How Well can Machine Learning Models Classify Individuals with and without Non-Specific Chronic Neck Pain based on Cervical Movements during Protraction and Retraction?

Ui-jae Hwang, PT, Ph.D.; Jun-hee Kim, PT, Ph.D.

1Department of Physical Therapy, College of Health Science, Laboratory of KEMA AI Research (KAIR), Yonsei University, Wonju, South Korea

INTRODUCTION

The human cervical spine, a complex structure responsible for supporting and facilitating the movement of the head, is susceptible to degenerative changes due to repetitive motions and extended postures required for daily activities. These degenerative alterations are particularly pronounced in cases of non-specific chronic neck pain (NSCNP). It is widely accepted that deviations in alignment and movement patterns can trigger the onset of painful and degenerative conditions. The purpose of this study was to explore the relationship between cervical movements during protraction and retraction and NSCNP using machine learning (ML) techniques for classification.
Classification NSCNP Using Machine Learning

conditions in the cervical spine.  
Cervical protraction and retraction, representing a combination of upper cervical extension and lower cervical flexion, play pivotal roles in cervical spine motion. Previous research has measured total head excursion, encompassing cervical protraction and retraction, as a sagittal translational motion in centimeters. Notably, individuals with neck pain exhibited distinct patterns, with a 7.5 cm total head excursion compared to 10.9 cm in healthy individuals. Moreover, retraction displayed a significantly greater range in the neck pain group. Cervical protraction and retraction have also been quantified as craniovertebral angles (CVAs) in healthy individuals. Given the ubiquity of cervical protraction in everyday activities and the therapeutic use of retraction to address neck pain, these movements present as valuable variables for classifying neck pain. Thus, it is reasonable to suspect a correlation among cervical protraction, retraction, and neck pain. However, despite the growing emphasis on proper movement, motor control, and coordination, it remains uncertain whether upper and lower cervical movements during cervical protraction and retraction differ between individuals with and without neck pain. Additionally, quantitative assessments of these kinematics in individuals with neck pain during cervical protraction and retraction are conspicuously absent from existing literature.

Previous studies investigating the linear relationship between neck pain and movement have produced inconsistent results, possibly indicative of a nonlinear relationship. Traditional methods such as linear and logistic regression may be ill-suited to capture such nonlinear relationships. In response, machine learning (ML) approaches are being increasingly adopted for classification tasks. The strength of ML lies in its ability to model both linear and highly nonlinear relationships, potentially yielding superior accuracy compared with conventional statistical methods.

Consequently, this study aimed to achieve two primary objectives: 1) to develop, assess, and compare the predictive performance of statistical ML models for classifying individuals with and without NSCNP using cervical movements during protraction and retraction and 2) to investigate the association between cervical movements and NSCNP.

METHODS

Participants

A total of 277 public service office workers (POWs) who did not have NSCNP and 463 POWs with NSCNP were carefully screened for eligibility. The inclusion criteria were that POWs had been employed in office settings, used computers for more than two years, and were recruited between September 2022 and March 2023. The eligibility criteria for individuals with NSCNP included two main factors: (1) they had to report an average neck pain intensity rating exceeding 3 out of 10 on a Numerical Rating Scale (NRS) for the preceding four weeks, and (2) neck pain intensity rating exceeding 2 out of 4 on the Northwick Park Neck Pain Questionnaire (NPQ). Asymptomatic individuals were eligible if they had not experienced neck pain that warranted treatment from a healthcare professional within the past two years. Exclusion criteria for both POWs, with or without NSCNP, encompassed a history of prior spinal surgery, rheumatologic conditions, ongoing or chronic respiratory conditions, or active compensation claims related to injuries. To conduct a statistical analysis of the impact of NSCNP involving up to eight variables, a sample size of no less than 80 participants was deemed necessary in accordance with the common rule of using one variable per 10 events. The data utilized in this study were derived from musculoskeletal screening tests conducted to prevent industrial accidents among POWs. These data were collected from April 2022 to February 2023, with a focus on examining the risk factors associated with musculoskeletal disorders in public service offices. It is important to note that for this study, informed consent was waived by the Yonsei University Mirae Campus Institutional Review Board prior to the commencement of data queries and analyses (IRB number:1041849-202306-BM-100-01). This waiver was granted because the study involved the analysis of preexisting data collected during the assessment of musculoskeletal disorder risk factors.

Cervical movement measurements using two-dimensional video analysis

We tracked the markers to estimate cervical movements during protraction and retraction. This tracking was accomplished using Python (Version 3.6.15; Python Software Foundation) with the primary computer vision library OpenCV. The tracking algorithm employed was the Channel and Spatial Reliability Tracking tracker, known for its reliability and robustness in tracking.

For video recording, we used a Samsung Galaxy S20 smartphone (Samsung Inc., Seoul, Korea) equipped with a 4K video recording application, capturing video at a resolution of 3,840×2,160 pixels at 60 frames per second. The smartphone was securely mounted on a tripod, positioned 100 cm from the side of the chair, and adjusted to the height corresponding to the level of the subject's tragus. To facilitate tracking, two spherical markers, each with a diameter of 20 mm, were affixed to the tragus of the ear and spinous
process of C7.

The head tilt angle (HTA) was calculated as the angle formed between the line connecting the lateral canthus of the eye, a marker on the tragus of the ear, and a vertical line passing through the midpoint of the tragus marker (canthus-tragus-horizontal) (Figure 1).\textsuperscript{11,18} The CVA was measured between the line connecting the two markers on the tragus of the ear and the spinous process of C7 and a horizontal line passing through the marker on the spinous process of C7 (Figure 1).\textsuperscript{11,18} The head excursion angle (HEA) was determined by combining HTA and CVA measurements. HEA was defined as the angle between the line connecting the two markers on the tragus of the ear and the spinous process of C7, and the line connecting the lateral canthus of the eye to the marker on the tragus of the ear (Figure 1).\textsuperscript{19} Protraction and retraction distances were assessed by measuring the displacement of the marker on the tragus between the initial resting position and the endpoints of cervical protraction and retraction (Figure 1).

For cervical movements during retraction, the participants were instructed to retract their heads as much as possible and then return to the starting position. Pro- and retraction were performed consecutively, and each procedure was repeated three times.

**Data analysis of cervical movements during protraction and retraction**

The head tilt, craniovertebral, and head excursion angles during protraction and retraction were measured while the subjects were in a seated position. For cervical movement analysis during protraction and retraction, measurements included HTA, CVA, HEA, and protraction and retraction distances at the endpoints of cervical protraction and retraction. The endpoints of protraction and retraction were defined as the positions where the most significant movement occurred in a positive or negative direction from the typical seated posture along the horizontal axis. We subtracted the HTA, CVA, and HEA at the endpoints of protraction and retraction from the HTA, CVA, and HEA in the resting position to quantify the movement of the upper and lower cervical spine.

\[
\text{HTA during protraction (PHTA) = HTA in end point of protraction – HTA in resting}
\]

\[
\text{CVA during protraction (PCVA) = CVA in resting – CVA in end point of protraction}
\]

\[
\text{HEA during protraction (PHEA) = HEA in end point of protraction – HEA in resting}
\]

\[
\text{HTA during retraction (RHTA) = HTA in end point of retraction – HTA in resting}
\]

\[
\text{CVA during retraction (RCVA) = CVA in resting – CVA in end point of retraction}
\]

\[
\text{HEA during retraction (RHEA) = HEA in end point of retraction – HEA in resting}
\]

As PHTA becomes less pronounced (more negative), the upper cervical extension increases during protraction. With increasing PCVA (greater positive values), the lower cervical flexion also increased during protraction. As the PHEA gains strength (more positive values), cervical protraction becomes more pronounced during this movement. With a

![Figure 1. Cervical kinematic measurements using two-dimensional video analysis during cervical protraction and retraction.](image-url)
growing RHTA (increasing positive values), upper cervical flexion intensifies during retraction. Conversely, a decrease in RCVA (more negative values) resulted in a greater lower cervical extension during retraction. Decreasing the RHEA (more negative values) leads to increased cervical retraction during this phase.

ML modeling

We conducted ML analysis using Orange data mining software (Version 3.3.0, developed in Ljubljana, Slovenia) in conjunction with Python (Version 3.6.15, developed by the Python Software Foundation).

1) Pre-processing and missing data handling

Eight numerical predictors (PHTA, PCVA, PHEA, RHTA, RCVA, RHEA, protraction, and retraction distances) were included in the present study. The target was transformed into a dichotomous variable with and without NSCNP. Exploratory data analysis was performed to detect missing data. Imputation for handling missing data was performed by eliminating instances with unknown values. We assessed the distribution of each variable using various visualization methods, including boxplots, scatterplots, and linear projections. This comprehensive examination aimed to identify and eliminate outliers using a local outlier factor (with parameters set at a contamination rate of 10%, 20 neighbors, and the Euclidean metric) because of its potential impact on the accuracy of the learning model.

2) ML algorithm

From the complete dataset consisting of 740 cases, we divided it into two subsets: a training set (comprising 80% of the data, with a total of 592 samples, where individuals with NSCNP (POWs) numbered 373 and those without NSCNP numbered 219) to construct our predictive models, and a test set (representing 20% of the data, totaling 148 samples, with 90 POWs having NSCNP and 58 POWs without NSCNP) for external validation to evaluate model performance. We employed six distinct ML algorithms: neural network, random forest, logistic regression, gradient boosting, decision tree, and support vector machine. These algorithms were trained using a 10-fold cross-validation approach in the training set.

3) Model validation

The primary measure of model performance focused on the calculation of the area under the curve (AUC), which was determined for both the training and test datasets, with emphasis on the target class being the average across all classes. Additionally, secondary indicators of model performance encompassed classification accuracy, recall, precision, and F1 score (which harmoniously combines recall and precision) for both the training and test data, with the same target class being the average across all classes. We categorized the predictive model's performance as excellent (AUC≥0.9), good (AUC between 0.8 and 0.9), fair (AUC between 0.7 and 0.8), or poor (AUC<0.7) based on the AUC value.

We calculated the importance of feature permutation using the training data to determine the significance of each predictive variable. This analysis involved evaluating the contribution of each feature to the model's performance by measuring its impact on the AUC, and consequently, any increase in the model's prediction error.

RESULTS

POW characteristics

A total of 740 POWs (99 men and 641 women) were included in the ML analysis, with a NSCNP proportion of 62.6% (n=463, men=37, women=426). Means and standard deviations of NRS was 5.9±2.8 and 0.6±1.1 in POWs with and without NSCNP, respectively. The means and standard deviations of all variables are provided in Table 1. Figure 2 shows the distribution of RCVA, PCVA and PHEA data between POWs with and without NSCNP.

Predictive models of ML

The performance of the six ML models for predicting the NSCNP during model training and testing is shown in Table 2, and the most important predictors of the best performance (random forest model) are described in Figure 3.

Six ML models in the training dataset classified POWs with and without NSCNP, performed in the order of high AUC, random forest (AUC,0.835 [good]; F1, 0.767; accuracy, 0.774), neural network (AUC, 0.711 [fair]; F1, 0.682; accuracy, 0.690), decision tree (AUC, 0.691 [poor]; F1, 0.695; accuracy, 0.693), gradient boosting (AUC, 0.676 [poor]; F1, 0.659; accuracy, 0.657), logistic regression (AUC, 0.590 [poor]; F1, 0.559; accuracy, 0.633), and support vector machine (AUC, 0.582 [poor]; F1, 0.589; accuracy, 0.593) (Table 2 and Figure 4). In the test dataset classification of POWs with and without NSCNP, six ML prediction models were performed in the following order: high AUC, random forest (AUC, 0.800 [good]; F1, 0.744; accuracy, 0.750), decision tree (AUC, 0.790 [fair]; F1, 0.772; accuracy, 0.770), gradient boosting (AUC, 0.701 [fair]; F1, 0.659; accuracy, 0.655), neural network (AUC,
DISCUSSION

This study represents a significant contribution to our understanding of NSCNP by investigating the relationship between cervical movements during protraction and retraction and the presence of NSCNP as well as by using ML models to classify individuals with and without NSCNP based on these kinematic variables. The findings of this study shed light on the intricate relationship between cervical movements and NSCNP. Traditionally, neck pain research has focused on association between neck posture and NSCNP or the linear relationship between neck pain and movement patterns and has often yielded inconsistent results. In this study, the authors recognized the potential for nonlinear relationships and employed ML techniques to capture these complex associations. The results indicate that the relationship between cervical movements and NSCNP is not straightforward, as evidenced by the variable performance of the ML models.

Six ML algorithms, including random forest, neural network, decision tree, gradient boosting, logistic regression, and support vector machine, were used to classify individuals with and without NSCNP. Random forest emerged as the top-performing model, demonstrating good accuracy and an AUC of 0.800 in the test dataset. Although direct comparison is difficult and related studies are lacking, in a study that classified individuals with and without neck pain using upper cervical rotation movement control test data, the AUC was 0.61, which was relatively lower than our study.20 The performance of this model indicates its potential utility in clinical settings for classifying individuals with NSCNP.

Table 1. Mean±standard deviation of baseline characteristics in POWs with and without NSCNP

<table>
<thead>
<tr>
<th>Variables</th>
<th>Without NSCNP (N=277)</th>
<th>With NSCNP (N=463)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>62/215</td>
<td>37/426</td>
<td>-</td>
</tr>
<tr>
<td>NRS ²</td>
<td>0.64±1.1</td>
<td>5.93±2.8</td>
<td>0.000</td>
</tr>
<tr>
<td>NPQ ³-pain intensity</td>
<td>0.85±0.3</td>
<td>2.61±0.6</td>
<td>0.000</td>
</tr>
<tr>
<td>PHTA ⁴</td>
<td>0.76±7.49</td>
<td>0.80±8.45</td>
<td>0.935</td>
</tr>
<tr>
<td>PCVA ⁵</td>
<td>18.71±5.06</td>
<td>17.13±5.85</td>
<td>0.000</td>
</tr>
<tr>
<td>PHEA ⁶</td>
<td>18.25±7.33</td>
<td>16.33±7.74</td>
<td>0.001</td>
</tr>
<tr>
<td>RHTA ⁷</td>
<td>0.57±7.21</td>
<td>1.16±6.66</td>
<td>0.313</td>
</tr>
<tr>
<td>RCVA ⁸</td>
<td>−6.79±4.08</td>
<td>−6.44±4.78</td>
<td>0.293</td>
</tr>
<tr>
<td>RHEA ⁹</td>
<td>−7.30±5.57</td>
<td>−7.55±5.80</td>
<td>0.556</td>
</tr>
<tr>
<td>Protrac distance</td>
<td>59.00±18.17</td>
<td>56.51±19.97</td>
<td>0.083</td>
</tr>
<tr>
<td>Retraction distance</td>
<td>21.11±11.45</td>
<td>20.72±12.03</td>
<td>0.654</td>
</tr>
</tbody>
</table>

³NSCNP, non-specific chronic neck pain; ²NRS, numerical rating scale; ³NPQ, Northwick Park neck pain questionnaire; ⁴PHTA, head tilt angle during protraction; ⁵PCVA, craniocervical angle during protraction; ⁶PHEA, head excursion angle during protraction; ⁷RHTA, head tilt angle during retraction; ⁸RCVA, craniocervical angle during retraction; ⁹RHEA, head excursion angle during retraction.

0.649 [poor]; F1, 0.632; accuracy, 0.655), logistic regression (AUC, 0.548 [poor]; F1, 0.494; accuracy, 0.595), and support vector machine (AUC, 0.542 [poor]; F1, 0.506; accuracy, 0.568) (Table 2 and Figure 4).

For feature permutation importance, in the order of high impact AUC, RCVA, PCVA, PHEA, protraction distance, RHTA, retraction distance, RHEA, and PHTA (Figure 3).
NSCNP based on cervical movements during protraction and retraction. Although the decision tree model demonstrated fair performance, it also showed promise in classifying NSCNP cases, achieving an AUC of 0.790 in the test dataset. Gradient boosting, neural networks, and logistic regression, while showing lower AUC values, still had some predictive capabilities. However, the support vector machine model exhibited the lowest performance among ML algorithms.

In terms of feature permutation importance, the random forest model highlighted RCV A, PCVA, and PHEA as significant predictors of NSCNP. Nevertheless, except for PCVA and PHEA, there were no statistically discernible distinctions between individuals with and without NSCNP concerning the remaining influential predictors. Specifically, the disparity in PCVA and PHEA between individuals with and without NSCNP was only 1.58° and 1.92°, respectively. Previous studies have reported that women with neck pain exhibit a horizontal axis displacement of 10 cm for protraction and retraction distances, which is significantly smaller than that observed in asymptomatic women (6.8 cm). Another study also reported that the subclinical neck pain group displayed a greater displacement in the horizontal axis for retraction distance compared to the normal group. However, decreased cervical retraction might be attributed to factors such as sternocleidomastoid and posterior neck muscle tightness, limited cervical spine gliding, and weakened deep neck flexors responsible for cervical

### Table 2. Performance metrics of six machine learning algorithms in the training and test set

<table>
<thead>
<tr>
<th>Model</th>
<th>Performance metrics of six machine learning algorithms in the training set</th>
<th>Performance metrics of five machine learning algorithms in the test set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.691</td>
<td>0.693</td>
</tr>
<tr>
<td>Gradient boosting</td>
<td>0.676</td>
<td>0.657</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0.590</td>
<td>0.633</td>
</tr>
<tr>
<td>Neural network</td>
<td>0.711</td>
<td>0.699</td>
</tr>
<tr>
<td>Random forest</td>
<td>0.835</td>
<td>0.774</td>
</tr>
<tr>
<td>Support vector machine</td>
<td>0.582</td>
<td>0.593</td>
</tr>
</tbody>
</table>

![Figure 3. Feature permutation importance of random forest model in the training set.](image-url)
stabilization. Consequently, numerous physical therapists and various clinicians have endeavored to evaluate cervical retraction or craniocervical flexion movements, along with upper cervical motor control, in individuals with NSCNP.

Similar to how clinicians assess these movements, the ML model also underscored the importance of RCV A in relation to cervical retraction as a significant predictor of NSCNP in the random forest model. Intriguingly, predictors associated with cervical protraction have also emerged as vital contributors to ML models. Therefore, further investigation is warranted to delineate the movement patterns of cervical protraction and retraction in individuals with NSCNP and to compare these patterns in individuals with and without NSCNP. Additionally, it is imperative to demonstrate the potential improvement in neck pain through interventions aimed at restoring normal movement patterns in individuals with NSCNP. The findings of this study have several implications for clinical practice and research. First, they emphasized the need to consider nonlinear relationships when studying NSCNP and other musculoskeletal conditions. ML techniques, as demonstrated here, offer a valuable tool for capturing such complexities. Second, the identification of key cervical kinematic variables such as RCV A, PCV A, and PHEA provides clinicians with potential targets for assessment and intervention in individuals with NSCNP. By monitoring and addressing abnormalities in these kinematic parameters, healthcare providers may be able to better diagnose and treat NSCNP. Third, the strong performance of the random forest model suggests its potential as a clinical tool for classifying NSCNP cases based on cervical movement. This could streamline the diagnostic process and assist treatment planning.

It’s important to acknowledge some limitations of this study. Because we did not control compensatory upper thoracic motion, cervical kinematics during protraction and retraction movements would be overestimated. Controlling compensatory motion data could enhance machine learning-based classification. The sample consisted primarily of office workers, which may limit the generalizability of the findings to other populations. Additionally, the retrospective design of the study using existing data may introduce selection bias and confounding variables. Future research in this area could focus on prospective studies with more diverse populations and consider additional clinical variables. Further investigation into the mechanisms underlying the relationship between cervical movements and NSCNP is warranted.

CONCLUSIONS

In conclusion, this study highlights the potential of ML models to enhance our understanding of NSCNP and offers valuable insights into the complex relationship between cervical movements during protraction and retraction and the presence of NSCNP. Cervical kinematic variables such as RCV A, PCVA and PHEA can be potential targets for assessment and intervention in individuals with NSCNP. Clinicians and researchers can leverage these findings to develop more effective diagnostic and treatment strategies for individuals with NSCNP.

Key Points

**Question** Is there a relationship between cervical movements during protraction and retraction and the presence of non-specific chronic neck pain (NSCNP)? How well can machine learning (ML) models classify individuals with and without NSCNP based on cervical movements during protraction and retraction?
Findings Random forests emerged as the best-performing ML model, achieving an AUC of 0.800. Key predictors of NSCNP identified by the random forest model included craniovertebral angle during retraction and protraction, and head excursion angle during protraction.

Meaning The random forest model, with its strong performance, may serve as a clinical tool to classify NSCNP cases, potentially aiding in treatment planning. Key cervical kinematic variables such as craniovertebral angle during retraction and protraction, and head excursion angle during protraction can be potential targets for assessment and intervention in individuals with NSCNP.

Article information
Conflict of Interest Disclosures: The authors declare that they have no competing financial interests or personal relationships that may have influenced the work reported in this study.
Funding/Support: Yonsei University Research Fund (grant numbers: 2023-51-0199) provided funding for this study.
Acknowledgment: We would like to thank LAIBI laboratory (SJ Yang, JH Han and YS Chu), which developed graphical user interface (GUI) of tracking system using Channel and Spatial Reliability Tracking tracker.
Ethic Approval: In the present study, Yonsei University Mirae Campus Institutional Review Board waived the requirement for explicit informed consent before the queries and analyses (1041849-202306-BM-100-01). This exemption was granted because of the study’s reliance on preexisting data acquired from investigating the risk factors of musculoskeletal disorders.

Author contributions
Conceptualization: UJ Hwang.
Data acquisition: UJ Hwang, JH Kim.
Design of the work: UJ Hwang.
Data analysis: UJ Hwang, JH Kim.
Project administration: UJ Hwang.
Interpretation of data: JH Kim.
Writing – original draft: UJ Hwang.
Funding acquisition: UJ Hwang.
Writing–review&editing: UJ Hwang, JH Kim

REFERENCES
2. Comerford M, Mottram S. Kinetic control. The Man-
15. Farooq MN, Naz S, Kousar A, Shahzad K. Cross-cultural adaptation and validation of the Northwick park neck pain questionnaire to Urdu language. BMC Mus-


