International Journal of Medical Informatics 211 (2026) 106335

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

BackTracker: Machine learning to identify kinematic phenotypes for personalised exercise management in non-specific low back pain

Zebang Liu[a,*], Yulia Hicks[a], Liba Sheeran[b]

[a] School of Engineering, Cardiff University, Queen's Buildings, Cardiff, CF24 3AA, United Kingdom
[b] School of Health Sciences, University of Southampton, Highfield Campus, Southampton, SO17 1BJ, United Kingdom

HIGHLIGHTS

- First video-based AI framework for classifying extension vs flexion NSLBP subgroups.
- The Track Anything segmentation model and the HigherHRNet pose estimation model are jointly used to capture spinal curvature features during movement.
- Achieves 91.9% accuracy (AUC 0.94) in distinguishing extension vs flexion NSLBP groups.
- Outperforms patient-reported outcomes and motion-only features in classification.
- Enables remote, personalised NSLBP rehabilitation via automated video analysis.

ARTICLE INFO

Keywords:
Non-specific low back pain
Kinematic phenotypes
Machine learning
Video analysis
Personalised exercise

ABSTRACT

Background: Low back pain (LBP) is a leading cause of global disability. Most cases are non-specific (NSLBP) and lack identifiable causes. Early active management is endorsed by clinical guidelines; however, exercises are rarely customised despite substantial variability in impairments. Existing classification systems can support targeted rehabilitation but require extensive clinical training and lengthy assessment procedures, limiting timely personalised care.

Objective: This study used AI methods to identify the two most common motor control impairments (MCIs)—flexion and extension patterns (FP/EP)—in NSLBP. The approach used spinal silhouettes extracted from movement videos to enable self-phenotyping and guide personalised exercise selection.

Methods: Data were collected from a research fellowship involving ninety NSLBP participants classified by an expert physiotherapist (LS) into FP or EP MCIs. Participants performed standard forward- and backward-bend tasks recorded in the sagittal plane. Pose estimation and instance segmentation techniques were applied to extract motion features and spine silhouettes. From each participant, a curated set of 80 black-and-white back images captured at specific bending angles was produced. These features were used to train a feedforward neural network. Model performance was assessed using five-fold cross-validation with accuracy, sensitivity, specificity, F1 score and AUC.

Results: The model achieved a diagnostic accuracy of 91.91% (95% CI 84.8–99.1) for backward-bend videos, exceeding reported inter-examiner agreement rates for trained physiotherapists. Robustness was supported by a mean AUC of 0.9422. Accuracy was lower for forward-bend images (86.69%), combined tasks (86.29%), or PROMs alone (63.82%). Adding PROMs to forward- or backward-bend tasks provided only modest improvements (66.32% and 71.62%, respectively).

Conclusion: The model reliably distinguished between FP and EP NSLBP subgroups, demonstrating the potential of AI to support timely personalised rehabilitation. The integration of PROMs with motion features reduced classification accuracy, suggesting that self-reported outcomes may provide limited benefit when tailoring exercises to specific physical impairments.

* Corresponding author.
Email addresses: LiuZ91@cardiff.ac.uk (Z. Liu), hicksYA@cardiff.ac.uk (Y. Hicks), L.Sheeran@soton.ac.uk (L. Sheeran).

https://doi.org/10.1016/j.ijmedinf.2026.106335
Received 19 September 2025; Received in revised form 24 November 2025; Accepted 2 February 2026
Available online 7 February 2026

1. Introduction

Low back pain (LBP) is a globally prevalent musculoskeletal disorder affecting individuals' health and quality of life and productivity [1]. It is a leading global cause of disability and poses a significant burden that is expected to rise [2]. Up to 90% of LBP is classified as non-specific LBP (NSLBP), characterised by the absence of a definitive pathology [3] making it difficult to tailor interventions as recommended by clinical guidelines [4–6].

NSLBP can be categorised using several clinical classification systems [7]. One clinically used approach is the Multidimensional Classification System (MDCS), a structured, mechanism-based assessment grounded in the biopsychosocial model [8]. It begins with a systematic examination comprising of taking subjective history, including symptom description and pain behaviour and objective physical examination, focusing on movement and posture. Patients are classified into subgroups based on dominant physical and behavioural drivers. Two primary groups are defined: (i) movement impairment (MI), characterised by restricted movement due to protective pain avoidance behaviours and (ii) motor control impairment (MCI), marked by less restricted but destabilising pain provoking movement patterns [9]. These primary classifications are then further divided into more specific directional patterns. MI subdivides into flexion pattern (FP-MI) and extension pattern (EP-MI), while MCI comprises five directional patterns including flexion (FP), active extension (AEP), passive extension (PEP), lateral shift, and multi-directional pattern, each with tailored intervention strategies [9–13].

Individuals within MDCS subgroups exhibit distinct postural and movement patterns, including differences in spinal alignment, muscle activation, and task-specific kinematics, supporting the validity of these phenotypes [14–20]. Physical therapists can classify NSLBP subgroups using the MDCS with high reliability and inter-examiner agreement [21,22]. However, achieving such accuracy requires clinician's level of training (~100 h) and continued clinical experience [22], limiting broader implementation and scalability.

Artificial intelligence (AI), particularly computer vision methods such as human pose estimation (HPE) and human instance segmentation (HIS), allows extraction of detailed kinematic and morphological features from video without markers [23,24]. Previous work, such as HigherHRNet-based classification, demonstrated feasibility but was limited by sparse keypoint coverage and insufficient modelling of spinal morphology [25]. The use case for HigherHRNet is shown in Fig. 1. To address these limitations, this study applies the Track Anything Model (TAM), a state-of-the-art HIS approach capable of automatically segmenting human silhouettes from video (Fig. 2) [24,26]. Research integrating PROMs with video-derived features for FP/EP classification remains limited despite their potential clinical relevance.

This study introduces a machine learning (ML) framework combining SOTA HPE–HIS methods to classify the two most common MCI subgroups—flexion (FP) and extension (EP) patterns, from standard movement videos. By leveraging kinematic features and incorporating patient-reported outcome measures (PROMs) on pain, disability, and pain-related psychological factors [27], the model provides an objective, scalable, and cost-effective approach to support accurate subgroup classification. The aims are to: (i) Explore the feasibility of using TAM and HPE for extracting silhouettes from NSLBP movement videos; (ii) Develop and evaluate an automated classifier for distinguishing between FP-/EP-MCI subgroups; and (iii) Assess the relative importance of different feature types and movement tasks in classifying subgroups.

2. Dataset

This study used a previously described BackMotion Clinical Dataset (BMCD) of 78 individuals with NSLBP [25,28]. The dataset includes spine and pelvis data captured using 3D kinematics (Vicon, Oxford Metrics, UK), inertial sensors (Xsens, Netherlands), and video recordings (GoPro HERO, GoPro Inc., USA) during forward- and backward-bend tasks, which have the highest expert consensus for distinguishing EP- and FP-MCI subgroups [29]. Full details on data collection, patient reported outcome measures (PROMs), and preprocessing are reported elsewhere [25,28].

3. Methods

3.1. Overview of the BackTracker framework

'BackTracker' is a machine learning pipeline integrating HPE and HIS methods to extract spinal motion and curvature from video and classify individuals into FP- or EP-MCI subgroups. The pipeline consists of four stages (Fig. 3): (I) Keypoint detection where the HigherHRNet HPE model [23] detects three anatomical landmarks (neck, hip and ankle) from the video frames to compute the spinal angle (θ) (Fig. 1), a surrogate for trunk posture; (II) Silhouette segmentation using TAM [24], guided by detected keypoints, full body silhouettes are extracted frame by frame (Figs. 2 and 3); (III) Feature extraction of (a) motion features from θ, where temporal-spatial movement parameters (mean, range, variance) are derived [30] and (b) silhouettes are cropped to obtain 'back masks' representing trunk curvature throughout movement (Fig. 3); and (IV) Classification using a ResNet18-based model, trained on these features, classifies patients into FP-MCI or EP-MCI. The workflow was reported in accordance with the CHAMAI checklist [31].

3.2. Spinal motion features

Spinal motion features replicate clinical assessment by capturing the dynamic range of motion (θ) during forward and backward bending, as previously validated in NSLBP populations [32]. Using HPE, the spinal

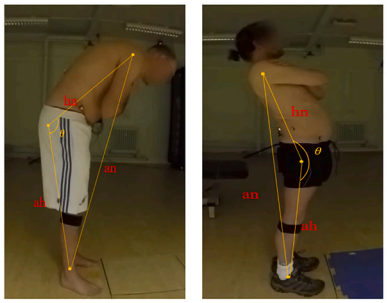

Fig. 1. Example of human pose estimation using HigherHRNet [23], showing detected anatomical keypoints (ankle, hip, neck) during (a) forward bending and (b) backward bending.

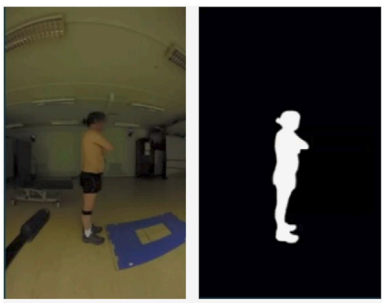

Fig. 2. Application of the track anything model (TAM) for human instance segmentation. (a) Original video frame. (b) Extracted human silhouette (mask) generated by TAM.

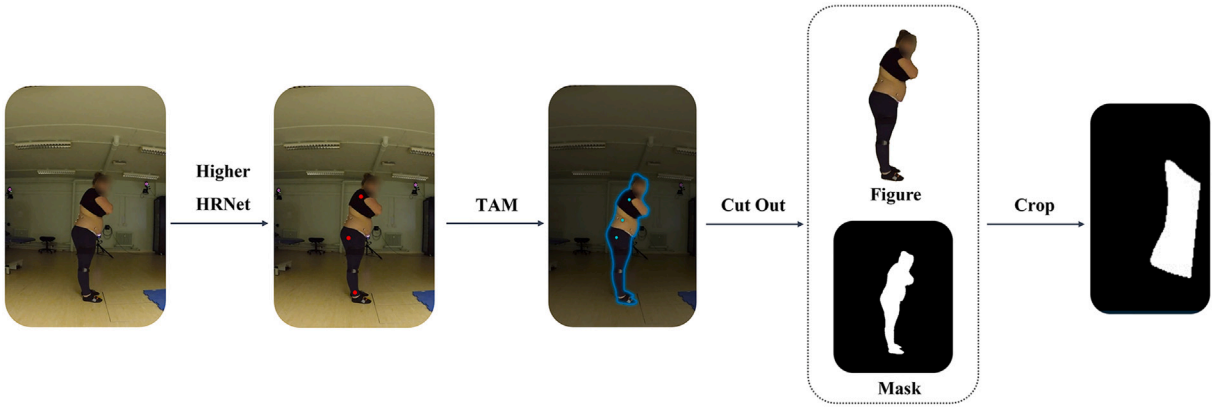

Fig. 3. Sequential workflow for spine postural analysis using the track anything model [24].

Table 1
Forward and backward bend mask counts in EP-MCI and FP-MCI subgroups.

		Forward bend masks				Backward bend masks			
Group	**N Participants**	**Total**	**Mean**	**SD**	**Range**	**Total**	**Mean**	**SD**	**Range**
EP-MCI	38	75993	1000.4	392.7	(37, 1728)	56650	766.0	277.2	(36, 1636)
FP-MCI	40	86301	1079.2	648.8	(39, 3806)	59410	762.2	263.5	(38, 1249)

angle θ is computed from the 2D coordinates of the neck, hip, and ankle joints (Fig. 3) based on the Euclidean distances between these points:

$$\theta = \arccos\left(\frac{(\mathbf{p}_{\text{hip}} - \mathbf{p}_{\text{neck}}) \cdot (\mathbf{p}_{\text{ankle}} - \mathbf{p}_{\text{hip}})}{\|\mathbf{p}_{\text{hip}} - \mathbf{p}_{\text{neck}}\| \|\mathbf{p}_{\text{ankle}} - \mathbf{p}_{\text{hip}}\|}\right) \quad (1)$$

From the angle series, we computed statistical descriptors (mean, range, min–max, variance, standard deviation) and temporal metrics (stability time, depth variance, repetition time variance), which have been shown to distinguish NSLBP subgroups [32].

3.3. Spinal curvature features

Spinal curvature features are extracted from segmented silhouettes following the pipeline in Fig. 3. HigherHRNet first identifies the three anatomical landmarks (neck, hip, and ankle) to guide TAM-based segmentation across video frames, generating background-free silhouettes ('Figure' and 'Mask' in Fig. 3). Each silhouette is then cropped along the shoulder-hip axis to produce a 'back mask' ('Back Mask' in Fig. 3).

Curvature features are derived from forward bend (~80 ° range) and backward bend (~30 ° range) videos, divided into 5-degree intervals, and one mask was extracted per interval. On average, 1040 masks were generated per participant during forward bend and 764 during backward bend, reflecting individual variations in range (Table 1).

3.4. Patient-reported outcome measures features

Patient-reported outcome measures (PROMs), including pain, disability, and psychological questionnaires, provide subjective insights into the patient's experience of LBP. As noted by Hartley et al. [25], such measures complement objective kinematic analysis by capturing psychosocial dimensions relevant to classification.

3.5. Classification framework design

The classification framework was built using a ResNet18-Binary network (Fig. 4), fine-tuned on the BMCD training dataset. Transformer-based alternatives [33], including Vision Transformer (ViT) [34], were excluded due to higher data and computational demands. Previous analyses show that CNNs remain more efficient on datasets below the million-sample scale, whereas ViT typically requires tens of millions (10^7–10^8) of images or large-scale pretraining to surpass CNN performance [35]. Although the cohort was small (N = 78), each participant contributed 80 backward-bend frames (6240 images total), which is generally sufficient for training CNNs on domain-specific tasks [35]. We also conducted preliminary experiments using a ViT-based architecture and observed rapid overfitting, consistent with their sensitivity to limited data [35].

To ensure consistency and capture movement variability, a fixed number of randomly selected back masks ('M' in Fig. 4) were passed through the classifier. Key modifications included a single-neuron output ('Output' in Fig. 4) to produce continual prediction, and a shared-weight scheme to apply the same model across all masks for a participant [36].

Predictions across the selected frames were averaged and passed through a sigmoid function, with a threshold of 0.5 used to determine class: >0.5 = FP-MCI; ≤0.5 = EP-MCI. This design prioritises interpretability and efficiency.

3.6. Evaluation strategy

The BackTracker framework was evaluated for distinguishing EP-MCI and FP-MCI subgroups in NSLBP patients using three distinct feature streams: motion-related features (from HPE), spinal curvature features (from 'back masks'), and patient-reported outcome measures (PROMs) (Fig. 5). Each stream was tested independently and then in combinations to assess the contribution of multimodal input.

In the motion feature stream, HPE-derived motion features (M = 1 or 9) were processed via a dedicated fully connected network (FCN) with a LeakyReLU activation function. In the spinal curvature stream, a fixed number of K back masks per patient were processed using a shared-weight ResNet18-Binary model. The K outputs were averaged to produce a single patient-level representation. In the PROMs stream, PROM features (N = 1 or 8) were normalised and passed through an FCN with a LeakyReLU.

Each stream captured a different dimension of the patient's representation: motion (functional), curvature (structural), and PROMs (subjective). In multimodal models, outputs from each stream were

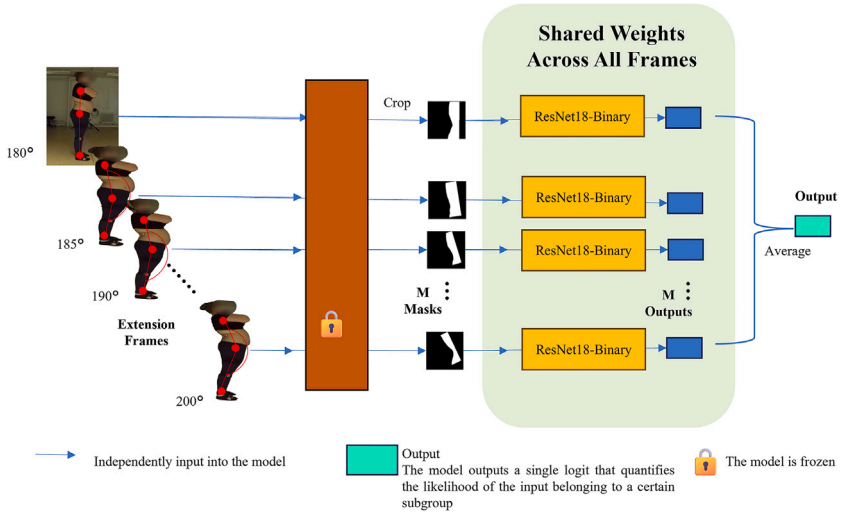

Fig. 4. Overview of the BackTracker classification framework, including pose estimation, silhouette extraction, image preprocessing, and ResNet18-based prediction.

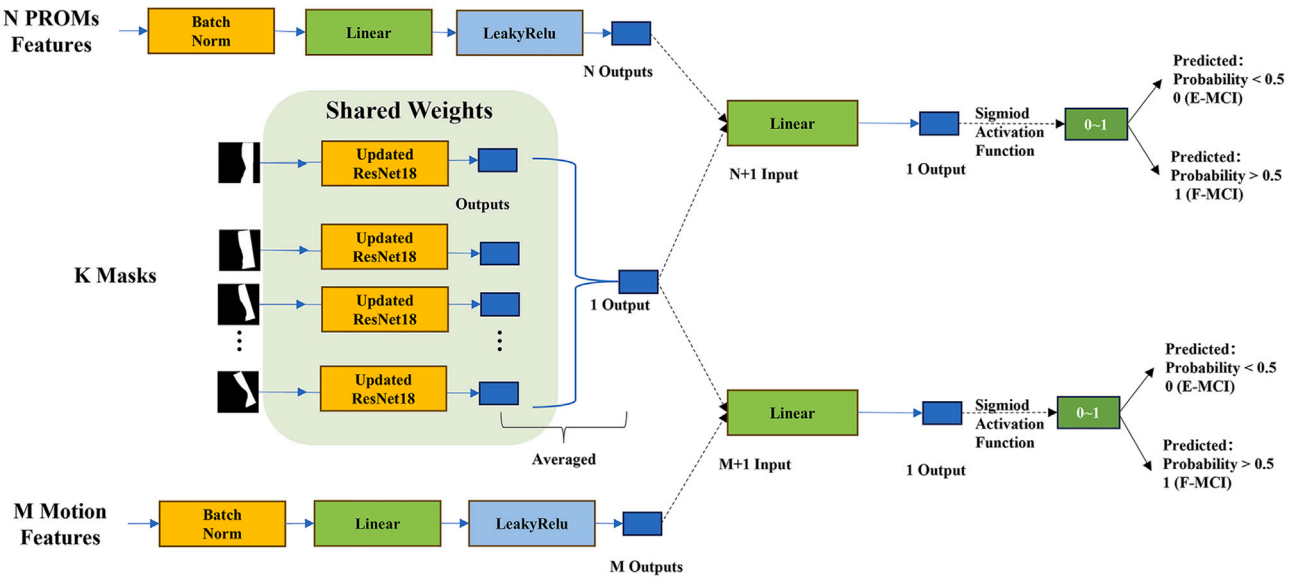

Fig. 5. Evaluation strategy for multimodal classification. K, N, and M denote the number of selected image-based features, motion features, and PROM-based features, respectively. $N \in \{1, 8\}$ and $M \in \{1, 9\}$, K will be decided through further investigation.

merged via an additional FCN layer, with final predictions generated using a sigmoid function and a 0.5 threshold.

We report macro-averaged metrics with 95% CIs across five folds (Table 2). In addition, we provide a micro-averaged confusion matrix aggregated over all 78 participants (Table 5).

3.7. Hyperparameters and evaluation metrics

The BackTracker model was trained using the Adam optimizer (learning rate of 0.0001), a mini-batch size of 32, and 200 epochs. These hyperparameters were selected empirically based on pilot experiments.

Generalisability was evaluated using five-fold cross-validation with the dataset randomly split into five groups, with each group used once for testing and the remaining four for training. To prevent data leakage, all frames belonging to the same participant were kept within a single fold. Metrics averaged across folds included accuracy, sensitivity, specificity, and the F1 score. Classifier performance across thresholds was evaluated using Receiver Operating Characteristic (ROC) curve analysis [37], with Area Under the Curve (AUC) used to quantify overall performance. An AUC of 1 indicates perfect classification, while 0.5 reflects chance-level performance [37].

3.8. Ethics statement

This study adhered to Cardiff University ethical standards and was approved by the School of Engineering Research Ethics Committee (Reference: 2023-PGR-LZ-R0) on 6 March 2024. All participants provided written informed consent.

4. Results

Classification performance was evaluated using unimodal and multimodal models across three feature types—spinal curvature image model (back masks), motion features and PROMs. Tables 2–4 and Fig. 6 summarise the performance outcomes.

Table 2
Performance of 5-fold cross-validation for the back mask image classifier model using forward and backward bend movements to classify EP- and FP-MCI.

Movement	Number of images	Accuracy	Sensitivity (EP)	Specificity (FP)	F1 Score
Forward bend	80	85.44%	93.82%	75.06%	0.8033
	160	84.27%	83.33%	79.34%	0.8024
	320	86.69%	93.33%	78.34%	0.8328
Backward bend	40	85.34%	86.67%	79.06%	0.8614
	80	**91.91%**	**90.00%**	**92.42%**	**0.9291**
	120	86.38%	88.48%	85.40%	0.8641
Forward + Backward	320 + 80	82.29%	91.79%	72.42%	0.7941
	160 + 80	86.29%	80.17%	80.00%	0.7958
	80 + 80	83.62%	73.00%	90.95%	0.8343

Key: 'Forward bend', 'Backward bend', and 'PROMs' represent all sub-features of this type of feature that are evaluated on the feedforward neural network model. The feature combinations with a "+" sign represent two types of features combined for evaluation.

Table 3
Performance of model based only on motion/PROMs features for EP-/FP-MCI classification.

Features	Accuracy	Sensitivity (EP)	Specificity (FP)	F1 Score
Forward bend	65.01%	84.00%	39.24%	0.4875
Backward bend	**74.29%**	71.43%	77.03%	0.7386
PROMs	63.82%	89.33%	32.02%	0.4257
Forward bend + PROMs	66.32%	66.67%	57.21%	0.5436
Backward bend + PROMs	71.62%	62.09%	75.57%	0.7054

Table 4
Performance of EP- and FP-MCI classification using an integrated model predicted on back mask image combining motion features or PROMs features (patient reported outcome measures) respectively.

Features	Accuracy	Sensitivity (EP)	Specificity (FP)	F1 Score
80 Backward Bend Images	**91.91%**	90.00%	92.42%	0.9291
80 Images + Min	82.89%	78.09%	85.40%	0.8432
80 Images + SD	**83.81%**	81.81%	84.25%	0.8435
80 Images + Var	82.48%	80.28%	84.25%	0.8275
80 Images + Max	81.05%	81.81%	78.54%	0.8058
80 Images + Stability Time	80.95%	79.43%	81.40%	0.8135
80 Images + Repetition Time	80.95%	73.71%	87.11%	0.8262
80 Images + Repetition Time Var	82.29%	79.43%	84.25%	0.8324
80 Images + Depth Var	80.95%	72.38%	83.11%	0.8094
80 Images + Range	79.71%	77.81%	85.51%	0.7996
80 Images + All Motion	81.14%	81.81%	76.84%	0.7986
80 Images + VAS	83.91%	81.81%	84.61%	0.8454
80 Images + Oswestry	81.33%	81.81%	80.45%	0.8188
80 Images + PCS	79.62%	79.43%	79.18%	0.7970
80 Images + PSEQ	79.62%	76.95%	71.40%	0.8004
80 Images + SBT	**85.24%**	87.14%	80.61%	0.8436
80 Images + TSK	80.95%	80.95%	76.12%	0.8042
80 Images + CSQ	79.62%	80.28%	82.28%	0.7894
80 Images + PASS-20	82.38%	79.79%	84.25%	0.8306
80 Images + All PROMs	74.38%	69.24%	76.48%	0.7397
All Forward Features	79.48%	88.50%	70.24%	0.7307
All Backward Features	**82.57%**	86.57%	81.97%	0.8221

Key: '80 Backward Bend Images' achieved the best performance on image-based model.

4.1. Classification using spinal curvature (Back masks)

Table 2 reports the performance of classifiers trained on spinal curvature features extracted from back mask images during forward bend, backward bend, or both movements. The highest accuracy (91.91%, 95% CI 84.8–99.1) was achieved using 80 back masks from the backward bend task. This configuration yielded a sensitivity (90.00%, 95% CI 71.49–100.00), specificity (92.42%, 95% CI 83.72–100.00), and an F1 score of 0.9291 (95% CI 0.8751–0.983). ROC analysis confirmed the robustness of this model, with an average AUC of 0.9422 across five folds (Fig. 6).

In comparison, models using forward bend–only back masks showed lower classification performance across all test image counts. Combining forward and backward bend masks further reduced accuracy and F1 scores, suggesting that backward bend provides more discriminative curvature patterns for differentiating EP- and FP-MCI subgroups (Table 2).

Table 5
Confusion matrix of the BackTracker model.

True / Predicted	FP	EP	Total
FP	36	3	39
EP	3	36	39
Total	39	39	78

4.2. Classification using motion features and PROMs

Table 3 presents results from classifiers that use numerical features. Motion features from backward bend achieved the best results, with 74.29% accuracy, 71.43% sensitivity, 77.03% specificity, and an F1

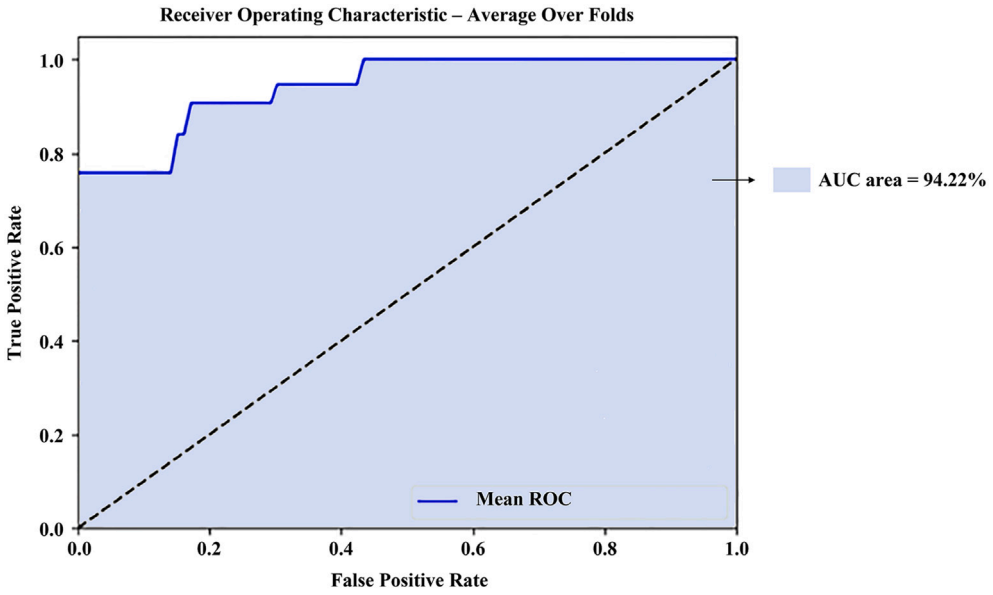

Fig. 6. Mean receiver operating characteristic (ROC) curve from 5-fold cross-validation for backward-bend silhouette images. The shaded region indicates the standard deviation across folds, and the enclosed area corresponds to the AUC value.

score of 0.7386. Motion features from forward bend yielded lower accuracy (65.01%) despite high sensitivity (84.00%) but low specificity (39.24%) and an F1 score of 0.4875. Models trained solely on PROMs showed the weakest performance with 63.82% accuracy and 32.02% specificity, although sensitivity remained high at 89.33% (Table 3). Combining PROMs and motion features led to marginal improvements in some cases, but performance remained lower than that of the spinal curvature classifier (Tables 2 and 3). These findings highlight the superior performance of backward bend features, while reaffirming the spinal curvature model as the most effective for classifying EP/FP subgroups.

4.3. Multimodal classification using images, motion, and PROMs

Table 4 details classification performance combining curvature images with motion or PROMs. While 80 backward bend back masks alone yielded the highest classification accuracy (91.91%, F1 = 0.9291), integrating additional features generally reduced performance. Combining back masks with individual motion features such as minimum or standard deviation reduced accuracy to approximately 83%, with corresponding F1 scores between 0.84 and 0.85. Incorporating all motion features simultaneously further reduced performance to 81.14% accuracy and an F1 score of 0.7986.

Back masks combined with individual PROMs performed best with the STarT Back Tool (SBT) (accuracy = 85.24%, F1 = 0.8436), whereas combining all PROMs decreased accuracy to 74.38% (F1 = 0.7397) (Table 4). Combining all PROMs and all motion features with back masks did not improve performance over the image-only model and typically resulted in degradation.

4.4. Error analysis

The overall confusion matrix (Table 4) shows that BackTracker correctly classified 72 out of 78 participants (92.31%). Of the six misclassifications, three FP cases were predicted as EP and three EP cases were predicted as FP, indicating a symmetric error pattern and comparable uncertainty between subgroups.

5. Discussion

This study evaluated the BackTracker classifier for distinguishing FP- and EP-pattern MCI subgroups using spinal curvature image features, motion features, and PROMs. Backward bend spinal curvature images alone yielded the highest accuracy (91.91%, F1 = 0.9291), outperforming all other modalities. Adding motion features or PROMs generally reduced performance. Among combined models, the best result was achieved by integrating backward bend images with STarTBACK Tool (SBT) PROM (85.24% accuracy). Motion-based features performed moderately (backward bend 74.29%, forward bend 65.01%) while combining all features did not improve accuracy. Overall, backward bend spinal curvature provides the strongest standalone discriminative signal, while additional numerical features offered limited benefit and sometimes diluted image-based performance.

The image-based model using sagittal spinal silhouette images from backward bend achieved accuracy comparable to physiotherapists' inter-examiner agreement (Kappa = 0.82, 86% agreement [27]), and exceeded the performance typically reported for clinician classification based on videos (Kappa = 0.55–0.71) [26]. Backward bend curvature and motion features consistently outperformed forward bend-based data, aligning with previous findings that backward bend kinematics more clearly differentiate FP- and EP-MCI subgroups [10].

The BackTracker showed higher classification accuracy for FP-MCI than EP-MCI, with specificity of 92.42% and sensitivity of 90.00%. This may reflect slight class imbalance (FP-MCI: n = 40, 145,711 masks; EP-MCI: n = 38, 132,643 masks), offering more training data for FP-MCI [28]. It may also relate to clear distinguishing features in FP-MCI such as flatter thoraco-lumbar curvature during backward bend [9,30], while EP-MCI presentations tend to be more variable [9,30]. This pattern mirrors previous findings where clinicians also classified FP-MCI more accurately (68%) than EP-MCI (57.2%) [21].

Our findings suggest that backward bend-based postural features may offer valuable discriminative cues for classifying FP- and EP-MCI, potentially complementing the current clinical emphasis on forward bend-based observation [18,21,38]. While clinicians may prioritise movement patterns, these results indicate that posture—though visually less prominent—can provide useful classification indicators.

5.1. Factors influencing classifier performance

Several factors influenced classifier performance. First, image volume played a critical role. Using 80-frame images per participant yielded the highest accuracy (91.91%), outperforming both 40 or 120-frame

configurations (85.34% and 86.38%, respectively). This suggests that 80 images struck an optimal balance between sufficient feature representation and noise avoidance—an issue well-documented in image classification tasks [18].

Second, the greater range of motion in forward bend likely increased variability and error, making feature extraction more difficult [39]. Combining forward and backward bend images led to reduced accuracy, possibly due to conflicting information or model saturation, where static postural features were diluted by excessive input [40–42].

Although multimodal integration is often recommended for complex conditions [42], adding motion or PROMs data did not enhance classification performance. Unlike video-derived kinematic features that capture objective movement patterns, PROMs are low-dimensional, intercorrelated, and sometimes inconsistent with actual motor performance. This may reflect the multifactorial nature of NSLBP where psychosocial factors such as self-efficacy and fear avoidance influence presentation [43,44], making it difficult for PROMs alone to improve detection of physical impairments. Previous studies similarly reported limited benefit of PROMs for identifying movement impairments [25]. These findings reinforce the need for multidimensional approaches to NSLBP classification [30].

Despite robust five-fold cross-validation, the small and relatively homogeneous cohort (N = 78) limits generalisability. Variations in age, sex, body mass index, and motor performance may have influenced movement patterns and predictions, introducing confounding effects that partly explain the remaining misclassifications and wide confidence intervals.

Visual review of the misclassified cases indicated three recurring factors: (i) inter-individual differences in forward-bending strategies, such as variable pelvic-to-lumbar contribution [45]; (ii) transitional or mixed motion around the thoracic–lumbar and lumbar–pelvic junctions, which blur FP/EP distinctions [45]; and (iii) asymmetric trunk control or reduced core stability, producing subtle curvature deviations [46]. These factors indicate that both anatomical variability and motor-control strategies can affect model predictions.

6. Strengths and limitations

This study presents a novel AI-based framework for classifying physical impairments in NSLBP using video-derived features via application of computer vision HPE and HIS methods. A key strength is the use of standard video cameras, supporting feasibility for routine clinical deployment and remote use. The framework demonstrated strong classification performance, indicating potential for guiding tailored rehabilitation based on individual MCI types.

Among the limitations, sample size, while sufficient for initial development, limits generalisability across the broader NSLBP population [9,30,42]. Given the limited dataset (N=78), directly concatenating multimodal features may exacerbate data imbalance and modality dominance effects [47]. Combining forward and backward bend images unexpectedly reduced performance, likely due to conflicting movement characteristics. The challenge of applying this technology in clinical or remote rehabilitation settings is that differences in lighting and camera position can affect feature quality, especially the standardisation of movements during self-recording.

7. Future work

Future studies should validate BackTracker on larger and more diverse NSLBP samples to enhance generalisability. Developing a mobile application that allows patients to record and upload motion videos, enabling self-management based on classification results, would help to democratise data collection and promote wider clinical application. Integrating wearable sensor data may further improve model precision by capturing joint torque and muscle activation, offering a richer biomechanical profile. Future work will explore movement-specific models and advanced fusion strategies, such as attention-based modality gating and latent embedding alignment, together with denoising and refined feature selection to improve robustness and clinical reliability. Combining these sources with video could support more accurate NSLBP classification and enable tailored rehabilitation at the point of need.

8. Conclusion

This study presents a novel AI-driven framework for classifying motor control impairments in NSLBP using computer vision techniques applied to standard video recordings. The model achieved high accuracy, particularly when using sagittal spine silhouettes during backward bend, outperforming motion- and PROM-based inputs. To ensure real-world applicability, future research should validate the approach using videos captured in home and workplace environments and evaluate integration into clinical practice to support earlier, more precise NSLBP management.

CRediT authorship contribution statement

Zebang Liu: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Yulia Hicks:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Liba Sheeran:** Writing – review & editing, Validation, Supervision, Resources, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data for this article can be found online at doi:10.1016/j.ijmedinf.2026.106335.

References

[1] D. Hoy, C. Bain, G. Williams, E. al., A systematic review of the global prevalence of low back pain, Arthritis Rheum. 64 (6) (2012) 2028–2037.
[2] M.L. Ferreira, G.C. Machado, J. Latimer, E. al., Global, regional, and national burden of low back pain, 1990–2020, its attributable risk factors, and projections to 2050: a systematic analysis of the global burden of disease study 2021, Lancet Rheumatol. 5 (6) (2023) e316–e329.
[3] C. Maher, M. Underwood, R. Buchbinder, Non-specific low back pain, Lancet 389 (10070) (2017) 736–747.
[4] National Institute for Health and Care Excellence (NICE), Low back pain and sciatica in over 16s: assessment and management, 2020. https://www.nice.org.uk/guidance/ng59. (Accessed 25 April 2025).
[5] J.A. Hides, W.R. Stanton, S. McMahon, E. al., Convergence and divergence of exercise-based approaches that incorporate motor control for the management of low back pain, J. Orthop. Sports Phys. Ther. 49 (6) (2019) 437–452.
[6] J.H. van Dieën, H. Flor, P.W. Hodges, E. al., Motor control changes in low back pain: divergence in presentations and mechanisms, J. Orthop. Sports Phys. Ther. 49 (6) (2019) 370–379.
[7] N.V. Karayannis, G.A. Jull, P.W. Hodges, Physiotherapy movement based classification approaches to low back pain: comparison of subgroups through review and developer/expert survey, BMC Musculoskelet. Disord. 13 (2012) 1–15.
[8] W.O. Spitzer, F. LeBlanc, M. Dupuis, Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. Report of the quebec task force on spinal disorders, Spine 12 (7 Suppl) (1987) S1–S59.
[9] P. O'Sullivan, Diagnosis and classification of chronic low back pain disorders: maladaptive movement and motor control impairments as underlying mechanism, Man. Ther. 10 (4) (2005) 242–255.
[10] L. Sheeran, colleagues, Classification-guided versus generalized postural intervention in subgroups of nonspecific chronic low back pain: a pragmatic randomized controlled study, 2013. Lippincott Williams & Wilkins.
[11] L. Sheeran, V. Sparkes, M. Bussey, E. al., The effect of classification-based cognitive functional therapy on spinal kinematics and function in subgroups of chronic low back pain, Spine J. 16 (4) (2016) S45–S46.
[12] W. Dankaerts, P. O'Sullivan, A. Burnett, E. al., The use of a mechanism-based classification system to evaluate and direct management of a patient with non-specific chronic low back pain and motor control impairment—a case report, Man. Ther. 12 (2) (2007) 181–191.

[13] K. Vibe Fersum, P. O'Sullivan, J.S. Skouen, E. al., Efficacy of classification-based cognitive functional therapy in patients with non-specific chronic low back pain: a randomized controlled trial, Eur. J. Pain 17 (6) (2013) 916–928.
[14] W. Dankaerts, P. O'Sullivan, A. Burnett, E. al., Differences in sitting postures are associated with nonspecific chronic low back pain disorders when patients are subclassified, Spine 31 (6) (2006) 698–704.
[15] W. Dankaerts, P. O'Sullivan, A. Burnett, E. al., Discriminating healthy controls and two clinical subgroups of nonspecific chronic low back pain patients using trunk muscle activation and lumbosacral kinematics of postures and movements: a statistical classification model, Spine 34 (15) (2009) 1610–1618.
[16] Z. Bacon, T. Hartley, L. Sheeran, E. al., Automatic low back pain classification using inertial measurement units: a preliminary analysis, Proc. Comput. Sci. 176 (2020) 2822–2831.
[17] E. Elgueta-Cancino, S.M. Schabrun, P.W. Hodges, E. al., Characterisation of motor cortex organisation in patients with different presentations of persistent low back pain, Eur. J. Neurosci. 54 (11) (2021) 7989–8005.
[18] R. Hemming, L. Sheeran, V. Sparkes, E. al., Non-specific chronic low back pain: differences in spinal kinematics in subgroups during functional tasks, Eur. Spine J. 27 (2018) 163–170.
[19] L. Sheeran, V. Sparkes, M. Bussey, E. al., Spinal position sense and trunk muscle activity during sitting and standing in nonspecific chronic low back pain: classification analysis, Spine 37 (8) (2012) E486–E495.
[20] L. Sheeran, V. Sparkes, M. Bussey, E. al., Identifying non-specific low back pain clinical subgroups from sitting and standing repositioning posture tasks using a novel Cardiff dempster–shafer theory classifier, Clin. Biomech. 70 (2019) 237–244.
[21] W. Dankaerts, P. O'Sullivan, A. Burnett, E. al., The inter-examiner reliability of a classification method for non-specific chronic low back pain patients with motor control impairment, Man. Ther. 11 (1) (2006) 28–39.
[22] K.V. Fersum, P. O'Sullivan, J.S. Skouen, E. al., Inter-examiner reliability of a classification system for patients with non-specific low back pain, Man. Ther. 14 (5) (2009) 555–561.
[23] B. Cheng, B. Xiao, J. Wang, E. al., HigherHRNet: scale-aware representation learning for bottom-up human pose estimation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
[24] J. Yang, Z. Deng, L. Li, E. al., Track anything: segment anything meets videos, arXiv preprint arXiv:2304.11968, 2023.
[25] T. Hartley, Z. Bacon, L. Sheeran, E. al., BACK-to-MOVE: machine learning and computer vision model automating clinical classification of non-specific low back pain for personalised management, PLOS ONE 19 (5) (2024) e0302899.
[26] P. O'Sullivan, Clinical instability of the lumbar spine, in: Grieve's Modern Manual Therapy: the Vertebral Column, 3 ed, Elsevier, Amsterdam, 2004.
[27] D.R. Fourney, G.B. Andersson, P.M. Arnold, E. al., Chronic low back pain: a heterogeneous condition with challenges for an evidence-based approach, Spine 36 (2011) S1–S9.
[28] Z. Liu, Y. Hicks, L. Sheeran, SpineSighter: an AI-driven approach for automatic classification of spinal function from video, Proc. Comput. Sci. 246 (2024) 3977–3989.
[29] L. Sheeran, M. Robling, Spinal function assessment and exercise performance framework for low back pain, in: Orthopaedic Proceedings, Bone & Joint, 2019.
[30] W. Dankaerts, P. O'Sullivan, The validity of o'sullivan's classification system (CS) for a sub-group of non-specific chronic low back pain with motor control impairment (MCI): overview of a series of studies and review of the literature, Man. Ther. 16 (1) (2011) 9–14.
[31] F. Cabitza, A. Campagner, The need to separate the wheat from the chaff in medical informatics: introducing a comprehensive checklist for the (self)-assessment of medical AI studies, Int. J. Med. Inform. 153 (2021) 104510.
[32] G. Christe, L. Redhead, P.W. Hodges, E. al., Patients with chronic low back pain have an individual movement signature: a comparison of angular amplitude, angular velocity and muscle activity across multiple functional tasks, Front. Bioeng. Biotechnol. 9 (2021) 767974.
[33] A. Vaswani, N. Shazeer, N. Parmar, E. al., Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).
[34] A. Dosovitskiy, L. Beyer, A. Kolesnikov, E. al., An image is worth 16 × 16 words: transformers for image recognition at scale, arXiv preprint arXiv:2010.11929, 2020.
[35] M. Raghu, T. Unterthiner, S. Kornblith, E. al., Do vision transformers see like convolutional neural networks? in: Proceedings of the 35th International Conference on Neural Information Processing Systems, Curran Associates Inc., 2021. Article 927.
[36] Z.-H. Zhou, J. Feng, Deep forest, Natl. Sci. Rev. 6 (1) (2019) 74–86.
[37] D. Chicco, G. Jurman, The advantages of the matthews correlation coefficient (MCC) over f1 score and accuracy in binary classification evaluation, BMC Genomics 21 (1) (2020) 6.
[38] R.G. Astfalck, P. O'Sullivan, L. Straker, E. al., A detailed characterisation of pain, disability, physical and psychological features of a small group of adolescents with non-specific chronic low back pain, Man. Ther. 15 (3) (2010) 240–247.
[39] D.A. Winter, Biomechanics and Motor Control of Human Movement, John Wiley & Sons, 2009.
[40] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (1) (2019) 1–48.
[41] A. Kline, M. Afshar, A. Chen, E. al., Multimodal machine learning in precision health: a scoping review, npj Digit. Med. 5 (1) (2022) 171.
[42] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, vol. 112, Springer, 2013.
[43] C. Kehl, J. Düssler, D. Pieper, E. al., Associations between pain-related fear and lumbar movement variability during activities of daily living in patients with chronic low back pain and healthy controls, Sci. Rep. 14 (1) (2024) 22889.
[44] T. Miki, K. Ota, H. Suzuki, E. al., Factors associating with disability of non-specific low back pain in different subgroups: a hierarchical linear regression analysis, Sci. Rep. 11 (1) (2021) 18278.
[45] L. Sheeran, Y. Hicks, C.-Y. Lin, E. al., Assessment of spinal and pelvic kinematics using inertial measurement units in clinical subgroups of persistent non-specific low back pain, Sensors 24 (7) (2024) 2127.
[46] V. Mohammadi, A. Letafatkar, A.A. Jafarnezhadgero, Effects of core stability training on kinematic and kinetic variables in patients with chronic low back pain, Phys. Treat. – Specif. Phys. Ther. 13 (1) (2023) 55–66.
[47] T. Baltrušaitis, C. Ahuja, L.-P. Morency, Multimodal machine learning: a survey and taxonomy, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2017) 423–443.