The Journal of Anatomical Sciences
Email: journalofanatomicalsciences@gmail.com

J. Anat Sci 17(2) June

Submitted: April 5th, 2026
Revised: May 28th, 2026
Accepted: June 5th, 2026

Machine learning model for predicting the risk of back pains among porters in Singer and Kurmi Market, Kano State, Nigeria

*Musa Ado Bashir and Usama Umar Aliyu

Department of Human Anatomy, Faculty of Basic Medical Sciences, Bayero University Kano, Nigeria

***Corresponding Author:** Email: musabashir34@gmail.com;
Tel: +234 703 687 8467

ABSTRACT

Porters in large Nigerian markets are routinely exposed to repetitive lifting and load carriage. This study aims to build and deploy a machine learning model to predict the risk of back pain from anthropometric parameters. A descriptive cross-sectional study was conducted among 300 active porters. Back pain was measured using an interviewer-administered questionnaire adapted from the Standardized Nordic Musculoskeletal Questionnaire. Body weight, height, and circumferences of neck, calf, and waist were measured using standardized protocols. The Tidymodels framework of the R software was used for all stages of the model development. The model was deployed as an R Shiny web application and hosted on shinyapp.io. The mean age, height, weight, body mass index, waist circumference, and waist-to-height ratio of porters with back pains were respectively 26.4 ± 7.2 years, 1.70 ± 0.07 meters, 64.8 ± 11.1 kg, 22.4 ± 4.1 kg/m^2, 82.9 ± 7.6 cm, and 0.49 ± 0.05. The respective values for porters without back pains were 24.8 ± 6.1 years, versus 1.74 ± 0.08 meters, 62.4 ± 9.3 kg, 21.0 ± 3.5 kg/m^2, 81.5 ± 8.2cm, and 0.47 ± 0.05. The prevalence of back pain among the subjects was 87.3%. The sensitivity, specificity, overall accuracy, and area under the ROC curve of the model for the test data were found to be 85%, 42%, 79% and 66%, respectively. In conclusion, predictive analytics using simple anthropometric measurements has the potential to be used in screening the risk of back pain among porters in the large markets of Kano State, Nigeria.

Keywords: porters, back pain, machine learning, Kano, shiny application

INTRODUCTION

Back pain, particularly low back pain, is one of the most common musculoskeletal and occupational health problems globally. The World Health Organization has identified low back pain as the leading cause of disability worldwide, underscoring its importance as a major public health concern[1]. In occupational settings, repeated lifting, bending, twisting, and prolonged physical loading increase spinal stress and contribute substantially to work-related back pain[2].

In low- and middle-income countries, porters represent an occupational group that is especially vulnerable to back pain because their daily work is heavily dependent on physical strength and repetitive manual material handling. In major open markets, porters routinely carry heavy goods over short and long distances under poor ergonomic conditions and with little occupational health protection. Nigerian evidence has shown that manual material handlers and market-based load porters bear a high burden of musculoskeletal disorders, particularly in the lower back, neck, and shoulders[3,4].

Anthropometric parameters such as body weight, height, waist circumference, and ratios derived from them, such as waist to height ratio and body mass index, are known to increase the risk of back pains[5,6]. While the reason for this association has traditionally been linked to mechanical loading properties of excessive weight on vertebral column structures (intervertebral discs, facet joints, and paraspinal muscles)[7–9], recent reviews have implicated metabolic dysregulation, chronic inflammation, and adipokines secreted by fat cells[10].

Traditional statistical approaches are useful for identifying associations between risk factors and outcomes, but they may be limited when several interacting predictors are present. Machine learning provides a complementary framework for risk prediction because it can integrate multiple predictors, support classification, and improve decision-making in complex datasets[11,12]. In occupational health, such approaches are increasingly relevant for screening workers at high risk and for targeting prevention strategies. Machine learning has emerged as a useful method in musculoskeletal health research because it can model non-linear relationships, rank predictors, and support classification tasks. Bhak and colleagues developed machine learning models for low back pain detection using a more robust predictor set, including dietary information and level of physical activity. They reported model performance of 80% (ROC is under the curve)[12]. Snyder and co-workers applied deep learning to detect unsafe lifting behavior and hence the risk of back pain in an industrial setting[13]. Together, these studies indicate that computational modeling approaches can complement conventional epidemiological analysis in back-pain research.

The study environment of Singer and Kurmi Markets in Kano State provides an important setting for investigating this problem. These markets are among the most active commercial centers in northern Nigeria and depend heavily on informal labor for the movement of goods[14,15]. Porters in these markets are frequently exposed to cumulative biomechanical stress through lifting, carrying, trunk flexion, awkward postures, and long working hours[16]. Despite these risks, published evidence specific to porters in Kano remains limited. Published studies of back-pain prediction among porter populations in the Singer-Kurmi market context in Kano State are virtually non-existent, hence the need for this study.

MATERIALS AND METHODS

Study design

The study is a descriptive cross-sectional

Study area and population

This study was carried out in Singer and Kurmi Markets, Kano metropolis, Kano State, Nigeria. Kano is one of the major commercial centres in northern Nigeria and remains a regional hub for trade, transportation, and small-scale enterprise. Singer and Kurmi Markets were selected because of their high concentration of buying and selling activities and the regular use of manual porters for transporting goods within and around the markets. The study population comprised active porters working in the two markets during the period of data collection. Porters who were actively engaged in carrying goods, consented to participate, and were physically able to undergo anthropometric measurements were included. Individuals who declined consent, had obvious deformity or severe illness, or could not complete the interview and measurements were excluded.

Sample size calculation

The minimum sample size of one hundred and seventy-four (174) was calculated using Epitools[17] online calculator of sample size for prevalence studies based on a previously reported 87% prevalence of musculoskeletal disorders reported among load carriers in Kwari, Sabon-gari, and Singa markets of Kano metropolis, Nigeria[18], 95% confidence level, and a precision rate of 0.05. A total of 300 porters were eventually recruited into the study to maximize the performance of the developed machine learning model.

Sampling methods

A purposive non-probability sampling approach was used. Singer and Kurmi Markets were purposively selected, after which eligible porters present during data collection were approached consecutively until the sample size was attained. This approach was considered appropriate because the porter population is mobile and operates in an informal work environment without a complete sampling frame.

Ethical considerations

Ethical approval was sought and obtained from the Bayero University Health Research Ethics Committee (BUK-HREC), with approval reference number [NHREC/BUK-HREC/773/10/23II]. The study was conducted in accordance with the ethical standards of the institution. All human subjects involved in this research provided informed consent before participation, and strict confidentiality and anonymity were maintained throughout the study.

Anthropometric measurements

Body weight

Body weight was measured using an electronic weighing scale (Seca, Germany) having an accuracy of 0.1kg. With the scale reading zero, weight was read with the subject standing on the centre of the scale without support and with the weight distributed evenly on both feet[19].

Height

Height was measured using a portable stadiometer (Seca, Germany) with a maximum of 200cm and an accuracy of 0.1cm. The subjects were asked to stand with the feet together and the heels, buttocks, and upper part of the back against the wall, and the head placed in the Frankfort plane[19].

Body mass index

The body mass index (BMI) was calculated as weight (in kilograms) divided by the square of height (in meters). It is measured in kg/m2.

Waist circumferences

Waist circumference was measured using an inelastic measuring tape (MEDLINE, 1- 800-Medline, USA). Subjects were asked to assume a relaxed standing position with the arms folded across the thorax, measurement was taken at the mid-point between the lower costal (10th rib) border and the iliac crest at the end of expiration[19].

Waist-to-height ratio

Waist-to-height ratio was calculated as waist circumference in meters divided by height in meters. It is a dimensionless quantity.

Back pains measurements

Back pain was assessed by adapting the Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms[20]. Back pain was recorded as any back pain (lower back, upper back, or posterior neck).

Machine learning modeling

Tidymodels framework[21] of the R statistical environment[22] was used in data processing, feature selection, hyperparameter tuning, model fitting, and model evaluation. The steps followed were:

i. Splitting the whole data into training and testing data based on 70%:30% ratio.

ii. Specifying a tidymodels recipe for preprocessing. This step included specifying the form of the formula used in the model, normalization of all of the predictors, application of SMOTE[23] algorithms within a cross-validation to handle class imbalance with tuned parameters for over-ratio and 5 for the number of neighbors, and setting the polynomial order of the height variable as a tunable hyperparameter

iii. Specifying the tidymodels model as a gradient boosted decision tree ensemble from “xgboost” engine.

iv. Building a tidymodels workflow object and adding the recipe and model to it.

v. Creation of a grid for the height variable polynomial order as 10 integers from 1 to 10, tree depth in the range 3 to 8, and SMOTE over-sampling ratio from 1 to 3

vi. Creation of resamples for cross-validation with 4 folds.

vii. Tuning the hyperparameters and training all the models based on the possible hyperparameter values

viii. Extraction of the best hyperparameters

ix. Fitting the model with the best hyperparameters

x. Assessing the prediction quality based on the test data using sensitivity, specificity, positive and negative predictive values, accuracy, and area under the ROC curve.

Deployment of the model

The machine learning model developed was deployed as a web application using the R Shiny web application framework[24]. The website was hosted on the shinyapp.io platform[25] with the URL of the website shared to potential users.

RESULTS

Age and anthropometric parameters distribution of the subjects

Out of the 297 subjects who had complete data for analysis, 257 (87%) reported back pains, whereas 40 (13%) did not report any back pains. Three (3) subjects have incomplete data on the outcome of interest (back pains) and are thus excluded from the analysis. Figure 1 shows, in the form of box plots, the distribution of the anthropometric parameters among those with back pains (coded “1”) and those without back pains (coded “0”).

In general, compared with those without back pains, porters with back pains are older (age = 26.4 ± 7.2 years versus 24.8 ± 6.1 years), shorter (height = 1.70 ± 0.07 meters versus 1.74 ± 0.08 meters), heavier (weight = 64.8 ± 11.1 kg versus 62.4 ± 9.3 kg) and have higher BMI (22.4 ± 4.1 kg/m2 versus 21.0 ± 3.5 kg/m2), waist circumference(82.9 ± 7.6 cm versus 81.5 ± 8.2cm) and waist-to-height ratio(0.49 ± 0.05 versus 0.47 ± 0.05).

Output of model training and evaluation

The results of the hyperparameter tuning were shown in Table 1. polynomial degree and tree depth of 1, and a SMOTE over-sampling ratio of 1 were found to be the best based on the area under the ROC curve. So 151 negative synthetic samples were generated, and the resulting class proportion was 50:50. The final model developed was of the form:

$$BackPains \sim Age + Weight + Height + BodyMassIndex$$

$$+ WaisttoHeightRatio + NeckCircumference + CalfCircumference$$

The confusion matrix of the model performance on training is shown as a heatmap image in Figure 2. The corresponding heatmap image of the model performance on testing data is shown in Figure 3, which visually reveals the failure of the model to classify the negative outcome (poor specificity).

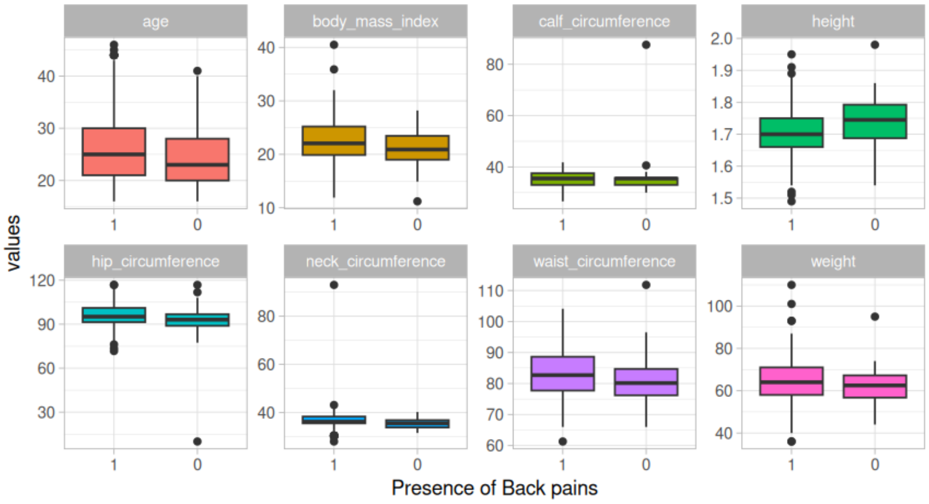

Figure 1: Age and Anthropometric Parameters of Porters with Back Pains (coded '1') and those without Back Pains (coded '0')

Table 1: Result of the Hyperparameter Tuning

Hyperparameter	Tuned_Value
Tree Depth	1
SMOTE Oversampling Ratio	1
Polynomial Degree of Height Variable	1

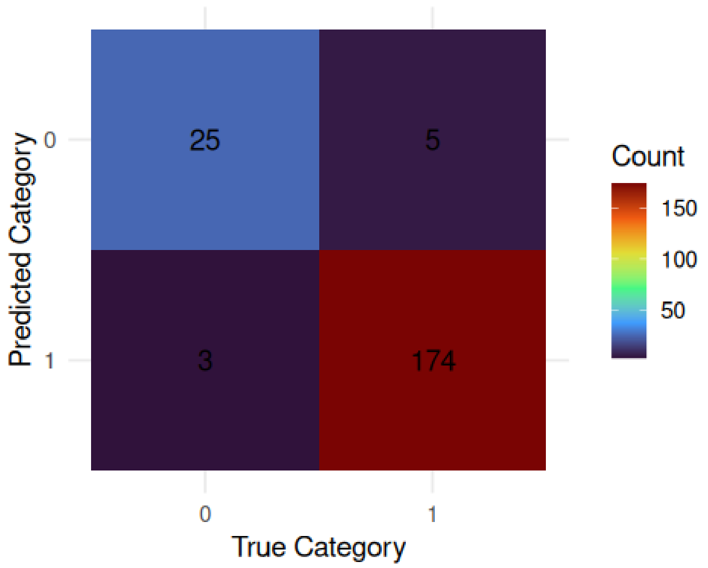

Figure 2: Confusion Matrix Heatmap of Model Performance on Training Data. Presence of Back Pains coded as 1 and Absence of Back Pains coded as 0.

The computed performance metrics of the trained model on training and unseen test data are shown in Table 2. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F-1 score, and accuracy of the model on the training data were, respectively 85%, 42%, 90%, 29% and 87%. The respective values on the test data were 97%, 89%, 98%, 83%, 98% and 97%. The overall area under the curve of the model was 99% and 66% on the training and test data, respectively (Figure 4).

The R Shiny web application for deploying the model

Figure 5 shows the homepage of the web application used to deploy the model. The required parameters to make a prediction were age, height, weight, and waist circumference. Figure 6 shows the same application after a prediction was generated.

The application can also make batch predictions for a list of subjects prepared as a comma-delimited file with the required columns, as shown in Figure 7. Figure 8 shows the application after the required file was uploaded and the predictions generated.

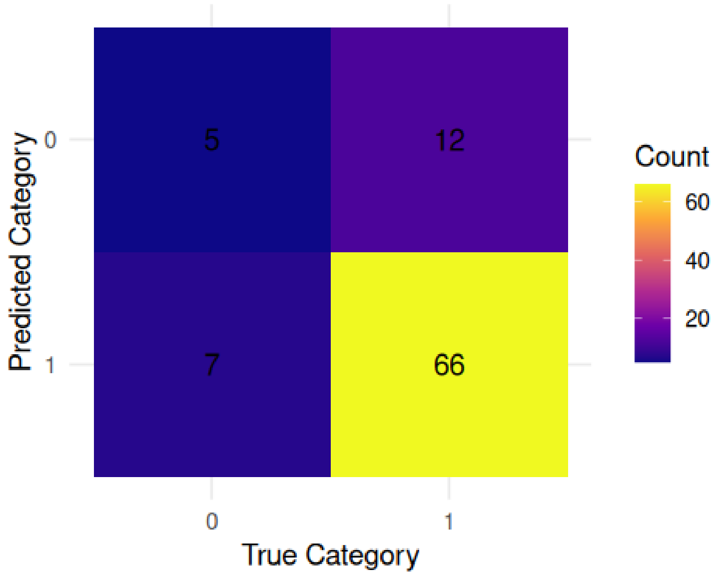

Figure 3: Confusion Matrix Heatmap of Model Performance on Testing Data. Presence of Back Pains coded as 1 and Absence of Back Pains coded as 0

Table 2: Performance Measures of the Model on Training and Testing Data

Dataset	Sensitivity	Specificity	PPV	NPV	F-1.Score	Accuracy	AUC
Training	97	89	98	83	98	97	99
Testing	85	42	90	29	87	79	66

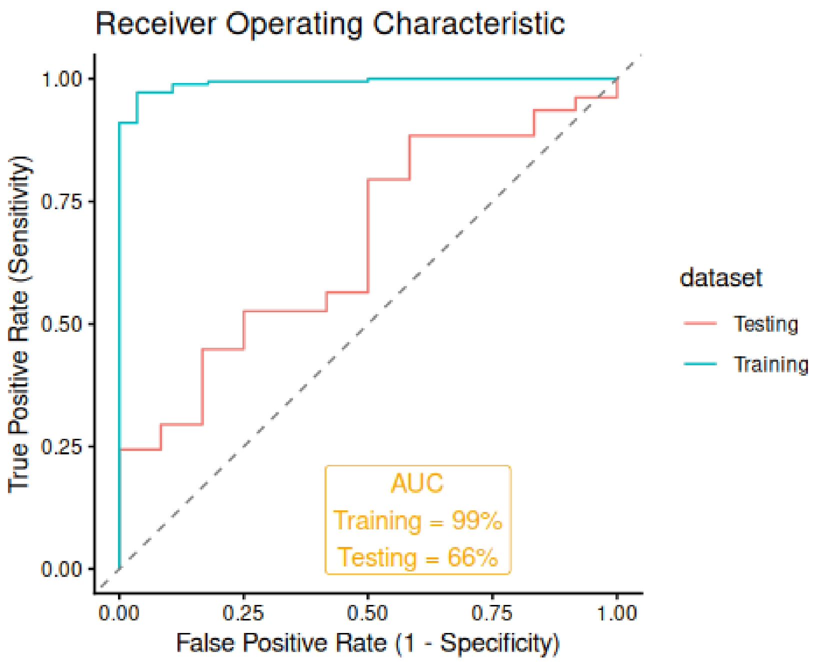

Figure 4: Receiver Operating Characteristics Curve for Training and Testing Data

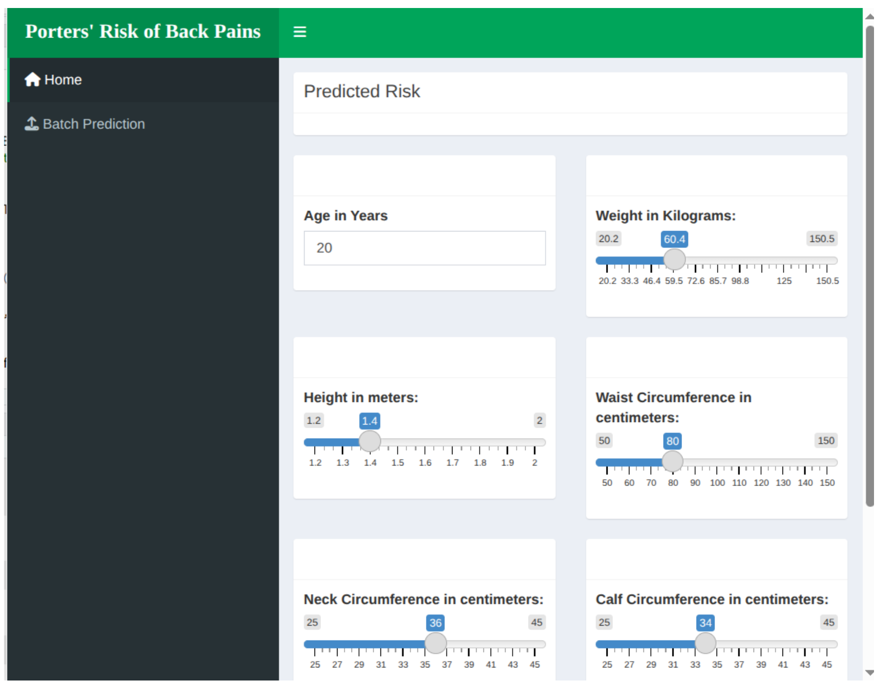

Figure 5: The Application before Prediction is Generated

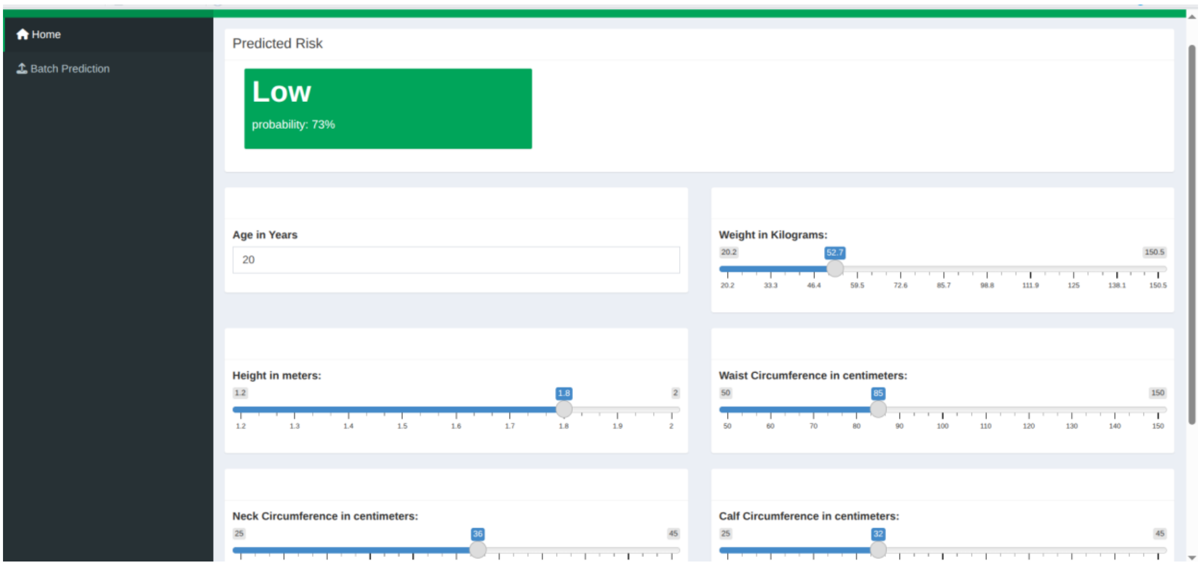

Figure 6: The Application after Prediction is Generated

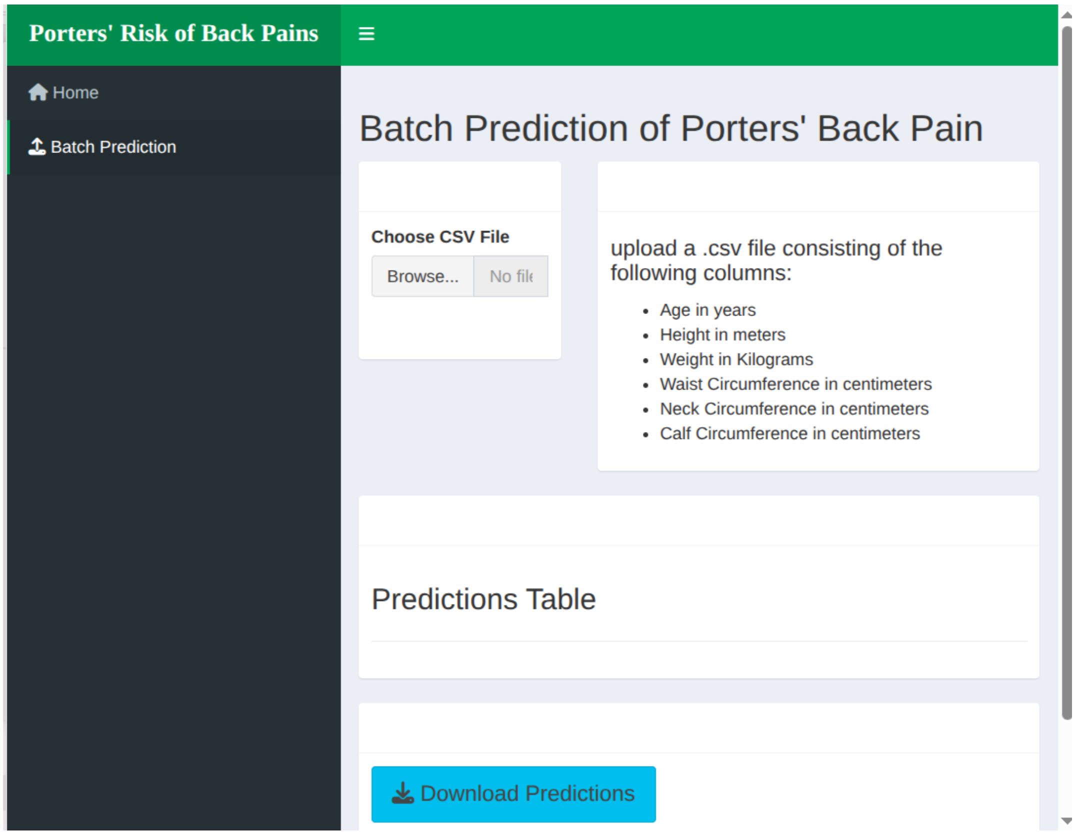

Figure 7: Interface for Batch Prediction of Back Pains

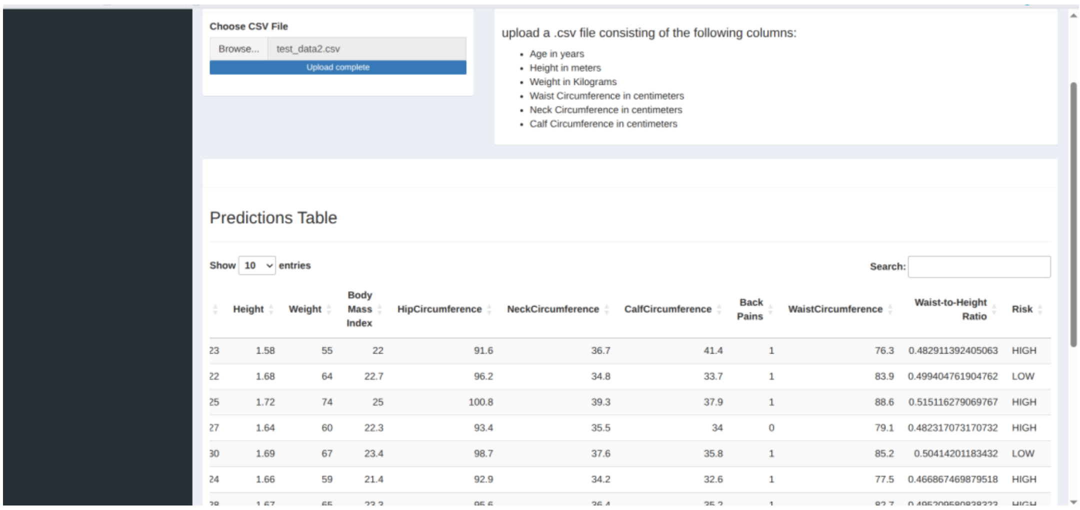

Figure 8: Batch Prediction of Back Pains

DISCUSSION

High sensitivity, positive predictive value, F-1 score, and accuracy, and low specificity and negative predictive value of the model

The high sensitivity, positive predictive value, F-1 scores, and accuracy on the test data paint a picture of high model performance. For example, a sensitivity of 85% means the model correctly predicts 85% of those with back pains as having the risk of back pains, while misclassifying the 15% at having low risk of back pains[26]. On the other hand, positive predictive value measures the worth of positive prediction[27]. The interpretation of a positive predictive value of 90% in the test data is that when the model predicts that a particular porter is at risk of back pain, then almost certainly they are. Such a worker can then be targeted for various preventive measures to lower their risk accordingly[28].

The high positive predictive value of the model in both training and test data (97% and 91%, respectively) is a mere reflection of the fact that the intended outcome (back pains) is highly prevalent in the study population[29]. This is because positive predictive value has a direct positive relationship with the prevalence of the outcome of interest. This is directly attributable to low false positive cases in such situations[30]. The high prevalence of back pains among the study subjects is consistent with occupational literature, which repeatedly shows that heavy lifting, repetitive bending, non-neutral postures, and manual handling are important determinants of back pain among such and similar categories of workers with exposure to repeated manual load handling. [2–4,31–33]. This high prevalence of back pains resulted in a high positive predictive value despite the use of the SMOTE algorithm, which was intended to counter the class imbalance and hence reduce positive predictive value[34]. The SMOTE algorithm used on the training data resampled those without back pains to have the same frequency as the class of subjects with back pains. The same high positive predictive value was reported on testing data on which the SMOTE algorithm was not applied (to avoid data leakage). This further confirms that the high positive predictive value results from a property of the dataset (high prevalence of back pains in particular).

The low specificity and negative predictive value on the test data (42% and 29%, respectively) indicate the model's poor ability to discriminate and predict the absence of back pain. A specificity of 42% means the model correctly predicts 42% of those without back pains as not having back pains while misclassifying the remaining 58% (false positives). The negative predictive value is an entirely different metric measuring the worth of a negative prediction, i.e., among those the model predicts as having no back pains, only 29% are really free from back pains, misclassifying the remaining 71% (false negatives)[26,27]. This poor performance in the negative class reflects the class imbalance. The classifier learned to predict the majority class most of the time. A high sensitivity and low specificity model might not be a major problem for a screening classifier such as the one in this study, but only if the overall accuracy of the model is high.

Low area under the ROC curve and the problem of class imbalance in the dataset

The high accuracy rate and F-1 scores on the test data (79% and 87%, respectively) give an impression of high model performance. However, they can also indicate a mathematical illusion because of the earlier-mentioned imbalance. This is because F1-score completely ignores the true negative count, thereby over-optimistically inflating performance metrics when evaluating heavily skewed distributions and hence masking the model's inability to see the rare minority group[35].

The low area under the ROC curve (66%) on the test data reveals the real vulnerability of the model. Such values are deemed poor[36] or at most fair[37]. The low value of the AUROC curve contrasts sharply with other measures of performance, such as the F-1 score and overall accuracy. The explanation for this poor performance can be attributed to either the class imbalance or the fact that the features (anthropometric parameters) occupy the same multidimensional space with no clear separating boundary.

Failure of SMOTE and tuned XGBoost to improve the classification performance of the model

SMOTE-Tomek and tuned XGBoost are industry-standard pipelines[34,38] used to address the class imbalance problem. A look at the performance of the model on the training data showed an 89% specificity and an 83% negative predictive value. This shows that class imbalance was addressed in the training data set. However, because SMOTE was not allowed in the testing data set (data leak prevented), the performance of predicting negative cases degraded significantly when the model was applied to the testing data. It appears that severe mathematical class overlapping (not class imbalance) is the true limiting bottleneck of the classifier performance. When features occupy the same multidimensional space (a dense and noisy feature space), data resampling techniques have zero impact on final testing parameters. This means class imbalance is a secondary issue; no amount of resampling will improve the AUC if the features lack discriminative capability[34,38].

The potential and limits of traditional anthropometry in predicting the risk of back pains

The role of anthropometric parameters in predicting risk of back pains has been well-documented [7–9,34,38] and explained as biologically plausible (through the effects of mechanical loading on intervertebral discs, paraspinal muscles, and facet joints, as well as through metabolic dysregulation and chronic inflammation)[10]. However, human physical dimensions exhibit massive collinearity (e.g., height scales with arm span; weight scales with waist circumference). When features are highly correlated but have a low signal-to-noise ratio regarding the specific target variable, XGBoost treats the classes as physically overlapping. Furthermore, traditional anthropometry captures size variation and misses most of the shape variation among individuals[39,40]. Consequently, XGBoost can easily map overall body size, but it completely misses body shape topology or tissue distribution[40]. Because the minority negative class is obscured by basic size variations, the model collapses back to its majority prediction pattern and hence has poor performance regarding predicting negative cases. In other words, human external body dimensions exhibit profound phenotypic overlap, and thus the low-to-moderate AUC is a reflection of biological reality rather than algorithmic failure.

Limitations of the research

This study demonstrated that applying a machine-learning framework that can rank traditional anthropometric features, handle interactions, and assist early identification of porters at elevated risk. However, despite deploying the SMOTE algorithm, hyperparameter tuning, and domain ratio scaling, the testing AUC consistently bounded at 66%. Therefore, traditional anthropometry combined with age provides a modest baseline signal, but is fundamentally insufficient for highly specific minority-class isolation without other variables such as shape variables (from geometric morphometry), load weight, carrying duration, distance traveled, and possibly internal metabolic parameters.

CONCLUSION

The study demonstrates that predictive analytics using simple anthropometric measurements has the potential to be used in screening the risk of back pain among porters in the large markets of Kano State, Nigeria. However, the low successful classification rates (ROC area under the curve) on the test data indicate that anthropometry alone is unlikely to capture the entire risk process.

Conflict of interest: The authors have no conflict of interest to declare

Source of funding: none

Acknowledgment: The authors acknowledge the Department of Human Anatomy, Faculty of Basic Medical Sciences, Bayero University Kano, and the leadership and participating porters in Singer and Kurmi Markets for their cooperation during the conduct of this study.

Authors' contribution: MAB: research conception, design, model training, and deployment; UUA: data acquisition, data cleaning, manuscript writing

REFERENCES

1. World Health Organization. WHO releases guidelines on chronic low back pain [Internet]. 2023 Dec 7 [cited 2026 Apr 4]. https://www.who.int/news/item/07-12-2023-who-releases-guidelines-on-chronic-low-back-pain

2. Kwon B, Roffey D, Bishop P, Dagenais S, Wai E. Systematic review: occupational physical activity and low back pain. Occup Med (Lond). 2011;61(8):541-548.

3. Tabiti OO, Ayomikun OO, Fapojuwo O, Adeyemi T. Prevalence and pattern of work-related musculoskeletal disorders among load carriers in Ibadan metropolis. J Biomed Investig. 2024;12(3):140-150.

4. Obed-Ojukwu LE, Dike NC. Prevalence of work-related musculoskeletal disorders among male manual material handling workers in markets across the Port Harcourt metropolis. Trends Educ Stud J. 2025;17(1):1-12.

5. Stribrny M, Svejdova M, Busch A, Novak J, Kavka T, Kobesova A. Anthropometric evaluation of trunk parameters in chronic low back pain: a comparative study with age- and sex-matched controls. BMC Musculoskelet Disord. 2025;26(1):959.

6. Saludes R, Acevedo P, García IZ, Gómez Carrión Á, Martínez Sebastián C, Núñez Fernández A, *et al*. Abdominal adiposity increases lordosis and doubles the risk of low back pain. Appl Sci. 2022;12(15):7616.

7. Heuch I, Heuch I, Hagen K, Zwart JA. A comparison of anthropometric measures for assessing the association between body size and risk of chronic low back pain: the HUNT study. PLoS One. 2015;10(10):e0141268.

8. You Q, Jiang Q, Li D, Wang T, Wang S, Cao S. Waist circumference, waist-hip ratio, body fat rate, total body fat mass and risk of low back pain: a systematic review and meta-analysis. Eur Spine J. 2022;31(1):123-135.

9. Shiri R, Karppinen J, Leino-Arjas P, Solovieva S, Viikari-Juntura E. The association between obesity and low back pain: a meta-analysis. Am J Epidemiol. 2010;171(2):135-154.

10. Ruiz-Fernandez C, Schol J, Ambrosio L, Sakai D. The vicious cycle of obesity and low back pain: a comprehensive review. Appl Sci. 2025;15(12):6660.

11. Garg A, Kapellusch J, Hegmann K. 0097 Risk factors for occupational low back pain, medicine use, and seeking care for low back pain: results from a prospective cohort study. Occup Environ Med. 2017;74(Suppl 1):A27.

12. Bhak Y, Ahn TK, Peterson TA, Han HW, Nam SM. Machine learning models for low back pain detection and factor identification: insights from a 6-year nationwide survey. J Pain. 2024;25(8):104497.

13. Snyder K, Thomas B, Lu ML, Jha R, Barim MS, Hayden M, *et al*. A deep learning approach for lower back-pain risk prediction during manual lifting. PLoS One. 2021;16(2):e0247162. doi:10.1371/journal.pone.0247162.

14. Greenhill CJ. Kurmi Market: the final report of the Kurmi Market study team on the future of the Kurmi Market, Kano City. Kano (Nigeria): Metropolitan Kano Planning and Development Board; 1972.

15. Ibrahim AM. Evolutionary trend, spatial distribution of, and issues associated with markets in Kano metropolis. Int J Phys Hum Geogr 2015;3(2):9-24.

16. Giginyu IM. Sorry tales of Kano Singer Market goods uploaders. Daily Trust [Internet]. 2023 Dec 2 [cited 2026 Jun 22]. https://dailytrust.com/sorry-tales-of-kano-singer-market-goods-uploaders/

17. Ausvet. Sample size to estimate a proportion or apparent prevalence with specified precision [Internet]. Epitools. 2026 [cited 2026 Apr 4]. https://epitools.ausvet.com.au/oneproportion

18. Bello B, Alhaji MA, Tamasini NA. Prevalence and pattern of musculoskeletal disorders among load carriers in Kano metropolis. Niger J Exp Clin Biosci. 2017;5(2):31. doi:10.4103/njecp.njecp_6_17

19. Centers for Disease Control and Prevention, National Center for Health Statistics. NHANES August 2021-August 2023 procedure manuals [Internet]. Hyattsville (MD): National Center for Health Statistics; 2023 [cited 2026 Apr 4]. https://cdc.gov

20. Kuorinka I, Jonsson B, Kilbom A, Vinterberg H, Biering-Sørensen F, Andersson G, *et al*. Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. Appl Ergon. 1987;18(3):233-237. doi:10.1016/0003-6870(87)90010-X.

21. Kuhn M, Wickham H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles [Internet]. 2020 [cited 2026 Apr 4]. Available from: https://www.tidymodels.org

22. R Core Team. R: a language and environment for statistical computing [Internet]. Vienna: R Foundation for Statistical Computing; 2024 [cited 2026 Apr 4]. Available from: https://www.R-project.org/

23. He H, Garcia EA. Learning from imbalanced data. IEEE Trans Knowl Data Eng. 2009;21(9):1263-1284. doi:10.1109/TKDE.2008.239.

24. Chang W, Cheng J, Allaire JJ, Sievert C, Schloerke B, Aden-Buie G, *et al*. Shiny: web application framework for R [Internet]. 2026 [cited 2026 Apr 4]. Available from: https://CRAN.R-project.org/package=shiny

25. Posit Software, PBC. Shinyapps.io [Internet]. 2026 [cited 2026 Apr 1]. Available from: https://www.shinyapps.io/

26. Altman DG, Bland JM. Diagnostic tests 1: sensitivity and specificity. BMJ. 1994;308(6943):1552.

27. Altman DG, Bland JM. Diagnostic tests 2: predictive values. BMJ. 1994;309:102.

28. Williams C, Sahagún BE, Su MK. Biostatistics and epidemiology principles for the toxicologist: the “testy” test characteristics part II: positive predictive value and negative predictive value. J Med Toxicol. 2023;19(3):298-302.

29. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity, and predictive values. Indian J Ophthalmol. 2008;56(1):45-50. doi:10.4103/0301-4738.37595.

30. Lau CS, Aw TC. Disease prevalence matters: challenge for SARS-CoV-2 testing. Antibodies (Basel). 2021;10(4):50. doi:10.3390/antib10040050.

31. Sauter M, Barthelme J, Müller C, Liebers F. Manual handling of heavy loads and low back pain among different occupational groups: results of the 2018 BIBB/BAuA employment survey. BMC Musculoskelet Disord. 2021;22(1):956. doi:10.1186/s12891-021-04819-z.

32. Onuah IA, Okoh MN, Maduku EJ, Alabo GI, Bestman AM, Gudi JA. Prevalence of back pain among baggage handlers at Port Harcourt International Airport, Nigeria. Orthop Res Surg. 2025;5(1).

33. Eke EI. Prevalence of back pain among casual dockworkers in Nigerian ports. Int J Econ Environ Dev Soc. 2023;4(2):203-212.

34. Batista GE, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor Newsl. 2004;6(1):20-29.

35. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient over F1 score and accuracy in binary classification evaluation. BMC Genomics. 2020;21(1):6.

36. Çorbacıoğlu ŞK, Aksel G. Receiver operating characteristic curve analysis in diagnostic accuracy studies: a guide to interpreting the area under the curve value. Turk J Emerg Med. 2023;23(4):195-198.

37. Šimundić AM. Measures of diagnostic accuracy: basic definitions. EJIFCC. 2009;19(4):203.

38. Abd Elrahman SM, Abraham A. A review of the class imbalance problem. J Netw Innov Comput. 2013;1:9.

39. World Health Organization. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. Geneva: World Health Organization; 1995.

40. Slice DE. Geometric morphometrics. Annu Rev Anthropol. 2007;36:261-281.