for informing their employees about this study.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Abdollahi M, Ashouri S, Abedi M, Azadeh-Fard N, Parnianpour M, Khalaf K, Rashedi E: Using a Motion Sensor to Categorize Nonspecific Low Back Pain Patients: A Machine Learning Approach. Sensors Multidisciplinary Digital Publishing Institute; 20:3600, 2020.

2. Abdullah AA, Yaakob A, Ibrahim Z: Prediction of Spinal Abnormalities Using Machine Learning Techniques. IEEE; page 1–62018.

3. Al Imran A, Rifat MRI, Mohammad R: Enhancing the classification performance of lower back pain symptoms using genetic algorithm-based feature selection. Springer; page 455–692020.

4. Anderson TW, Darling DA: Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. The Annals of Mathematical Statistics Institute of Mathematical Statistics; 23:193–212, 1952.

5. Andersson GB: Epidemiologic aspects on low-back pain in industry. Spine (Phila Pa 1976) 6:53–60, 1981.

6. Avman MA, Osmotherly PG, Snodgrass S, Rivett DA: Is there an association between hip range of motion and nonspecific low back pain? A systematic review. Musculoskeletal Science and Practice 42:38–51, 2019.

7. Barrett E, McCreesh K, Lewis J: Reliability and validity of non-radiographic methods of thoracic kyphosis measurement: a systematic review. Man Ther 19:10–7, 2014.

8. Breiman L: Random Forests. Machine Learning 45:5–32, 2001.

9. Brinjikji W, Diehn FE, Jarvik JG, Carr CM, Kallmes DF, Murad MH, Luetmer PH: MRI Findings of Disc Degeneration are More Prevalent in Adults with Low Back Pain than in Asymptomatic Controls: A Systematic Review and Meta-Analysis. AJNR Am J Neuroradiol 36:2394–9, 2015.

10. Collins GS, Reitsma JB, Altman DG, Moons KGM: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 162:55–63, 2015.

11. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, Pratt M, Ekelund U, Yngve A, Sallis JF, Oja P: International Physical Activity Questionnaire: 12-Country Reliability and Validity. Medicine & Science in Sports & Exercise 35:1381, 2003.

12. Deyo RA, Gray DT, Kreuter W, Mirza S, Martin BI: United States trends in lumbar fusion surgery for degenerative conditions. Spine (Phila Pa 1976) 30:1441–5; discussion 1446-1447, 2005.

13. Dreischarf B, Koch E, Dreischarf M, Schmidt H, Pumberger M, Becker L: Comparison of three validated systems to analyse spinal shape and motion. Sci Rep 12:10222, 2022.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

14. Fujiwara A, Tamai K, Yamato M, An HS, Yoshida H, Saotome K, Kurihashi A: The relationship between facet joint osteoarthritis and disc degeneration of the lumbar spine: an MRI study. Eur Spine J 8:396–401, 1999.

15. Gardner B, Abraham C, Lally P, de Bruijn G-J: Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. International Journal of Behavioral Nutrition and Physical Activity 9:102, 2012.

16. Ge L, Pereira MJ, Yap CW, Heng BH: Chronic low back pain and its impact on physical function, mental health, and health-related quality of life: a cross-sectional study in Singapore. Sci Rep 12:20040, 2022.

17. Grotle M, Småstuen MC, Fjeld O, Grøvle L, Helgeland J, Storheim K, Solberg TK, Zwart J-A: Lumbar spine surgery across 15 years: trends, complications and reoperations in a longitudinal observational study from Norway. BMJ Open British Medical Journal Publishing Group; 9:e028743, 2019.

18. Guermazi M, Ghroubi S, Kassis M, Jaziri O, Keskes H, Kessomtini W, Ben Hammouda I, Elleuch M-H: [Validity and reliability of Spinal Mouse to assess lumbar flexion]. Ann Readapt Med Phys 49:172–7, 2006.

19. Hadi MA, McHugh GA, Closs SJ: Impact of Chronic Pain on Patients' Quality of Life: A Comparative Mixed-Methods Study. J Patient Exp 6:133–41, 2019.

20. Hnatešen D, Pavić R, Radoš I, Dimitrijević I, Budrovac D, Čebohin M, Gusar I: Quality of Life and Mental Distress in Patients with Chronic Low Back Pain: A Cross-Sectional Study. Int J Environ Res Public Health 19:10657, 2022.

21. Holm S: A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics [Board of the Foundation of the Scandinavian Journal of Statistics, Wiley]; 6:65–70, 1979.

22. Iguti AM, Guimarães M, Barros MBA: Health-related quality of life (SF-36) in back pain: a population-based study, Campinas, São Paulo State, Brazil. Cad Saude Publica 37:e00206019, 2021.

23. Jin-Heeku: Analysis of sitting posture using wearable sensor data and support vector machine model. Medico-Legal Update 1:334–8, 2018.

24. Karabulut EM, Ibrikci T: Effective automated prediction of vertebral column pathologies based on logistic model tree with SMOTE preprocessing. Journal of Medical Systems 38:50, 2014.

25. Kraemer J: Natural course and prognosis of intervertebral disc diseases. International Society for the Study of the Lumbar Spine Seattle, Washington, June 1994. Spine (Phila Pa 1976) 20:635–9, 1995.

26. Kursa MB, Rudnicki WR: Feature Selection with the Boruta Package. Journal of Statistical Software 36:1–13, 2010.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

27. Lamé IE, Peters ML, Vlaeyen JWS, Kleef M v, Patijn J: Quality of life in chronic pain is more associated with beliefs about pain, than with pain intensity. Eur J Pain 9:15–24, 2005.

28. Larsson B, Dragioti E, Gerdle B, Björk J: Positive psychological well-being predicts lower severe pain in the general population: a 2-year follow-up study of the SwePain cohort. Ann Gen Psychiatry 18:8, 2019.

29. Lee W, Alexeyenko A, Pernemalm M, Guegan J, Dessen P, Lazar V, Lehtiö J, Pawitan Y: Identifying and Assessing Interesting Subgroups in a Heterogeneous Population. Biomed Res Int 2015:462549, 2015.

30. Lonsdale C, Hodge K, Rose EA: The behavioral regulation in sport questionnaire (BRSQ): Instrument development and initial validity evidence. Journal of Sport & Exercise Psychology US: Human Kinetics; 30:323–55, 2008.

31. Lötsch J, Ultsch A: Machine learning in pain research. Pain 159:623–30, 2018.

32. Mansfield M, Thacker M, Taylor JL, Bannister K, Spahr N, Jong ST, Smith T: The association between psychosocial factors and mental health symptoms in cervical spine pain with or without radiculopathy on health outcomes: a systematic review. BMC Musculoskeletal Disorders 24:235, 2023.

33. Mathew B, Norris D, Hendry D, Waddell G: Artificial intelligence in the diagnosis of low-back pain and sciatica. Spine 13:168–72, 1988.

34. Meucci RD, Fassa AG, Faria NMX: Prevalence of chronic low back pain: systematic review. Rev Saude Publica 49:1, 2015.

35. Meyerding HW: Spondyloptosis. Surgery, Gynecology & Obstetrics :371–7, 1932.

36. Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR: Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. Radiology 166:193–9, 1988.

37. Mwangi B, Tian TS, Soares JC: A review of feature reduction techniques in neuroimaging. Neuroinformatics 12:229–44, 2014.

38. Nijs J, Daenen L, Cras P, Struyf F, Roussel N, Oostendorp RAB: Nociception affects motor output: a review on sensory-motor interaction with focus on clinical implications. Clin J Pain 28:175–81, 2012.

39. Noroozi Z, Orooji A, Erfannia L: Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction. Sci Rep Nature Publishing Group; 13:22588, 2023.

40. Parsaeian M, Mohammad K, Mahmoudi M, Zeraati H: Comparison of logistic regression and artificial neural network in low back pain prediction: second national health survey. Iranian Journal of Public Health 41:86, 2012.

41. Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, Boccia S: Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. Eur J Public Health 29:23–7, 2019.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

42. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N: Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine 26:1873–8, 2001.

43. Rajput D, Wang W-J, Chen C-C: Evaluation of a decided sample size in machine learning applications. BMC Bioinformatics 24:48, 2023.

44. Rampazo ÉP, da Silva VR, de Andrade ALM, Back CGN, Madeleine P, Arendt-Nielsen L, Liebano RE: Sensory, Motor, and Psychosocial Characteristics of Individuals With Chronic Neck Pain: A Case Control Study. Physical Therapy 101:pzab104, 2021.

45. Riveros NAM, Espitia BAC, Pico LEA: Comparison between K-means and self-organizing maps algorithms used for diagnosis spinal column patients. Informatics in Medicine Unlocked 16:100206, 2019.

46. Roland M, Fairbank J: The Roland–Morris Disability Questionnaire and the Oswestry Disability Questionnaire. Spine 25:3115, 2000.

47. Roldán-Jiménez C, Pérez-Cruzado D, Neblett R, Gatchel R, Cuesta-Vargas A: Central Sensitization in Chronic Musculoskeletal Pain Disorders in Different Populations: A Cross-Sectional Study. Pain Med 21:2958–63, 2020.

48. Rusu AC, Kreddig N, Hallner D, Hülsebusch J, Hasenbring MI: Fear of movement/(Re)injury in low back pain: confirmatory validation of a German version of the Tampa Scale for Kinesiophobia. BMC Musculoskelet Disord 15:280, 2014.

49. Sandag GA, Tedry NE, Lolong S: Classification of lower back pain using K-Nearest Neighbor algorithm. IEEE; page 1–52018.

50. Schizas C, Theumann N, Burn A, Tansey R, Wardlaw D, Smith FW, Kulik G: Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the dural sac on magnetic resonance images. Spine (Phila Pa 1976) 35:1919–24, 2010.

51. Shilo S, Rossman H, Segal E: Axes of a revolution: challenges and promises of big data in healthcare. Nat Med Nature Publishing Group; 26:29–38, 2020.

52. Shim J-G, Ryu K-H, Cho E-A, Ahn JH, Kim HK, Lee Y-J, Lee SH: Machine Learning Approaches to Predict Chronic Lower Back Pain in People Aged over 50 Years. Medicina Multidisciplinary Digital Publishing Institute; 57:1230, 2021.

53. Tagliaferri SD, Angelova M, Zhao X, Owen PJ, Miller CT, Wilkin T, Belavy DL: Artificial intelligence to improve back pain outcomes and lessons learnt from clinical classification approaches: three systematic reviews. npj Digital Medicine 3:93, 2020.

54. Tagliaferri SD, Fitzgibbon BM, Owen PJ, Miller CT, Bowe SJ, Belavy DL: Brain structure, psychosocial, and physical health in acute and chronic back pain: a UKBioBank study. Pain 163:1277–90, 2022.

55. Tagliaferri SD, Ng S-K, Fitzgibbon BM, Owen PJ, Miller CT, Bowe SJ, Belavy DL: Relative contributions of the nervous system, spinal tissue and psychosocial health to non-specific low back pain: Multivariate meta-analysis. Eur J Pain , 2021.

56. Tagliaferri SD, Owen PJ, Miller CT, Angelova M, Fitzgibbon BM, Wilkin T, Masse-Alarie H, Van Oosterwijck J, Trudel G, Connell D, Taylor A, Belavy DL: Towards data-

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

driven biopsychosocial classification of non-specific chronic low back pain: a pilot study. Sci Rep Nature Publishing Group; 13:13112, 2023.

57. Tagliaferri SD, Wilkin T, Angelova M, Fitzgibbon BM, Owen PJ, Miller CT, Belavy DL: Chronic back pain sub-grouped via psychosocial, brain and physical factors using machine learning. Sci Rep 12:15194, 2022.

58. Takeyachi Y, Konno S, Otani K, Yamauchi K, Takahashi I, Suzukamo Y, Kikuchi S: Correlation of low back pain with functional status, general health perception, social participation, subjective happiness, and patient satisfaction. Spine (Phila Pa 1976) 28:1461–6; discussion 1467, 2003.

59. Tanaka K, Nishigami T, Mibu A, Imai R, Manfuku M, Tanabe A: Combination of Pain Location and Pain Duration is Associated with Central Sensitization-Related Symptoms in Patients with Musculoskeletal Disorders: A Cross-Sectional Study. Pain Pract 21:646–52, 2021.

60. Topalidou A, Tzagarakis G, Souvatzis X, Kontakis G, Katonis P: Evaluation of the reliability of a new non-invasive method for assessing the functionality and mobility of the spine. Acta Bioeng Biomech 16:117–24, 2014.

61. Vaughn ML, Cavill SJ, Taylor SJ, Foy MA, Fogg AJ: Direct explanations for the development and use of a multi-layer perceptron network that classifies low-back-pain patients. International Journal of Neural Systems 11:335–47, 2001.

62. von der Lippe E, Krause L, Porst M, Wengler A, Leddin J, Müller A, Zeisler M-L, Anton A, Rommel A: Prevalence of back and neck pain in Germany. Results from the BURDEN 2020 Burden of Disease Study. J Health Monit 6:2–14, 2021.

63. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP: The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. Preventive Medicine 45:247–51, 2007.

64. Von Korff M, Ormel J, Keefe FJ, Dworkin SF: Grading the severity of chronic pain. PAIN 50:133, 1992.

65. Waddell G, Newton M, Henderson I, Somerville D, Main CJ: A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. PAIN 52:157, 1993.

66. Ware JE, Sherbourne CD: The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 30:473–83, 1992.

67. World Medical Association: World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 310:2191–4, 2013.

68. Wright MN, Ziegler A: ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software 77:1–17, 2017.

69. Yoshihara H, Yoneoka D: National trends in the surgical treatment for lumbar degenerative disc disease: United States, 2000 to 2009. Spine J 15:265–71, 2015.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.29.24316352; this version posted October 30, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

70. Zhang W, Chen Z, Su Z, Wang Z, Hai J, Huang C, Wang Y, Yan B, Lu H: Deep learning-based detection and classification of lumbar disc herniation on magnetic resonance images. JOR SPINE 6:e1276, 2023.