

1 Integrating Multidimensional Data Analytics for Precision Diagnosis 2 of Chronic Low Back Pain

3 Sam Vickery^{1*}, Frederick Junker¹, Rebekka Döding¹, Daniel L Belavy¹, Maia Angelova^{2,3},
4 Chandan Karmakar³, Louis Alexander Becker^{4,5}, Nima Taheri^{4,5}, Matthias Pumberger⁴,
5 Sandra Reitmaier⁵, Hendrik Schmidt⁵

6
7 ¹ Division of Physiotherapy / Department of Applied Health Sciences | Hochschule für
8 Gesundheit (University of Applied Sciences) | Bochum | Germany

9 ² Aston Digital Futures Institute | Aston University| Birmingham | United Kingdom

10 ³ School of Information Technology | Deakin University | Geelong | Australia

11 ⁴ Center for Musculoskeletal Surgery, Charité – Universitätsmedizin Berlin, Berlin, Germany

12 ⁵ Julius Wolff Institut, Berlin Institute of Health - Charité at Universitätsmedizin Berlin |
13 Berlin | Germany

15 **Abstract**

16 Low back pain (LBP) is a leading cause of disability worldwide, with up to 25% of cases
17 become chronic (cLBP). Optimal diagnostic tools for cLBP remains unclear. Here we leveraged
18 a comprehensive multi-dimensional data-set and machine learning-based feature importance
19 selection to identify the most effective diagnostic tools for cLBP patient stratification. The
20 dataset included questionnaire data, clinical and functional assessments, and spino-pelvic
21 magnetic resonance imaging (MRI), encompassing a total of 144 parameters from 1,161 adults
22 with (n=512) and without cLBP (n=649). Boruta and random forest were utilised for variable
23 importance selection and cLBP classification respectively. Boruta feature selection led to
24 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NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

25 comparable to using all variables across all modality datasets. Multi-modality models
26 performed better than single modality models. Boruta selected key variables from
27 questionnaire, clinical, and MRI data were the most effective in distinguishing cLBP patients
28 from controls with an AUC (area under the receiver operating characteristic curve) of 0.699
29 (95% confidence interval [CI], 0.669 – 0.729). The most robust features (n=9) across the whole
30 dataset identified were psychosocial factors, neck and hip mobility, as well as lower lumbar
31 disc herniation and degeneration. These critical variables (AUC = 0.664, 95% CI = 0.514 –
32 0.814) outperformed all parameters (AUC = 0.602, 95% CI = 0.538 – 0.666) in an unseen
33 holdout dataset, demonstrating superior patient delineation. Paving the way for targeted
34 diagnosis and personalized treatment strategies, ultimately enhancing clinical outcomes for
35 cLBP patients.

36

37 **Key words**

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

40

41 **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 |

42

43 **Glossary**

44 AUC - area under the receiver operating characteristic curve

45 BMI – body mass index

46 C – Clinical examination

47 CI – confidence interval

48 cLBP – Chronic low back pain

49 FABQ - Fear-Avoidance Belief Questionnaire

50 IPAQ - International Physical Activity Questionnaire

51 IVD – intervertebral disc

52 M – MRI

53 MET – metabolic equivalents

54 MRI - magnetic resonance imaging

55 Q – Questionnaire

56 RMDQ - Disability Questionnaire: Roland and Morris

57 S - Back shape and function

58 SF-36 - Short-form 36 Health Status Questionnaire

59 SRBAI - Self-Report Behavioural Automaticity Index

60 TSK-GV - Tampa Scale for Kinesiophobia

61

62 **Introduction**

63

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

72

73 Addressing chronic pain as a multidimensional phenomenon is a major challenge in practice.
74 This challenge is exacerbated ^{41,51} by the number of potential factors that may differentiate pain
75 from pain-free and a lack of consensus as to what key parameters are. Studies have shown that
76 permanent changes in central nervous system sensitisation contribute to the persistence of pain
77 ^{47,59}, as opposed to acute pain, which typically resolves as tissue heals Previous meta-analyses
78 ⁵⁵ underscored that experience of cLBP is related to physical, psychological and social
79 elements, highlighting the need for a multifaceted approach to pain assessment.

80

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

87

88 Machine learning algorithms provide models for identifying distinct subgroups that elucidate
89 the occurrence and characteristics of a disease^{29,31}. A previous systematic review by our team
90 using machine learning applications in LBP⁵³ highlighted that a narrow range of mechanistic
91 domains have been assessed, and sample sizes in these studies were consistently small, ranging
92 up to only 171 participants. Consequently, probing limited data and modalities limits the
93 robustness and applicability of such models. Through gathering many data points across
94 multiple modalities one can ascertain which variables and modalities are the most informative
95 at distinguishing cLBP patients from asymptomatic controls. Reducing the number of variables
96 to those that are most informative has been previously employed in predictive and classification
97 modelling to improve accuracy^{37,39}. This approach can be applied as the main outcome and not
98 only in model preprocessing, in order to obtain a data-driven decision on the most important
99 variables in multi-dimensional clinical data. Such a systematic data-informed investigation of
100 back pain diagnosis in a large multi-modality sample is lacking to help inform future studies in
101 selecting which data to acquire and for clinicians in which tests to conduct. Therefore, the aim
102 of this study was to identify and compare the most informative domains and variables in
103 delineating patients with and without cLBP utilising a large multi-modality dataset.
104

105 Materials and methods

106 Study design

107 The prospective cross-sectional study draws its data from the ongoing “Berliner Rückenstudie”
108 (“Berlin Back Study”; https://spine.charite.de/en/spine_study/; running time: 01/01/2022 to
109 31/12/2025), which was registered at the German Clinical Trial Register (DRKS-ID:
110 [DRKS00027907](https://www.drks.de/DRKS000027907)). Recruitment procedures vary from local promotion (i.e., postal flyers, notice
111 boards, internet approaches, and social media) at the Charité-Universitätsmedizin Berlin, in
112 the general public (i.e., newspapers, magazines, podcasts) to cooperation with local companies,
113 administrative authorities, and word-of-mouth. The protocol is in accordance with the Helsinki
114 Declaration of ethical principles⁶⁷ and has been approved by the Ethics Committee of the
115 Charité – Universitätsmedizin Berlin (registry numbers: EA4/011/10, EA1/162/13). Written
116 informed consent was obtained from all participants. The STROBE guideline⁶³ (Supplemental
117 Table 1) and TRIPOD statement¹⁰ (Supplemental Table 2) for prediction model development
118 were used to report this study. Data collection started on 1st January 2022 and cut-off for
119 inclusion in the current analysis was 5th April 2024. Data collection occurred in a research
120 centre within a university-hospital.

121

122 Participants

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

127 to a clinical examination by a trained medical doctor, which included physical examinations,
128 questions, as well as a back shape and function test. The examinations and questionnaires took
129 a total of 90 minutes to complete. Additionally, participants were offered a magnetic resonance
130 imaging (MRI) within 14 days of the spino-pelvic region. During the clinical assessment the
131 participants were classified by the clinician as asymptomatic (no back pain), symptomatic
132 (cLBP), or previously suffering from cLBP. To ensure a more robust cLBP patient classification
133 previous symptomatic subjects were removed from the sample. Furthermore, participants who
134 revoked their inclusion in the study and those who were missing demographic data; age, sex,
135 body mass index (BMI), and patient status were removed. This resulted in a study sample of
136 1161 subjects that included 649 asymptomatic (19 – 72 years old, mean age = 40.7 ± 12.6 ,
137 females = 353) and 512 cLBP (19 – 65 years old, mean age = 43.5 ± 11.7 , females = 306)
138 participants. This sample was sub-divided into four modalities; questionnaires (Q), clinical
139 physical assessment (C), back shape and function (S), and MRI (M). Only participants with all
140 data within a particular dataset modality were considered for modelling and no interpolation
141 was conducted.

142

143 Quantitative variables and data collection

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

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

153 154 Questionnaires

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

163 The following questionnaires were completed within 30 minutes:

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

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

178

179 Demographic data

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

184

185 Clinical examination

186 The clinical examination included the evaluation of organ functions (inspection, palpation,
187 percussion, and auscultation), the general impression, and the vital parameters of the patient
188 (temperature, heart rate, blood pressure, etc.). It was performed by an experienced orthopaedic
189 consultant. The neurological status was assessed by the examination of the coordination,
190 reflexes, sensitivity, and motor function. The evaluation of the functional parameters, that is,
191 the assessment of posture, shape, orientation, and movement of the lumbar spine and pelvis,
192 was based on current clinical standards (e.g., Ott and Schober test, 3-step hyperextension test,
193 passive lumbar extension test, etc.) and self-assessment by the persons investigated. Data were
194 documented according to their dimension using distances in cm, degrees of angle, and number
195 of repetitions per defined time interval or bivalent whether pain provocation occurred. Self-
196 assessment of functional restrictions of the back was recorded according to a scale from 1 (best)
197 – 10 (worst).

198

199 Back shape and function

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

209

210 Spino-pelvic MRI

211 MRIs were conducted using a 1.5 MRI scanner. Following sequences were evaluated: 1) Sag
212 T1 (4 mm slices), 2) Sag T2 (4 mm slices), 3) Cor STIR-T2 (4 mm slices) and 4) Axial T2
213 (3 mm slices). MRIs were evaluated for intervertebral disc degeneration (Pfirrmann
214 classification ⁴²), disc herniation (Kramer classification ²⁵), facet joint arthrosis (Fujiwara
215 classification ¹⁴), osteochondrosis intervertebralis ³⁶, spondylolisthesis (Meyerding
216 classification ³⁵ and spinal canal stenosis (Schizas classification ⁵⁰) at each level of the lumbar
217 spine. The spino-pelvic MRI evaluation was performed blinded by two spine surgeons and a
218 radiologist, all of whom have many years of experience in the evaluation of spinal pathologies.
219 The inter-rater reliability was good-to-excellent for all measurement parameters.

220

221 Data storage

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

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

232

233 Potential sources of bias and minimisation

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

243

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

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

253

254 **Outcome**

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

261

262 **Data handling, preprocessing, cleaning and missing data**

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

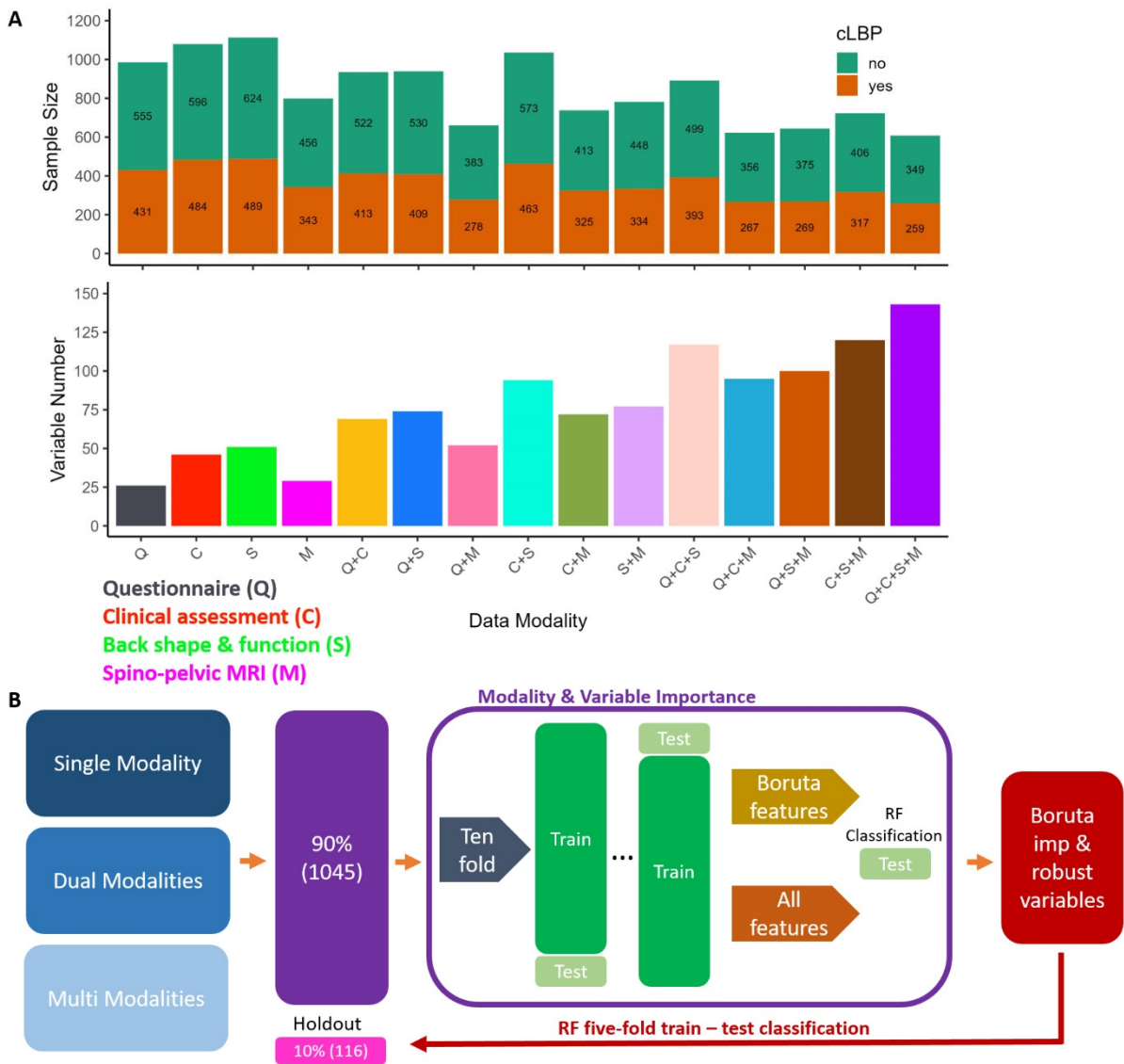
270

271 Total and sub-scores for the self-administered questionnaires were used after removal of pain
272 biased questionnaires (see *Potential sources of bias and minimisation*). This includes the SF-
273 36⁶⁶, the IPAQ¹¹, the SRBAI¹⁵, and the BRSQ³⁰. The SF-36 was used to collect statements

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

292
293 Modelling preprocessing was conducted by checking variables within each modality for very
294 low variance and collinearity. Features with close to zero (variance < 1) variance were removed.
295 Furthermore, features presenting a Spearman correlation greater than 0.9 were also removed to
296 reduce collinearity between features. The decision on which of the correlated features to
297 removed, was the feature showing less correlation to the target, cLBP patient status. Following
298 these cleaning steps, single data modalities were joined with demographic data and then for the
299 dual- and multi-modalities datasets join with other datasets. Subjects having any missing values

300 were removed from our analyses and therefore no imputation was conducted. Following
 301 cleaning and preprocessing modality dataset presented different number of subjects and features
 302 with a similar age and sex distribution across cLBP patients and asymptomatic controls (Fig.
 303 1A).
 304



305
 306 **Figure 1. Modality dataset distributions and machine learning workflow.** A – Top shows the chronic
 307 low back pain (cLBP) sample size distribution across all 15 dataset modalities. Bottom presents the
 308 number of variables used for cLBP classification and variable importance selection across the 15 dataset
 309 modalities. B – Represents the machine learning workflow implemented to compare the different
 310 modalities and determine the most important variables for cLBP patient delineation using a random
 311 forest binary classification algorithm for training and testing.
 312

313 Univariate statistics

314 The univariate statistics were carried out separately for continuous, ordinal and nominal data to
315 compare patients suffering from cLBP against asymptomatic controls using R (version 4.3.1;
316 www.r-project.org). As most continuous variables did not follow a normal distribution
317 according to the Anderson-Darling test⁴, we implemented the non-parametric Wilcoxon-Mann-
318 Whitney test to determine significant difference between cLBP patients and asymptomatic
319 controls for ordinal and continuous data. Hence, u-values, z-values, r-value (effect sizes), as
320 well as the p-value are reported from. Nominal data were compared using the Chi-Square test
321 and reported with Chi² values, Cohen's ω -values (effect size), and p-values. Statistical
322 significance was determined at $p \leq 0.05$ following family wise error (FWE)²¹ correction for
323 multiple comparisons within each modality (demographics, questionnaires, clinical
324 examinations, superficial spine morphology and motion, and spino-pelvic MRI).

325 Machine learning

326 Boruta feature selection

327 We used the Boruta method²⁶ for importance feature selection. Boruta utilises random forest
328 (RF) classification algorithm⁸ with both the real variables and set of 'dummy' or 'shadow'
329 variables, that are created by shuffling the feature values. This creates random variables
330 (dummy features) that have the same distribution as the original features, although represent
331 the classification accuracy of this feature randomly sampled. As these dummy features
332 represent random noise, they have had their possible correlation to the target (cLBP) removed.
333 All real and dummy features are used to classify cLBP and feature importance is calculated.
334 Feature importance is calculated as the Z-score of the mean decrease in classification accuracy
335 following the removal of this feature from the model. The importance of the dummy features
336 can be used as a reference to test the feature importance of the real features. Through an iterative

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

345

346 Random forest classification algorithm

347 We utilised RF ⁸ implemented using the ranger package ⁶⁸ in R (version 4.3.1, [www.r-](http://www.r-project.org)
348 project.org) to classify cLBP patients and pain-free controls. As RF provides feature importance
349 measures and can handle categorical, ordinal, and continuous data it represents the ideal choice
350 to deal with the different data types present in the Berlin Back dataset. Ten-fold cross validation
351 was conducted during model training and hyper-parameter tuning utilising 1000 tree RF. We
352 conducted hyper-parameter tuning using a tuning parameter search grid, which contained
353 number of variables to be sampled at each split (mtry) of 1 – square root of the number of
354 features, and a minimum node size of 5 and 10. Therefore, using a gini split rule a grid search
355 was conducted with all combination of hyper-parameters to determine the best. A ten-fold train-
356 test loop (Fig. 1B) was conducted to determine the cLBP classification performance within each
357 of the 15 datasets utilising all and Boruta selected variables independently. Model performance
358 was calculated as follows; accuracy = $(TP + TN) / (TP + FP + TN + FN)$, sensitivity = $TP / (TP$
359 $+ FN)$, specificity = $TN / (TN + FP)$, and AUC (area under the receiver operating characteristic
360 (ROC) curve). Whereby, TP – true positive, TN – true negative, FP – false positive, and FN –

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

363

364 Robust variable selection workflow

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

380

381 Data and code availability

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

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

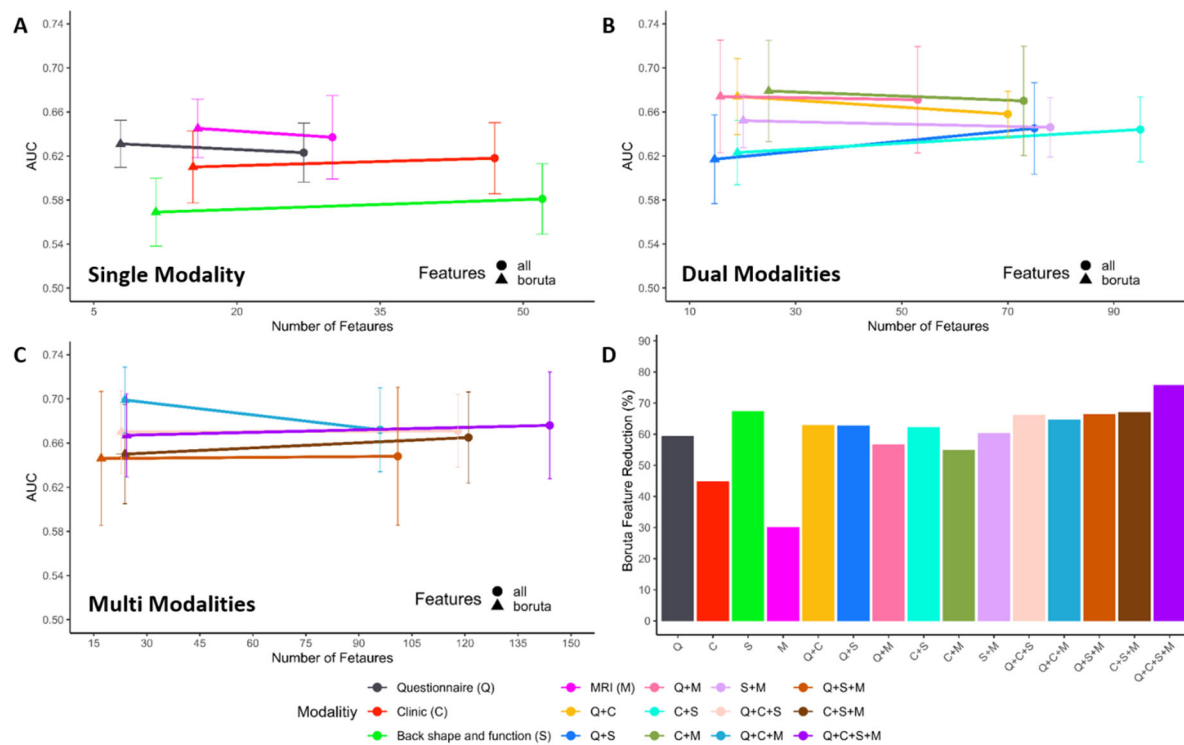
388

389 Results

390 Chronic low back pain classification

391 The best single modality for cLBP classification was Boruta selected features MRI (Fig. 2A)
392 with a mean AUC of 0.645 with a 95% confidence interval (CI) of 0.618 – 0.672 and accuracy
393 of 0.657 (95% CI, 0.636 – 0.678). The Boruta selected questionnaire dataset modality produced
394 only minimally worse classification performance (mean AUC = 0.631, 95% CI = 0.610 – 0.652)
395 than Boruta reduced and all MRI variables (mean AUC = 0.637, 95% CI = 0.599 – 0.675) using
396 the least amount of variables (mean = 8) across all modality datasets (Fig. 2A-C). Dual and
397 multi modalities generally performed better than single modality models in cLBP classification.
398 Questionnaire, clinical physical assessment, and MRI (Q + C + M, Fig. 2C) modality with
399 Boruta selected features represents the best performing modality model with a mean AUC of
400 0.699 (95% CI, 0.669 – 0.729) and a mean accuracy of 0.709 (95% CI, 0.679 – 0.739).
401 Moreover, this model showed the highest sensitivity (mean = 0.622, 95% CI = 0.568 – 0.676).
402 The modality model showing the highest specificity was using all features and all modalities
403 (Q + C + S + M, mean = 0.840, 95% CI = 0.799 – 0.881). The three best dual modalities models,
404 C + M (mean AUC = 0.679, 95% CI = 0.633 – 0.725), Q + M (mean AUC = 0.674, 95% CI =
405 0.623 – 0.725), and Q + C (mean AUC = 0.674, 95% CI = 0.639 – 0.709) all with Boruta
406 selected features (Fig. 2B), performed only slightly worse than the best model (Boruta – Q + C
407 + M). Overall back shape and function dataset using Boruta selected features produces the worst
408 classification performance (Fig. 2A, AUC = 0.569, 95% CI = 0.538 – 0.60). Additionally, the
409 dual modalities continuing back shape and function data always performed worse than those
410 without (Fig. 2B), when using both all and Boruta selected features. Classification performance
411 metrics across all models is provided in Supplementary Table 7.

412



413

414 **Figure 2. Boruta feature reduction performance.** A – C shows RF classification model performance
 415 (AUC) following the reduction of features using Boruta and all variables in the single, dual, and multi
 416 data modalities respectively. This shows the change in performance follow feature reduction. Error bars
 417 represent 95% CI in AUC over 10-fold train-test splits. D – Shows the amount of feature reduction by
 418 using Boruta as a percentage of the total number of features within each modality dataset.
 419

420 Boruta feature importance

421 Boruta feature importance selection resulted in a median of 62.7% reduction in the number of
 422 variables across all 15 datasets (Fig. 2D). This large reduction in variables from Boruta
 423 performed comparable (eight slightly worse and seven better) compared to using all features,
 424 with high overlap in confidence intervals. Furthermore, three of the top five performing
 425 modality models where those employing Boruta selected features (Supplementary Table 7). The
 426 greatest performance improvement following Boruta feature selection was found in the Q + C
 427 + M datasets with an AUC increase of 0.270 and an average reduction of 72.2 features (Fig.
 428 2C). The smallest feature reduction was shown in MRI (30%) with an AUC increase of 0.008,

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

431
432 Boruta selected variables represented the best patient delineation performance in single, dual,
433 and multi-modality datasets (Fig. 2A-C). MRI was found to be the best single modality model
434 using Boruta selected features (Fig. 2A). The most important variables (Supplementary Table
435 11) were intervertebral disc (IVD) herniation L4 – L5, spinal canal width L2, IVD degeneration
436 L3 – L4 and L4 – L5, and spinal canal width L1, showing a mean importance across the ten
437 iterations of 19.49, 9.94, 9.46, 9.14, and 8.13 respectively. The best dual modalities cLBP
438 patient stratification model, Boruta selected C + M (Fig. 2B), showed the second highest AUC
439 (mean = 0.679). The MRI variable IVD herniation L4 – L5 was found to be the most important
440 (mean = 13.72), with clinical mobility assessments of the hip, cervical spine, and the whole
441 body showing high importance (Supplementary Table 15). The most important and robust
442 variables of the best performing model (Boruta – Q + C + M, Fig. 2C) contained assessments
443 from all three modalities (Supplementary Table 18). The Short-form 36 Health Status
444 Questionnaires (SF-36) psychological well-being, SF-36 social function, and hip pain presented
445 a mean importance of 13.36, 13.07, and 9.71 respectively. The clinical assessment, cervical
446 axial rotation (left) and MRI variable IVD herniation L4 – L5 each showed an importance of
447 9.47 and 9.12. These represent the top five most important variables and the rest are provided
448 in Table 3. The questionnaire Boruta reduced model provides a sparse model with decent
449 performance, meaning the modality performs comparably well with a small amount of data. On
450 average, the questionnaire Boruta model used 8 variables with a mean AUC reduction of 0.068
451 compared to the best model (Q + C + M). The most robust and important features
452 (Supplementary Table 8) were SF-36 social function (mean = 21.42), SF-36 psychological well-
453 being (mean = 20.02), hip pain (mean = 15.0), smoking in pack years (mean = 11.8), and family

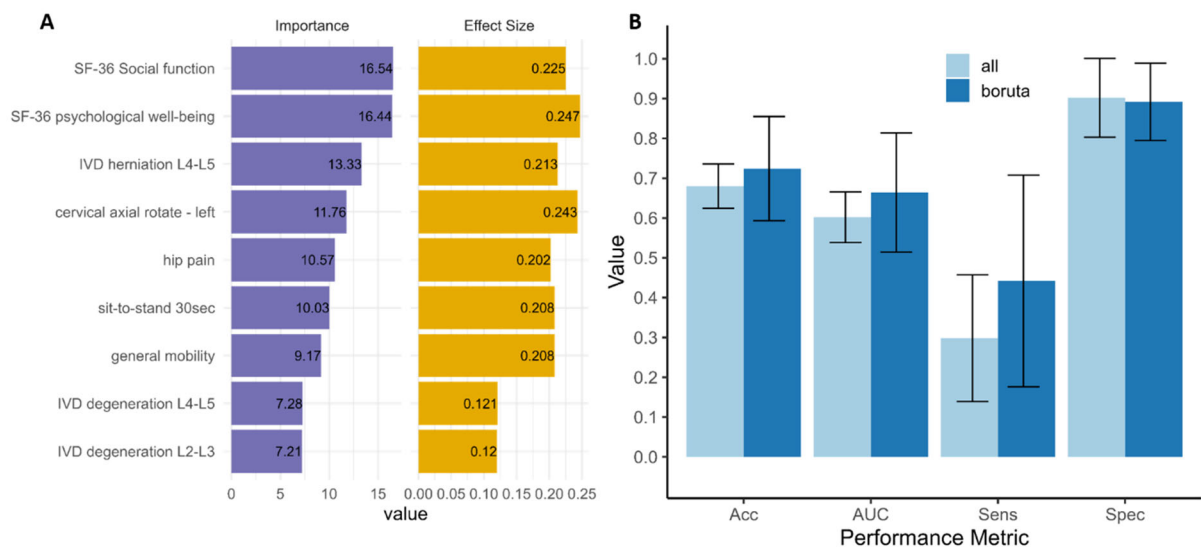
454 history of back pain (mean = 7.27). All Boruta selected important variables across the additional
455 datasets are provided in Supplementary Table 8 - 22.

456

457 Most robust and important variables

458 To select the most robust and important variables for cLBP patient classification, the percentage
459 of selection and average importance score was calculated across the 15 datasets (Supplementary
460 Table 23). The features that were selected at every opportunity (100%) across the multiple
461 iterations and datasets were defined as most important and solely implemented in cLBP
462 classification compared to all variables. This resulted in nine robust and important variables
463 (Fig. 3A). These variables represented psychosocial factors, IVD herniation and degeneration
464 of the lower lumbar spine, presence of hip pain, as well as mobility of the neck and general
465 mobility of the whole body. The performance of these nine variables were compared to all
466 variables utilising the hold-out dataset (Fig. 1) in a five-fold train-test workflow to provide an
467 unbiased comparison. The best nine variables showed better mean accuracy (Fig. 3B, Boruta =
468 0.724, 95% CI = 0.593 – 0.855, All = 0.680, 95% CI = 0.625 – 0.736), AUC (Boruta = 0.664,
469 95% CI = 0.514 – 0.814, All = 0.602, 95% CI = 0.538 – 0.666), and sensitivity (Boruta = 0.442,
470 95% CI = 0.176 – 0.708, All = 0.298, 0.139 – 0.457), while all variables provided better
471 specificity (Boruta = 0.892, 95% CI = 0.795 – 0.989, All = 0.902, 95% CI = 0.803 – 1.0).
472 Moreover, all these nine variables showed significant univariate statistically significant
473 differences between cLBP patients and asymptomatic controls in the questionnaire, clinical,
474 and MRI datasets (Fig. 3A). Utilising a Wilcoxon-Mann-Whitney test revealed reduced scores
475 of the SF36 for social function ($u = 183751$, $z = -7.08$, $p < 0.001$, effect size $r = -0.225$) and
476 psychological well-being ($u = 178506.5$, $z = -7.74$, $p < 0.001$, $r = -0.247$) in people suffering
477 from cLBP. In addition, the occurrence of hip pain was also altered comparing people with and
478 without cLBP ($\chi^2(3, N=986) = 40.17$, $p = 0.004$, $\omega = 0.202$). Comparing clinical examinations

479 further revealed reduced cervical axial rotation to the left ($u = 221595$, $z = -7.98$, $p < 0.001$, $r =$
 480 -0.243), reduced sit to stand 30 second repetition ($u = 226818.5$, $z = -6.83$, $p < 0.001$, $r = -$
 481 0.208), as well as altered general mobility ($\chi^2 (2, N=1080) = 46.81$, $p = 0.006$, $\omega = 0.208$) in
 482 cLBP patients. Regarding MRI investigations, increased IVD degeneration at L2-L3 ($u =$
 483 147168 , $z = 3.38$, $p = 0.011$, $r = 0.120$) and L4-L5 ($u = 147691$, $z = 3.43$, $p = 0.010$, $r = 0.121$),
 484 as well as increased disc herniation at L4-L5 ($u = 153470.5$, $z = 6.02$, $p < 0.001$, $r = 0.213$) were
 485 found in people suffering from cLBP compared to asymptomatic controls. Univariate statistical
 486 results for all variable can be found in Supplementary Table 24 – 31.
 487



488
 489 **Figure 3. Most robust and important variables chronic low back pain classification performance.**
 490 A – Presents a bar plot of the nine most robust variables in order of average Boruta importance score
 491 (left). The right bar plot shows the absolute effect size (Cohen’s r or ω depending on data type)
 492 comparing controls and cLBP patients of the nine robust variables. B – Column plot showing RF
 493 classification performance as mean of five-fold train-test iterations in hold-out set using Boruta selected
 494 and all features. Column plot error bars represent 95% CI. IVD – intervertebral disc, SF-36 – short form
 495 36 health status questionnaire, Acc – accuracy, AUC – Area under the receiver operating characteristic
 496 curve, Sens – sensitivity, Spec – specificity.
 497

498 Discussion

499

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

512

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

518

519 Social functioning describes the ability of a person to engage in social activities, which we have
520 shown to be an important marker for delineating cLBP patients from asymptomatic controls.
521 Using cross-sectional data from 180 chronic low back pain patients, Ge and colleagues¹⁶
522 showed that these patients reported more limitations in performing (major life tasks and) social
523 activities as compared to subjects without cLBP, even after adjusting for influencing factors,

524 such as socio-demographics, lifestyle and number of diseases. Furthermore, Tagliaferri et al.⁵⁷
525 were able to separate 4156 chronic back pain patients from the UK Biobank dataset into five
526 sub-groups based on their scores of social isolation and depressive symptoms. Interestingly,
527 increased social isolation was only a feature of three sub-groups, encompassing 26% of all back
528 pain patients (n = 1085), while the remaining subgroups showed either no changes (4.1%; n =
529 776) or a reduction in social isolation scores (12%; n = 2296). This prevalence in patients with
530 LBP may indicate that reduced social functioning was identified by some studies, while others
531 did not find similar changes as compared to asymptomatic controls²². However, as levels of
532 social function (here: social participation) were found to be correlated with self-perceived
533 physical health status⁵⁸, a direct impact of social functioning on personal functional
534 impairments remains feasible.

535
536 In addition to social function, the psychological health or well-being was shown to be an
537 important variable in cLBP patient delineation. Using longitudinal data from the SwePain
538 cohort, including 9361 participants with and without chronic pain, psychological well-being
539 scores at baseline were able to predict pain intensity after 2 years²⁸. Within this study, positive
540 well-being was predictive of lower pain severity in participants without and with chronic pain.
541 Similar conclusions were drawn from the comparison of back pain patients with different levels
542 of mental distress, in which patients with higher mental distress showed, among other things,
543 reduced psychological well-being, reduced social function, and higher severe pain than patients
544 with lower mental distress²⁰. Furthermore, patients suffering from chronic pain exhibit
545 significantly lower quality of life scores across all sub-domains, including psychological well-
546 being¹⁹. Alterations in quality of life are stronger associated with changes in social functioning
547 and psychological well-being (via pain catastrophizing) than pain intensity itself²⁷, indicating
548 the high importance of psycho-social aspects for daily living with painful conditions such as
549 back pain.

550

551 Our findings suggest that spinal herniation and degeneration observed on MRI may contribute
552 to pain mechanisms in the bio-psychosocial model of cLBP, consistent with previous meta-
553 analyses⁹ indicating a higher prevalence of disc herniation and degeneration on MRI in adults
554 with cLBP compared with asymptomatic individuals. However, it is clear that MRI findings
555 alone do not fully explain pain presence in cLBP, highlighting the need for caution when
556 interpreting MRI results at the individual patient level. Interestingly, cervical spine rotation but
557 not lumbar back motion assessments were shown to be robust important examinations for cLBP
558 delineation. This contradictory finding is likely a result of two factors. First, the poor cervical
559 rotation may be the result of neck pain that has high comorbidity with cLBP⁶² and can lead to
560 decreased axial rotation⁴⁴. Second, poor psychological health has been associated with neck
561 pain³², which we found to also present high importance in patient stratification and relates to
562 the bio-physical-psychosocial interplay present in cLBP patients⁵⁵. A systematic review on hip
563 mobility in LBP patients⁶ showed small to no changes in hip flexion compared to controls. As
564 all studies had less than 110 subjects, they were likely under powered to uncover the decreased
565 mobility we show here and may represent a diagnostic test for LBP. Previous studies have
566 shown that clinical kinematic data can effectively stratify cLBP patients into high, low and
567 intermediate risk groups¹, suggesting that pain correlates with reduced physical function.
568 Persistent nociceptive input from aggravated spinal joints/muscles may lead to reduced motor
569 output and spinal cord excitability³⁸, potentially resulting in a reduced ability to recruit specific
570 muscles and necessitating compensatory movement strategies. Our findings underscore that
571 detailed movement analysis could serve as a diagnostic biomarker for LBP, potentially rivalling
572 medical imaging in diagnostic accuracy and improving patient care by identifying sub-
573 populations likely to respond well to specific therapies or at risk of adverse outcomes.

574

575 Classification of cLBP patients has often been conducted on relatively small sample sizes (<
576 200) as well as utilising a single data domain⁵³. Performance of such models are subject to
577 overfitting due to their small samples and would likely perform poorly at out-of-sample
578 classification in unseen external datasets⁴³. Furthermore, several studies have established
579 classification models with high accuracy (> 0.8) at determining particular LBP symptoms
580 ^{2,3,24,33,45,49,61,70}, although these lack clinical applicability in understanding the most appropriate
581 variables and modalities in classification of cLBP as well as the classification of the disorder in
582 general. Classification models created using large datasets (n > 1000) either contained
583 psychosocial and demographic variables, without imaging and physical features^{40,52}. On the
584 other hand, Jin-Heekun and colleagues²³ employed only physical features without considering
585 important psychosocial factors, which have shown to be important in previous research⁵⁴⁻⁵⁷ as
586 well as in our current study (Fig. 3). Our best model (Boruta – Q + C + M) performed slightly
587 worse than Parsaeian et al.⁴⁰ (AUC 0.693 – 0.75) and compared to Shim et al.⁵² (AUC, 0.693
588 – 0.716), we utilised more plentiful data points per subject in a significantly smaller sample size
589 (approximately 34x and 6x smaller respectively) to address the clinically relevant question of
590 what modalities and variables a best suited for cLBP classification.

591
592 The multimodal nature of the data utilised and the amount of subjects that have participated in
593 physical, imaging, and questionnaire measurements are major strengths in our current study.
594 Utilising the “Berlin Back Study” dataset that contains more than 500 subjects in the different
595 domains highlighted as lacking in a recent review by our group⁵³, which enables us to robustly
596 investigate the importance of different modalities as well as specific variables in cLBP patient
597 stratification. The large sample size enabled us to minimise model overfitting and data leakage
598 through cross-validation and hold-out testing samples of good size and variable distributions.
599 A more accurate representation of a models performance is provided by out-of-sample testing
600 that uses a new sample population containing comparable variables. This provides a test set

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

614

615 Conclusion

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

622 **Disclosures**

623 **Funding:** This study is part of the Research Unit FOR 5177 funded by the German Research
624 Foundation (DFG), Hendrik Schmidt: SCHM 2572/11-1, SCHM 2572/12-1, SCHM 2572/13-
625 1; Sandra Reitmaeier: RE 4292/3-1, Matthias Pumberger: PU762/1-1. The analyses and
626 contribution from the Hochschule für Gesundheit were funded, in part, by grant number
627 50WK2273A (to DLB) from the German AeroSpace Center (DLR).

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

629

630 **Acknowledgments**

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

633

634 **References**

- 635
- 636 1. Abdollahi M, Ashouri S, Abedi M, Azadeh-Fard N, Parnianpour M, Khalaf K, Rashedi
637 E: Using a Motion Sensor to Categorize Nonspecific Low Back Pain Patients: A
638 Machine Learning Approach. *Sensors Multidisciplinary Digital Publishing Institute*;
639 20:3600, 2020.
- 640 2. Abdullah AA, Yaakob A, Ibrahim Z: Prediction of Spinal Abnormalities Using Machine
641 Learning Techniques. *IEEE*; page 1–62018.
- 642 3. Al Imran A, Rifat MRI, Mohammad R: Enhancing the classification performance of
643 lower back pain symptoms using genetic algorithm-based feature selection. *Springer*;
644 page 455–692020.
- 645 4. Anderson TW, Darling DA: Asymptotic Theory of Certain “Goodness of Fit” Criteria
646 Based on Stochastic Processes. *The Annals of Mathematical Statistics Institute of*
647 *Mathematical Statistics*; 23:193–212, 1952.
- 648 5. Andersson GB: Epidemiologic aspects on low-back pain in industry. *Spine (Phila Pa*
649 *1976)* 6:53–60, 1981.
- 650 6. Avman MA, Osmotherly PG, Snodgrass S, Rivett DA: Is there an association between
651 hip range of motion and nonspecific low back pain? A systematic review.
652 *Musculoskeletal Science and Practice* 42:38–51, 2019.
- 653 7. Barrett E, McCreesh K, Lewis J: Reliability and validity of non-radiographic methods of
654 thoracic kyphosis measurement: a systematic review. *Man Ther* 19:10–7, 2014.
- 655 8. Breiman L: Random Forests. *Machine Learning* 45:5–32, 2001.
- 656 9. Brinjikji W, Diehn FE, Jarvik JG, Carr CM, Kallmes DF, Murad MH, Luetmer PH: MRI
657 Findings of Disc Degeneration are More Prevalent in Adults with Low Back Pain than in
658 Asymptomatic Controls: A Systematic Review and Meta-Analysis. *AJNR Am J*
659 *Neuroradiol* 36:2394–9, 2015.
- 660 10. Collins GS, Reitsma JB, Altman DG, Moons KGM: Transparent Reporting of a
661 multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the
662 TRIPOD statement. *Ann Intern Med* 162:55–63, 2015.
- 663 11. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, Pratt M,
664 Ekelund U, Yngve A, Sallis JF, Oja P: International Physical Activity Questionnaire: 12-
665 Country Reliability and Validity. *Medicine & Science in Sports & Exercise* 35:1381,
666 2003.
- 667 12. Deyo RA, Gray DT, Kreuter W, Mirza S, Martin BI: United States trends in lumbar
668 fusion surgery for degenerative conditions. *Spine (Phila Pa 1976)* 30:1441–5; discussion
669 1446-1447, 2005.
- 670 13. Dreischarf B, Koch E, Dreischarf M, Schmidt H, Pumberger M, Becker L: Comparison
671 of three validated systems to analyse spinal shape and motion. *Sci Rep* 12:10222, 2022.

- 672 14. Fujiwara A, Tamai K, Yamato M, An HS, Yoshida H, Saotome K, Kurihashi A: The
673 relationship between facet joint osteoarthritis and disc degeneration of the lumbar spine:
674 an MRI study. *Eur Spine J* 8:396–401, 1999.
- 675 15. Gardner B, Abraham C, Lally P, de Bruijn G-J: Towards parsimony in habit
676 measurement: Testing the convergent and predictive validity of an automaticity subscale
677 of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and*
678 *Physical Activity* 9:102, 2012.
- 679 16. Ge L, Pereira MJ, Yap CW, Heng BH: Chronic low back pain and its impact on physical
680 function, mental health, and health-related quality of life: a cross-sectional study in
681 Singapore. *Sci Rep* 12:20040, 2022.
- 682 17. Grotle M, Småstuen MC, Fjeld O, Grøvle L, Helgeland J, Storheim K, Solberg TK,
683 Zwart J-A: Lumbar spine surgery across 15 years: trends, complications and
684 reoperations in a longitudinal observational study from Norway. *BMJ Open British*
685 *Medical Journal Publishing Group*; 9:e028743, 2019.
- 686 18. Guerhazi M, Ghroubi S, Kassis M, Jaziri O, Keskes H, Kessomtini W, Ben Hammouda
687 I, Elleuch M-H: [Validity and reliability of Spinal Mouse to assess lumbar flexion]. *Ann*
688 *Readapt Med Phys* 49:172–7, 2006.
- 689 19. Hadi MA, McHugh GA, Closs SJ: Impact of Chronic Pain on Patients' Quality of Life:
690 A Comparative Mixed-Methods Study. *J Patient Exp* 6:133–41, 2019.
- 691 20. Hnatešen D, Pavić R, Radoš I, Dimitrijević I, Budrovac D, Čebohin M, Gusar I: Quality
692 of Life and Mental Distress in Patients with Chronic Low Back Pain: A Cross-Sectional
693 Study. *Int J Environ Res Public Health* 19:10657, 2022.
- 694 21. Holm S: A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian*
695 *Journal of Statistics* [Board of the Foundation of the Scandinavian Journal of Statistics,
696 Wiley]; 6:65–70, 1979.
- 697 22. Iguti AM, Guimarães M, Barros MBA: Health-related quality of life (SF-36) in back
698 pain: a population-based study, Campinas, São Paulo State, Brazil. *Cad Saude Publica*
699 37:e00206019, 2021.
- 700 23. Jin-Heeku: Analysis of sitting posture using wearable sensor data and support vector
701 machine model. *Medico-Legal Update* 1:334–8, 2018.
- 702 24. Karabulut EM, Ibrikci T: Effective automated prediction of vertebral column pathologies
703 based on logistic model tree with SMOTE preprocessing. *Journal of Medical Systems*
704 38:50, 2014.
- 705 25. Kraemer J: Natural course and prognosis of intervertebral disc diseases. *International*
706 *Society for the Study of the Lumbar Spine Seattle, Washington, June 1994. Spine (Phila*
707 *Pa 1976)* 20:635–9, 1995.
- 708 26. Kurasa MB, Rudnicki WR: Feature Selection with the Boruta Package. *Journal of*
709 *Statistical Software* 36:1–13, 2010.

- 710 27. Lamé IE, Peters ML, Vlaeyen JWS, Kleef M v, Patijn J: Quality of life in chronic pain is
711 more associated with beliefs about pain, than with pain intensity. *Eur J Pain* 9:15–24,
712 2005.
- 713 28. Larsson B, Dragioti E, Gerdle B, Björk J: Positive psychological well-being predicts
714 lower severe pain in the general population: a 2-year follow-up study of the SwePain
715 cohort. *Ann Gen Psychiatry* 18:8, 2019.
- 716 29. Lee W, Alexeyenko A, Pernemalm M, Guegan J, Dessen P, Lazar V, Lehtiö J, Pawitan
717 Y: Identifying and Assessing Interesting Subgroups in a Heterogeneous Population.
718 *Biomed Res Int* 2015:462549, 2015.
- 719 30. Lonsdale C, Hodge K, Rose EA: The behavioral regulation in sport questionnaire
720 (BRSQ): Instrument development and initial validity evidence. *Journal of Sport &
721 Exercise Psychology US: Human Kinetics*; 30:323–55, 2008.
- 722 31. Lötsch J, Utsch A: Machine learning in pain research. *Pain* 159:623–30, 2018.
- 723 32. Mansfield M, Thacker M, Taylor JL, Bannister K, Spahr N, Jong ST, Smith T: The
724 association between psychosocial factors and mental health symptoms in cervical spine
725 pain with or without radiculopathy on health outcomes: a systematic review. *BMC
726 Musculoskeletal Disorders* 24:235, 2023.
- 727 33. Mathew B, Norris D, Hendry D, Waddell G: Artificial intelligence in the diagnosis of
728 low-back pain and sciatica. *Spine* 13:168–72, 1988.
- 729 34. Meucci RD, Fassa AG, Faria NMX: Prevalence of chronic low back pain: systematic
730 review. *Rev Saude Publica* 49:1, 2015.
- 731 35. Meyerding HW: Spondyloptosis. *Surgery, Gynecology & Obstetrics* :371–7, 1932.
- 732 36. Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR: Degenerative disk disease:
733 assessment of changes in vertebral body marrow with MR imaging. *Radiology* 166:193–
734 9, 1988.
- 735 37. Mwangi B, Tian TS, Soares JC: A review of feature reduction techniques in
736 neuroimaging. *Neuroinformatics* 12:229–44, 2014.
- 737 38. Nijs J, Daenen L, Cras P, Struyf F, Roussel N, Oostendorp RAB: Nociception affects
738 motor output: a review on sensory-motor interaction with focus on clinical implications.
739 *Clin J Pain* 28:175–81, 2012.
- 740 39. Noroozi Z, Orooji A, Erfannia L: Analyzing the impact of feature selection methods on
741 machine learning algorithms for heart disease prediction. *Sci Rep Nature Publishing
742 Group*; 13:22588, 2023.
- 743 40. Parsaeian M, Mohammad K, Mahmoudi M, Zeraati H: Comparison of logistic regression
744 and artificial neural network in low back pain prediction: second national health survey.
745 *Iranian Journal of Public Health* 41:86, 2012.
- 746 41. Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, Boccia S:
747 Benefits and challenges of Big Data in healthcare: an overview of the European
748 initiatives. *Eur J Public Health* 29:23–7, 2019.

- 749 42. Pfirmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N: Magnetic resonance
750 classification of lumbar intervertebral disc degeneration. *Spine* 26:1873–8, 2001.
- 751 43. Rajput D, Wang W-J, Chen C-C: Evaluation of a decided sample size in machine
752 learning applications. *BMC Bioinformatics* 24:48, 2023.
- 753 44. Rampazo ÉP, da Silva VR, de Andrade ALM, Back CGN, Madeleine P, Arendt-Nielsen
754 L, Liebano RE: Sensory, Motor, and Psychosocial Characteristics of Individuals With
755 Chronic Neck Pain: A Case Control Study. *Physical Therapy* 101:pzab104, 2021.
- 756 45. Riveros NAM, Espitia BAC, Pico LEA: Comparison between K-means and self-
757 organizing maps algorithms used for diagnosis spinal column patients. *Informatics in*
758 *Medicine Unlocked* 16:100206, 2019.
- 759 46. Roland M, Fairbank J: The Roland–Morris Disability Questionnaire and the Oswestry
760 Disability Questionnaire. *Spine* 25:3115, 2000.
- 761 47. Roldán-Jiménez C, Pérez-Cruzado D, Neblett R, Gatchel R, Cuesta-Vargas A: Central
762 Sensitization in Chronic Musculoskeletal Pain Disorders in Different Populations: A
763 Cross-Sectional Study. *Pain Med* 21:2958–63, 2020.
- 764 48. Rusu AC, Kreddig N, Hallner D, Hülsebusch J, Hasenbring MI: Fear of
765 movement/(Re)injury in low back pain: confirmatory validation of a German version of
766 the Tampa Scale for Kinesiophobia. *BMC Musculoskelet Disord* 15:280, 2014.
- 767 49. Sandag GA, Tedry NE, Lolong S: Classification of lower back pain using K-Nearest
768 Neighbor algorithm. *IEEE*; page 1–52018.
- 769 50. Schizas C, Theumann N, Burn A, Tansey R, Wardlaw D, Smith FW, Kulik G:
770 Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the
771 dural sac on magnetic resonance images. *Spine (Phila Pa 1976)* 35:1919–24, 2010.
- 772 51. Shilo S, Rossman H, Segal E: Axes of a revolution: challenges and promises of big data
773 in healthcare. *Nat Med Nature Publishing Group*; 26:29–38, 2020.
- 774 52. Shim J-G, Ryu K-H, Cho E-A, Ahn JH, Kim HK, Lee Y-J, Lee SH: Machine Learning
775 Approaches to Predict Chronic Lower Back Pain in People Aged over 50 Years.
776 *Medicina Multidisciplinary Digital Publishing Institute*; 57:1230, 2021.
- 777 53. Tagliaferri SD, Angelova M, Zhao X, Owen PJ, Miller CT, Wilkin T, Belavy DL:
778 Artificial intelligence to improve back pain outcomes and lessons learnt from clinical
779 classification approaches: three systematic reviews. *npj Digital Medicine* 3:93, 2020.
- 780 54. Tagliaferri SD, Fitzgibbon BM, Owen PJ, Miller CT, Bowe SJ, Belavy DL: Brain
781 structure, psychosocial, and physical health in acute and chronic back pain: a
782 UKBioBank study. *Pain* 163:1277–90, 2022.
- 783 55. Tagliaferri SD, Ng S-K, Fitzgibbon BM, Owen PJ, Miller CT, Bowe SJ, Belavy DL:
784 Relative contributions of the nervous system, spinal tissue and psychosocial health to
785 non-specific low back pain: Multivariate meta-analysis. *Eur J Pain* , 2021.
- 786 56. Tagliaferri SD, Owen PJ, Miller CT, Angelova M, Fitzgibbon BM, Wilkin T, Masse-
787 Alarie H, Van Oosterwijck J, Trudel G, Connell D, Taylor A, Belavy DL: Towards data-

- 788 driven biopsychosocial classification of non-specific chronic low back pain: a pilot
789 study. *Sci Rep Nature Publishing Group*; 13:13112, 2023.
- 790 57. Tagliaferri SD, Wilkin T, Angelova M, Fitzgibbon BM, Owen PJ, Miller CT, Belavy
791 DL: Chronic back pain sub-grouped via psychosocial, brain and physical factors using
792 machine learning. *Sci Rep* 12:15194, 2022.
- 793 58. Takeyachi Y, Konno S, Otani K, Yamauchi K, Takahashi I, Suzukamo Y, Kikuchi S:
794 Correlation of low back pain with functional status, general health perception, social
795 participation, subjective happiness, and patient satisfaction. *Spine (Phila Pa 1976)*
796 28:1461–6; discussion 1467, 2003.
- 797 59. Tanaka K, Nishigami T, Mibu A, Imai R, Manfuku M, Tanabe A: Combination of Pain
798 Location and Pain Duration is Associated with Central Sensitization-Related Symptoms
799 in Patients with Musculoskeletal Disorders: A Cross-Sectional Study. *Pain Pract* 21:646–
800 52, 2021.
- 801 60. Topalidou A, Tzagarakis G, Souvatzis X, Kontakis G, Katonis P: Evaluation of the
802 reliability of a new non-invasive method for assessing the functionality and mobility of
803 the spine. *Acta Bioeng Biomech* 16:117–24, 2014.
- 804 61. Vaughn ML, Cavill SJ, Taylor SJ, Foy MA, Fogg AJ: Direct explanations for the
805 development and use of a multi-layer perceptron network that classifies low-back-pain
806 patients. *International Journal of Neural Systems* 11:335–47, 2001.
- 807 62. von der Lippe E, Krause L, Porst M, Wengler A, Leddin J, Müller A, Zeisler M-L,
808 Anton A, Rommel A: Prevalence of back and neck pain in Germany. Results from the
809 BURDEN 2020 Burden of Disease Study. *J Health Monit* 6:2–14, 2021.
- 810 63. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandembroucke JP: The
811 Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)
812 Statement: Guidelines for reporting observational studies. *Preventive Medicine* 45:247–
813 51, 2007.
- 814 64. Von Korff M, Ormel J, Keefe FJ, Dworkin SF: Grading the severity of chronic pain.
815 *PAIN* 50:133, 1992.
- 816 65. Waddell G, Newton M, Henderson I, Somerville D, Main CJ: A Fear-Avoidance Beliefs
817 Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain
818 and disability. *PAIN* 52:157, 1993.
- 819 66. Ware JE, Sherbourne CD: The MOS 36-item short-form health survey (SF-36). I.
820 Conceptual framework and item selection. *Med Care* 30:473–83, 1992.
- 821 67. World Medical Association: World Medical Association Declaration of Helsinki: ethical
822 principles for medical research involving human subjects. *JAMA* 310:2191–4, 2013.
- 823 68. Wright MN, Ziegler A: ranger: A Fast Implementation of Random Forests for High
824 Dimensional Data in C++ and R. *Journal of Statistical Software* 77:1–17, 2017.
- 825 69. Yoshihara H, Yoneoka D: National trends in the surgical treatment for lumbar
826 degenerative disc disease: United States, 2000 to 2009. *Spine J* 15:265–71, 2015.

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

830