Turk Neurosurg, 2022

Original Investigation

DOI: 10.5137/1019-5149.JTN.36068-21.3

Received: 26.07.2021
Accepted: 10.11.2021

Published Online: 14.03.2022

Comparative Analysis Between Machine Learning Algorithms and Conventional Regression in Predicting the Prognosis of Patients with Basilar Invagination: A Retrospective Cohort Study

Liwei PENG[1*], Chi PENG[2*], Fan YANG[3*], Wei ZUO[1], Chao CHENG[1], Peng WANG[1], Jin'an ZHANG[1], Weixin LI[1]

[1]Fourth Military Medical University, Tangdu Hospital, Department of Neurosurgery, Xi'an, China
[2]Naval Medical University, Department of Health Statistics, Shanghai, China
[3]Fourth Military Medical University, Tangdu Hospital, Department of Plastic Surgery and Burns, Xi'an, China

*Liwei Peng, Chi Peng, and Fan Yang contributed equally to this work.

Corresponding author: Weixin LI ✉ tangdunaowai@163.com

ABSTRACT

AIM: To identify predictors of basilar invagination (BI) prognosis and compare diagnostic properties between logistic modeling and machine learning methods.

MATERIAL and METHODS: We conducted a single-center retrospective study. Patients at our hospital who met the inclusion and exclusion criteria were identified between August 2015 and August 2020 for inclusion. Candidate predictors, such as demographics, clinical scores, radiographic parameters, and outcome, were included. The primary outcome was the prognosis evaluated by the change in patient-reported Japanese orthopaedic association (PRO-JOA) score. Conventional logistic regression models and machine learning algorithms were implemented. Models were compared, considering the area under the curve (AUC), sensitivity, specificity, positive and negative predictive values, and calibration curve.

RESULTS: Overall, the machine learning algorithms and traditional logistic regression models performed similarly. The postoperative cervicomedullary angle, head–neck flexion angle (HNFA), atlantodental interval, postoperative clivo-axial angle, age, postoperative clivus slope, postoperative cranial incidence, weight, postoperative HNFA, and postoperative Boogaard's angle (BoA) were identified as important predictors for BI prognosis. Among the surveyed radiographic parameters, postoperative BoA was the most important predictor of BI prognosis. In the validation dataset, the bagged trees model performed best (AUC, 0.90).

CONCLUSION: Through machine learning, we have demonstrated predictors of BI prognosis. Machine learning methods did not provide too many advantages over logistic regression in predicting BI prognosis but remain promising.

KEYWORDS: Machine learning; basilar invagination, craniocervical junction, sagittal parameter, patient-reported Japanese Orthopaedic Association score

ABBREVIATIONS: BI: Basilar invagination, **AAD:** Atlantoaxial dislocation, **CM:** Chiari malformation, **SVA:** Sagittal vertical axis, **AT:** Axial tilt, **CC:** Craniocervical tilt, **CXA:** Clivoaxial angle, **CMA:** Cervicomedullary angle, **BA:** Basal angle, **BoA:** Boogaard's angle, **HNFA:** Head-neck flexion angle, **OS:** Occipital slope; **CCA:** Craniocervical angle, **SCA:** Spino-cranial angle, **CI**: Cranial incidence, **ADI:** Atlanto-dental interval, **CLV:** Chamberlain's line violation, **CS:** Clivus slope, **LR:** Logistic regression, **BT:** Bagged trees, **ANN:** Artificial Neural Network, **RF:** Random forest, **NB:** Naive bayes, **AUC:** Area under the Curve, **PPV:** Positive predictive value, **NPV:** Negative predictive value.

Liwei PENG : 0000-0003-0273-7762
Chi PENG : 0000-0002-2527-8397
Fan YANG : 0000-0002-9595-4290
Wei ZUO : 0000-0003-0420-1793
Chao CHENG : 0000-0003-2930-5900
Peng WANG : 0000-0002-7625-8126
Jin'an ZHANG : 0000-0003-2168-618X
Weixin LI : 0000-0002-7399-8989

INTRODUCTION

Basilar invagination (BI), an abnormality that is either congenital or degenerative, is mainly characterized by a higher level of the odontoid process than normal and even protrusion into the foramen magnum, leading to chronic headaches, limited neck motion, and acute neurologic deterioration (3). BI is commonly seen in concert with a host of other congenital conditions, such as platybasia, atlantoaxial dislocation (AAD), Chiari malformation (CM), atlas occipitalization, and Klippel–Feil syndrome (28).

Exploring predictors and developing predictive models derived from clinical data can contribute to the early diagnosis and treatment of patients with adverse outcomes. However, there is a paucity of literature using variables based on radiographic parameters to establish a conventional regression model, and none of them have compared the performance of machine learning methods with that of logistic regressions (LRs). As a data-analysis technique, machine learning in essence develops models that "learn" from existing data, which can be applied to new datasets. In general, machine learning methods include a gradient-boosting machine, naive Bayes (NB), random forest (RF), bagged trees (BT), decision forest, and artificial neural networks (ANNs). In contrast with conventional logistic modeling, machine learning algorithms are not subject to distribution assumptions; therefore, they can deal with intricate or non-linear relations between predictive variables and the outcomes (21). Also, machine learning could capture complex interactions and may be effective in revealing strongly predictive factors among multiple predictive variables (8,10).

Owing to the clinical significance of identifying patients with a poor prognosis and the paucity of predictive models in BI prognosis, in the present study, we sought to use conventional regression models and a host of machine learning methods to identify predictors of the prognosis of patients with BI following reduction, decompression, and fusion and to compare their performance (9,14).

MATERIAL and METHODS

Study Design and Setting

A single-center, retrospective investigation was conducted. Eligible subjects were patients who were diagnosed with craniovertebral junction malformation at Tangdu Hospital of the Fourth Military Medical University from August 2015 to August 2020. The following were excluded: 1) patients with only AAD; 2) patients only with CM; 3) patients with os odontoideum; 4) patients without internal fixation; and 5) patients who underwent revision surgery. Because detailed data were collected from the hospital's electronic medical records system, the need for written informed consent was waived. This study was approved by the ethics committee of Tangdu Hospital (no. K202011-04).

Predictors

Predictive variable, including basic demographics, clinical symptoms, and radiographic parameters, were obtained. Specifically speaking, basic demographics included age, body mass index, and sex, while clinical symptoms incorporated posterior fossa cranial nerves symptoms, numbness, and dizziness. Separately, the radiographic parameters of interest were as follows (Figure 1A-D):

(1) C2–7 sagittal vertical axis, which is the distance between the plumb line passing through the center of gravity of C2 and the plumb line passing through the posterior edge of the superior endplate of C7;

(2) C0–1 angle, which is the angle between the McRae line and the line joining the anterior tubercle and posterior tubercle of C1 (32);

(3) C1–2 angle, which is the angle between the line joining anterior tubercle and posterior tubercle of C1 and the tangent line of the inferior endplate of C2;

(4) C0–2 angle, which is the angle between the McRae line and the tangent line of the inferior endplate of C2;

(5) C2–7 angle, which is the angle between the tangent line of the inferior endplate of C2 and the tangent line of the inferior endplate of C7;

(6) C2 slope, which is the angle between the tangent line of the inferior endplate of C2 and the horizontal line (26);

(7) C7 slope, which is the angle between the tangent line of the inferior endplate of C7 and the horizontal line (13);

(8) Cervical tilt, which is the angle between the line joining the midpoint of the superior endplate of T1 and the tip of odontoid and the vertical line to the tangent line of the superior endplate on T1;

(9) Cranial tilt, which is the line joining the midpoint of the superior endplate of T1 and the tip of odontoid and the vertical line passing through the tip of odontoid;

(10) Axial tilt, which is the angle between Chamberlain's line and the tangent line of the dorsal edge of the odontoid;

(11) Craniocervical tilt, which is the angle between the tangent line of anterior margin of the clivus and the tangent line of the ventral margin of the odontoid (4);

(12) Clivo-axial angle (CXA), which is the angle between the Wackenheim line and the tangent line of the dorsal edge of the odontoid;

(13) Cervicomedullary angle (CMA), which is the angle between the tangent line of ventral edge of the cervical spinal cord and the tangent line of the ventral edge of the medulla oblongata on midsagittal magnetic resonance imaging;

(14) Basal angle, which is the angle between the line joining the nasion and posterosuperior tip of the dorsum and the Wackenheim line;

(15) Boogaard's angle (BoA), which is the angle between the Wackenheim line and McRae line;

(16) Head–neck flexion angle (HNFA), which is the angle between the line joining the nasion and posterosuperior tip of the dorsum and the tangent line of the dorsal edge of the odontoid (2);

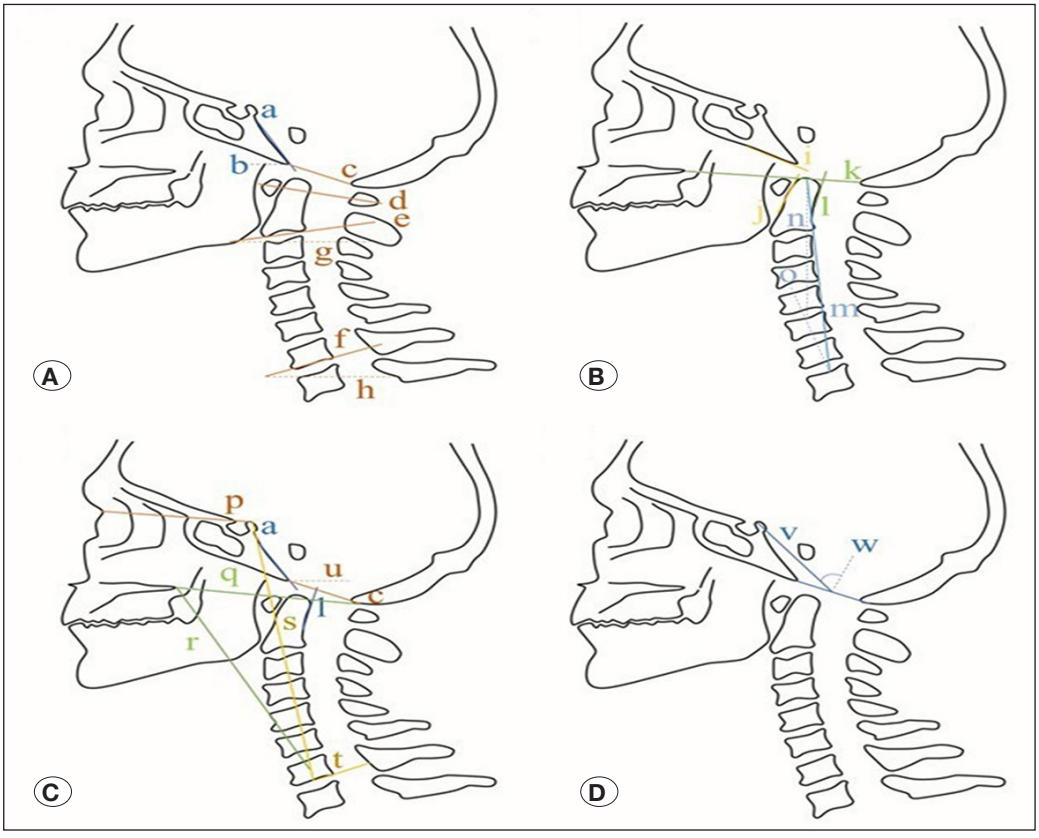

Figure 1: Measurement of craniovertebral and cervical sagittal alignment. **A)** The clivus slope is subtended between line a and a horizontal line b. The C0–1 angle is subtended between line c and line d, and the C1–2 angle is subtended between line d and line e. The C2–7 angle is subtended between line e and line f, and the C2 slope is subtended between line e and a horizontal line g. The C7 slope is subtended between line f and a horizontal line h. **B)** The craniocervical tilt is subtended between line i and line j. The axial tilt is subtended between line k and line l. The cervical tilt is subtended between line m and a vertical line o, and the cranial tilt is subtended between line m and a plumb line n. **C)** The clivo-axial angle is subtended between line a and line l. The basal angle is subtended between line a and line p, and Boogaard's angle is subtended between line a and line c. The head–neck flexion angle is subtended between line p and line l. The occipital slope is subtended between line c and a horizontal line u. The craniocervical angle is subtended between line q and line r, and the spinocranial angle is subtended between line s and line t. **D)** The cranial incidence is subtended between line v and line w.

(17) Occipital slope, which is the angle between the McRae line and horizontal line (34);

(18) Craniocervical angle, which is the angle between the line joining the posterior point of the hard palate and the center of the C7 vertebral body and McGregor line;

(19) Spinocranial angle, which is the angle between the line joining the midpoint of the posterosuperior tip of the dorsum and the middle point of the inferior endplate of C7 and the tangent line of the inferior endplate of C7;

(20) Cranial incidence (CI), which is the angle between the line joining the posterosuperior tip of the dorsum and the midpoint of the McRae line and the vertical bisection line of the McRae line (18);

(21) Atlantodental interval (ADI), which is the distance from the dorsal edge of the anterior arch of C1 to the ventral edge of the odontoid;

(22) Chamberlain's line violation, which is distance from the tip of the odontoid process to Chamberlain's line; and

(23) Clivus slope (CS), which is the angle between the Wackenheim line and the horizontal line (25).

All of the above parameters were measured on craniocervical radiographs with the patient in an upright position preoperatively and postoperatively by Surgimap (version 2.3.2.1; Nemaris Inc., New York, NY, USA). To reduce the possibility of measuring errors, every figure was recorded by an average value after two passes of quantifications (measured by L. P. and C. C.). Then, the measured value was assessed by another researcher who did not participate in the surgery with an aim to reduce the degree of potential subjective bias (P. W.). In total, 63 variables were assessed.

Outcomes

The primary outcome was the patient-reported Japanese Orthopaedic Association (PRO-JOA) score. The JOA score, originally introduced by the JOA in 1975, is widely used to offer a quantitative measurement for patients diagnosed with cervical compression myelopathy (12). Moreover, the PRO score is based on patients' self-reporting and thus is not

disturbed by the doctors' assessment, which may add to the objectiveness of the scores (24). In this passage, the outcome index was analyzed as a binary datapoint and categorized as "recovery rate over 60%" or not.

Statistical Analysis

First, continuous variables were expressed as mean ± standard deviation (normal distribution) or median (quartile) (skewed distribution) values, and categorical variables were expressed as percentages (%). When developing the predictive model, multiple methods, including conventional LR and four machine learning algorithms, were employed. It is notable that these machine learning methods were chosen because they have been successfully utilized before in clinical research (6,29).

Initially, the study population was extracted from patients who met the inclusion criteria. Then, the data were randomly assigned to a training (75%) and a testing (25%) dataset, respectively. More precisely, the study population was split into 71 patients in the development cohort and 23 patients in the validation cohort.

In this study, as a feature selection function, random forest was employed to rank all of the variables. Then, according to the feature-selection method, the rank was obtained to organize the features (10/20/30/40/50/ALL, where ALL is 63 variables, as listed in Table I). To find the optimal hyperparameters and ensure data robustness, we conducted k-fold cross-validation in the training cohort. To begin with, the data were divided into k subsets. In each iteration, each subset was used for validation, and k-1 was used for training. This process was repeated k times, and all data were used exactly k-times for training and once for testing. Finally, the mean k-time validation result was selected as a final estimate value. In this study, we used five-fold cross-validation. In this way, in the original training dataset, each sample would be involved in the training model and also participate in the testing model, so that all data were used to the greatest extent. With the ability to rank candidate variables according to their importance, this variable-selection method was suitable for high-dimensional data with multicollinearity (11).

Table I: The Basic Clinical Characteristics of Patients with BI

	Overall	Development Cohort	Validation Cohort	p	recovery rate > 60%	recovery rate < 60%	p
	n=94	n=71	n=23		n=68	n=26	
Age, years	39.0 (29.0;49.0)	39.0 (28.0;49.0)	39.0 (31.0;47.0)	0.728	33.0 (28.0;46.0)	48.0 (32.2;56.5)	0.003
Male (%)	39 (41.5%)	37 (49%)	57 (51%)	0.092	27 (39.7%)	12 (46.2%)	0.570
BMI (kg/m^2)	22.0 (20.0;27.0)	23.4 (20.0;27.0)	21.3 (18.0;25.2)	0.125	21.8 (20.0;26.2)	24.1 (20.0;27.0)	0.392
Posterior fossa cranial nerve symptoms	7 (7.45%)	8 (28%)	4 (0.21%)	0.846	6 (8.82%)	1 (3.85%)	0.702
NRS	2.00 (0.00;5.00)	2.00 (0.00;6.00)	2.00 (0.00;5.00)	0.849	2.00 (0.00;5.00)	0.00 (0.00;5.75)	0.089
NDI	0.23 (0.20;0.37)	0.23 (0.20;0.37)	0.23 (0.20;0.36)	0.813	0.20 (0.20;0.33)	0.37 (0.20;0.68)	0.040
Numbness	72 (76.6%)	73 (45%)	87 (34%)	0.177	50 (73.5%)	22 (84.6%)	0.256
Dizziness	41 (43.6%)	45 (50%)	39 (50%)	0.618	31 (45.6%)	10 (38.5%)	0.533
AAD	31 (33.0%)	32 (47%)	35 (49%)	0.832	19 (27.9%)	12 (46.2%)	0.093
Klippel-Feil syndrome	26 (27.7%)	27 (45%)	30 (47%)	0.732	20 (29.4%)	6 (23.1%)	0.539
CM	38 (40.4%)	39 (49%)	43 (51%)	0.731	26 (38.2%)	12 (46.2%)	0.484
Platybasia	27 (28.7%)	31 (47%)	22 (42%)	0.394	19 (27.9%)	8 (30.8%)	0.786
PFD	15 (16.0%)	20 (40%)	4 (21%)	0.080	10 (14.7%)	5 (19.2%)	0.825
Atlantoaxial fusion	7 (7.45%)	8 (28%)	4 (21%)	0.846	5 (7.35%)	2 (7.69%)	1.000
CLV (mm)	11.2 (8.62;13.8)	11.2 (8.60;13.4)	11.3 (8.80;15.4)	0.559	11.2 (8.95;13.7)	11.4 (7.78;14.3)	0.560
ADI (mm)	3.80 (2.60;5.50)	3.80 (2.65;5.35)	3.70 (2.35;5.90)	0.996	3.90 (2.80;5.82)	3.00 (2.50;5.12)	0.130
C2-7 SVA (mm)	12.1 (5.80;16.0)	11.9 (5.20;15.4)	12.9 (7.80;24.4)	0.159	12.1 (6.47;16.4)	12.2 (5.40;15.6)	0.806
C0-1 angle (°)	9.60 (5.03;14.9)	10.2 (6.25;15.4)	6.80 (4.30;13.9)	0.158	9.15 (5.33;13.6)	10.7 (5.00;15.4)	0.462
C1-2 angle (°)	23.9 (15.5;30.7)	23.2 (15.6;28.5)	27.3 (15.7;32.0)	0.518	22.9 (15.8;30.2)	24.9 (11.8;32.6)	0.823
C0-2 angle (°)	31.7 (22.9;37.8)	31.7 (23.0;38.3)	32.7 (23.0;37.5)	0.742	31.6 (23.1;37.4)	32.9 (22.9;39.4)	0.790
C2-7 angle (°)	20.1 (11.1;28.8)	18.8 (9.30;28.6)	23.7 (13.7;32.2)	0.130	18.3 (10.5;28.2)	21.4 (12.8;32.0)	0.277
C2 slope (°)	7.85 (3.52;14.0)	7.90 (3.65;14.0)	5.00 (3.40;13.9)	0.586	7.35 (3.58;13.3)	9.25 (3.33;17.2)	0.660
C7 slope (°)	21.5 (15.3;26.6)	21.2 (13.9;25.5)	26.1 (18.2;30.7)	0.018	21.9 (15.5;26.6)	20.9 (14.7;28.1)	0.823
Cranial tilt (°)	4.80 (2.55;8.10)	4.80 (2.50;7.60)	5.60 (3.20;10.4)	0.340	4.95 (2.65;8.10)	3.85 (2.55;8.38)	0.633

Table I: Cont.

	Overall	Development Cohort	Validation Cohort	p	recovery rate > 60%	recovery rate < 60%	p
	n=94	n=71	n=23		n=68	n=26	
Cervical tilt (°)	18.7 (13.7;25.0)	18.0 (13.1;23.6)	22.8 (15.9;27.3)	0.098	18.4 (13.8;23.3)	21.0 (14.2;31.8)	0.247
Axial tilt (°)	82.2 (74.1;89.1)	80.9 (73.9;88.9)	84.9 (74.6;89.8)	0.509	81.4 (73.8;87.9)	83.0 (78.7;89.4)	0.403
Craniocervical tilt (°)	103 (91.8;116)	104 (91.6;119)	100 (93.6;108)	0.189	102 (90.3;114)	109 (96.8;122)	0.111
CXA (°)	132 (121;144)	131 (121;144)	133 (119;144)	0.850	131 (122;144)	133 (121;143)	0.856
CS (°)	53.2 (43.1;62.4)	53.0 (42.9;62.3)	54.5 (45.3;61.4)	0.651	50.4 (42.9;61.3)	57.1 (43.7;62.5)	0.422
CMA (°)	132 (120;144)	134 (120;143)	128 (122;144)	0.651	130 (119;142)	138 (125;144)	0.370
Basal angle (°)	134 (125;146)	134 (125;146)	138 (122;146)	0.622	136 (126;146)	132 (122;146)	0.630
Boogaard's angle (°)	156 (147;164)	157 (147;164)	154 (148;162)	0.632	155 (147;164)	159 (145;163)	0.872
HNFA (°)	90.1 (77.0;101)	89.7 (79.2;101)	93.6 (73.2;104)	0.867	93.9 (79.5;102)	86.7 (74.1;95.2)	0.105
Occipital slope (°)	30.1 (23.5;34.8)	30.6 (23.1;35.2)	28.7 (24.4;33.2)	0.601	29.0 (22.7;34.0)	30.7 (24.7;37.2)	0.462
CCA (°)	55.3 (52.2;61.2)	56.3 (51.5;61.5)	55.0 (52.5;56.4)	0.455	56.0 (52.5;61.2)	54.5 (51.6;59.1)	0.366
SCA (°)	79.6 (73.0;85.1)	79.9 (74.6;86.3)	73.8 (69.2;80.9)	0.019	79.7 (73.1;86.0)	76.9 (72.0;81.4)	0.444
Cranial incidence (°)	69.9 (64.3;76.3)	71.0 (64.2;76.9)	67.7 (65.6;72.6)	0.449	70.2 (65.3;75.4)	69.0 (61.6;79.1)	0.586
pCLV (mm)	8.25 (6.00;11.7)	7.80 (5.55;10.9)	10.1 (7.10;12.2)	0.136	8.80 (6.30;11.2)	8.05 (5.70;12.6)	0.862
pADI (mm)	1.85 (1.50;2.70)	1.80 (1.50;2.55)	2.10 (1.60;3.05)	0.367	1.80 (1.50;2.70)	1.90 (1.42;3.00)	0.973
pC2-7 SVA (mm)	13.9 (7.23;20.6)	13.7 (6.55;20.0)	16.4 (10.9;21.4)	0.176	13.9 (10.1;19.0)	13.5 (4.92;22.0)	0.816
pC0-1 angle (°)	7.75 (4.23;13.5)	7.60 (4.05;13.5)	8.30 (4.75;13.7)	0.583	7.75 (4.27;14.0)	7.75 (4.35;10.7)	0.764
pC1-2 angle (°)	23.9 (15.9;31.3)	23.5 (16.6;30.5)	26.9 (14.7;35.5)	0.847	26.3 (16.4;31.6)	21.2 (14.5;28.8)	0.199
pC0-2 angle (°)	31.0 (24.1;40.9)	31.1 (24.1;37.9)	30.9 (22.4;44.0)	0.895	31.8 (24.1;41.8)	28.6 (21.5;36.6)	0.317
pC2-7 angle (°)	19.3 (10.2;25.6)	16.7 (10.0;24.5)	25.0 (19.0;31.9)	0.033	19.3 (10.2;25.5)	19.8 (8.53;25.3)	0.973
pC2 slope (°)	7.80 (3.95;13.3)	7.40 (3.90;11.6)	10.1 (4.65;15.1)	0.228	7.90 (4.25;13.9)	7.75 (3.52;12.5)	0.823
pC7 slope (°)	22.6 (16.7;28.0)	20.8 (15.2;27.4)	27.3 (21.6;28.4)	0.014	23.3 (17.8;28.1)	20.6 (14.8;27.5)	0.396
pCranial tilt (°)	4.75 (2.15;9.28)	4.20 (2.10;9.25)	5.50 (3.10;9.10)	0.457	4.50 (2.10;8.88)	6.45 (2.45;9.73)	0.467
pCervical tilt (°)	17.8 (11.4;21.9)	17.1 (10.6;21.4)	19.5 (16.4;26.1)	0.152	18.3 (10.6;22.8)	17.3 (14.2;21.0)	0.980
pAxial tilt (°)	83.2 (76.0;90.9)	81.8 (76.0;90.8)	84.2 (79.2;90.6)	0.657	82.8 (76.1;88.7)	87.2 (76.1;91.9)	0.333
pCraniocervical tilt (°)	107 (93.0;115)	106 (92.2;116)	108 (95.5;114)	0.840	106 (92.7;115)	107 (99.1;117)	0.407
pCXA (°)	130 (121;142)	131 (121;142)	130 (120;141)	0.718	128 (119;138)	139 (127;146)	0.009
pCS (°)	49.5 (43.7;63.1)	49.4 (43.7;61.9)	51.6 (43.9;63.8)	0.782	47.9 (40.3;58.8)	56.5 (49.8;65.0)	0.011
pCMA (°)	140 (130;151)	142 (132;151)	136 (125;147)	0.289	143 (131;156)	135 (124;144)	0.043
pBasal angle (°)	137 (126;148)	136 (124;148)	140 (126;150)	0.465	141 (129;150)	130 (123;148)	0.231
pBoogaard's angle (°)	158 (148;165)	157 (149;164)	161 (147;168)	0.498	160 (150;167)	152 (140;161)	0.011
pHNFA (°)	88.8 (75.7;103)	88.2 (76.7;103)	94.0 (70.4;103)	0.400	86.6 (75.2;103)	95.4 (83.2;103)	0.117
pOccipital slope (°)	28.2 (23.6;35.6)	27.3 (21.2;34.9)	30.5 (27.1;39.5)	0.184	28.1 (23.7;38.1)	29.4 (22.2;35.3)	0.899
pCCA (°)	57.2 (52.5;61.6)	57.4 (52.5;63.5)	56.9 (51.9;58.4)	0.423	57.3 (52.5;61.8)	57.0 (53.2;61.3)	0.993
pSCA (°)	79.2 (71.0;85.6)	79.5 (72.1;86.8)	76.6 (67.4;82.6)	0.088	79.2 (71.2;86.7)	78.7 (71.2;83.5)	0.742
pCranial incidence (°)	69.4 (61.9;73.0)	69.1 (60.9;73.0)	69.9 (62.5;71.9)	0.958	70.0 (63.8;73.2)	67.9 (59.7;70.8)	0.137

Values are median ± IQR or n (%) as applicable.
***BMI:** Body mass index; **NRS:** Numeric rating scales, **NDI:** Neck disability index, **AAD:** Atlantoaxial dislocation, **CM:** Chiari malformation, **PFD:** Posterior fossa decompression, **CLV:** Chamberlain's line violation, **ADI:** Atlanto-dental interval; **SVA:** Sagittal vertical axis, **CXA:** Clivo-axial angle, **CS:** Clivus slope, **CMA:** Cervico-medullary angle, **HNFA:** Head-neck flexion angle, **CCA:** Craniocervical angle, **SCA:** Spino-cranial angle, **p:** postoperative.*

Conventional LR model

LR is used to estimate the probability of binomial variables (17). As a simplified version of generalized linear modeling, LR utilizes a sigmoid function to predict the logistic transformation of the probability for each class of the dependent variable (16).

Machine learning models

BT and RF are common decision tree-based machine learning methods, which have been proven to produce reliable predictions for various datasets (7). Through a "bagging" (bootstrap aggregation) procedure, these two models are able to reduce model variance in order to increase prediction accuracy (19).

An ANN is a biologic network-based machine learning algorithm. The working mechanism of ANN simulates the way that brain neurons operate (33). The model is learned based on multilayer perceptions using a backpropagation algorithm, which relies on nodes and connections to make complex decisions (23).

Based on Bayes' theorem, NB is a high-bias, low-variance probabilistic classifier with the strong assumption that all variables or features are independent. The NB utilizes training samples and the conditional probability of each variable to estimate the parameters (5).

The area under the recover-operating characteristic curve (AUC) in the validation data was calculated to compare the performance of models generated by diverse machine learning algorithms. The closer the AUC was to 1, the better the calibration, and the calibration curve was also used to visualize the predictive power of the models. In addition, calculated by the obfuscation matrix, the sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) were obtained.

All analyses were performed using the R version 4.0.2 software program (http://www.R-project.org, The R Foundation For Statistical Computing, Vienna, Austria). In our study, we used the "Caret" R package to achieve the process. P-values of less than .05 (two-sided test) were considered to be statistically significant.

RESULTS

Baseline Characteristics

Of the 289 patients diagnosed with craniocervical junction during the study period, following application of the inclusion and exclusion criteria, 94 patients were finally included in the study dataset. The processes of data extraction, training preparation, and data testing via different machine learning algorithms are depicted in Figure 2. The demographics, clinical manifestations, and radiographic parameters are shown in Table I. Between the training set and the validation set, there were almost no differences in a variety of variables, which illustrated the fact that the baseline level was comparable, further proving that our results were robust. Patients who had a poor prognosis were more likely to be older with higher neck disability index (NDI) scores and a larger postoperative CXA, larger postoperative CS, smaller postoperative CMA, and smaller postoperative BoA. In the development cohort, 51 patients reported improved prognosis after reduction, decompression, and fusion, while, in the validation cohort, 17 patients reported these results.

Variable Selection

Based on the algorithm of backward selection and an outside sample division loop, the RF method was used to choose the best dataset of covariates associated with BI prognosis. In our study, we found that, when the observation results selected 10 independent variables, the prediction accuracy was optimized (Figure 3). The feature selection is shown in Figure 4. The 10 most important variables were postoperative CMA, HNFA, ADI, postoperative CXA, age, postoperative, CS postoperative CI, weight, postoperative HNFA, and postoperative BoA.

Model Comparisons

Figures 5 and 6 and Table II summarize the discrimination and calibration performance for ANN, NB, BT, RF, and LR, respectively. Considering the four machine learning algorithms applied to the BI patients, the model obtained by BT had the best discrimination (AUC, 0.90; sensitivity, 94.12%; specificity, 33.33%; PPV, 80.00%; and NPV, 66.67%). It is notable that the ANN, RF, and LR showed identical AUCs for predicting BI prognosis: (0.89, 0.87, and 0.89). However, the accuracy of LR (0.9130) outperformed those of ANN (0.7391) and RF (0.7826). The NB was least discriminative, with an AUC of 0.79.

Table II: Performance of Models Derived Using BT, ANN, RF, NB, and Logistic Regression, Assessed by a Separate Validation Dataset

	AUC	Sensitivity	Specificity	PPV	NPV	Accuracy
BT	0.9	0.9412	0.3333	0.8	0.6667	0.7826
ANN	0.89	1	0	0.7391	NA	0.7391
RF	0.87	0.9412	0.3333	0.8	0.6667	0.7826
NB	0.79	0.9412	0.6667	0.8889	0.8	0.8696
LR	0.89	0.8974	1	1	0.6667	0.913

BT: *Bagged Trees,* ***ANN:*** *Artificial neural network,* ***RF:*** *Random forest,* ***NB:*** *Naive bayes,* ***LR:*** *Logistic regression,* ***PPV:*** *Positive predictive value,* ***NPV:*** *Negative predictive value.*

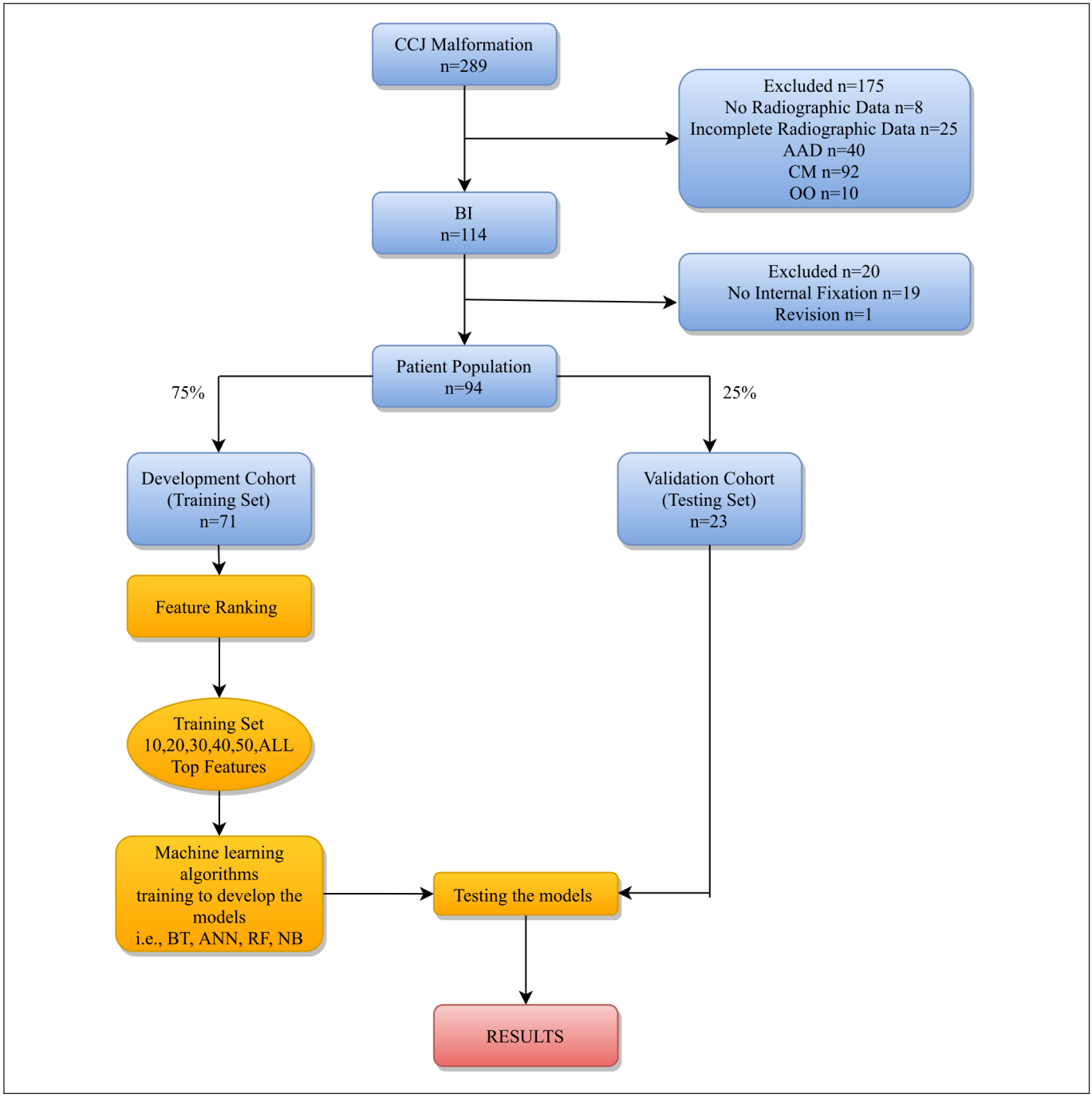

Figure 2: An overview of the methods used for data extraction, training, and testing. **CCJ:** Craniocervical junction, **AAD:** Atlantoaxial dislocation, **CM:** Chiari malformation, **OO:** Os odontoideum, **BT:** Bagged tree, **ANN:** Artificial neural network, **RF:** Random forest, **NB:** Naive bayes.

When it comes to calibration performance, by comparing the observed and model-predicted risks of BI prognosis, the calibration curves are as represented in Figure 6 (22). We also employed the Hosmer–Lemeshow goodness-of-fit test. The 80%- and 95%-confidence level calibration belts are plotted in light and dark grey, respectively. To be specific, the test calculated the chi-squared value based on the actually observed and model-predicted values for each group, and then yielded the corresponding p-value. When the 45° diagonal bisector did not cross the 95% confidence interval region, it suggested a good fit of the prediction model. A p-value of less than .05 for the belt plot of the calibration curve indicates a poor fit of the prediction model. As the plot depicted, the BT, RF, NB, and LR models had a good calibration—namely,

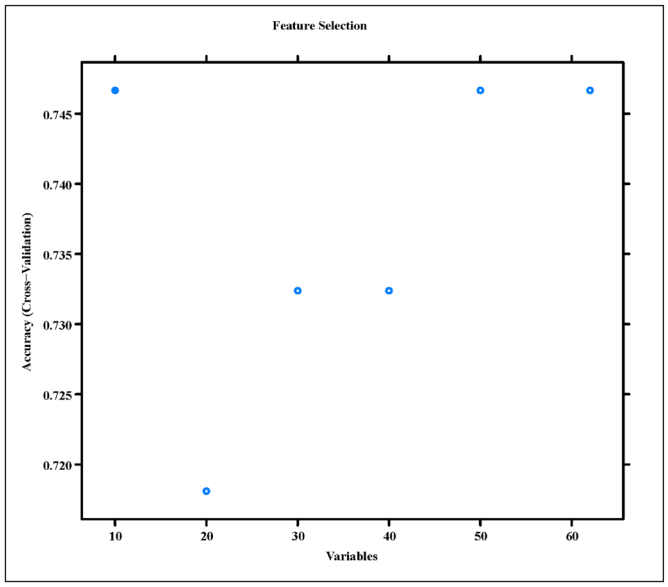

Figure 3: The association between the number of variables allowed to be considered at each split and the prediction accuracy in the random forest model.

a high accuracy of prediction, with p-values of .207, .220, .324, and .794, respectively. In contrast, ANN was the worst calibrated model, with p<.05.

In Figure 4, the top 10 predictor variables in the machine learning models are shown. Each variable had varying importance in BI prognosis relying on the machine learning approach. As a whole, the radiographic parameter with the greatest importance was postoperative BoA, followed by postoperative CS, postoperative CMA, postoperative CXA, and so forth.

DISCUSSION

In our study, we found that models based on machine learning can better predict BI prognosis compared to conventional logistic models. All models showed moderate discrimination, with AUC statistics above 0.79. The BT model, which was the best-performing machine learning model, achieved an AUC of 0.90, with a sensitivity of 94.12% and a specificity of 33.33%. The prediction models conducted by ANN, RF, and LR manifested similar discrimination capabilities, with AUCs of 0.89, 0.87, and 0.89, respectively. Further, the BT, RF, NB,

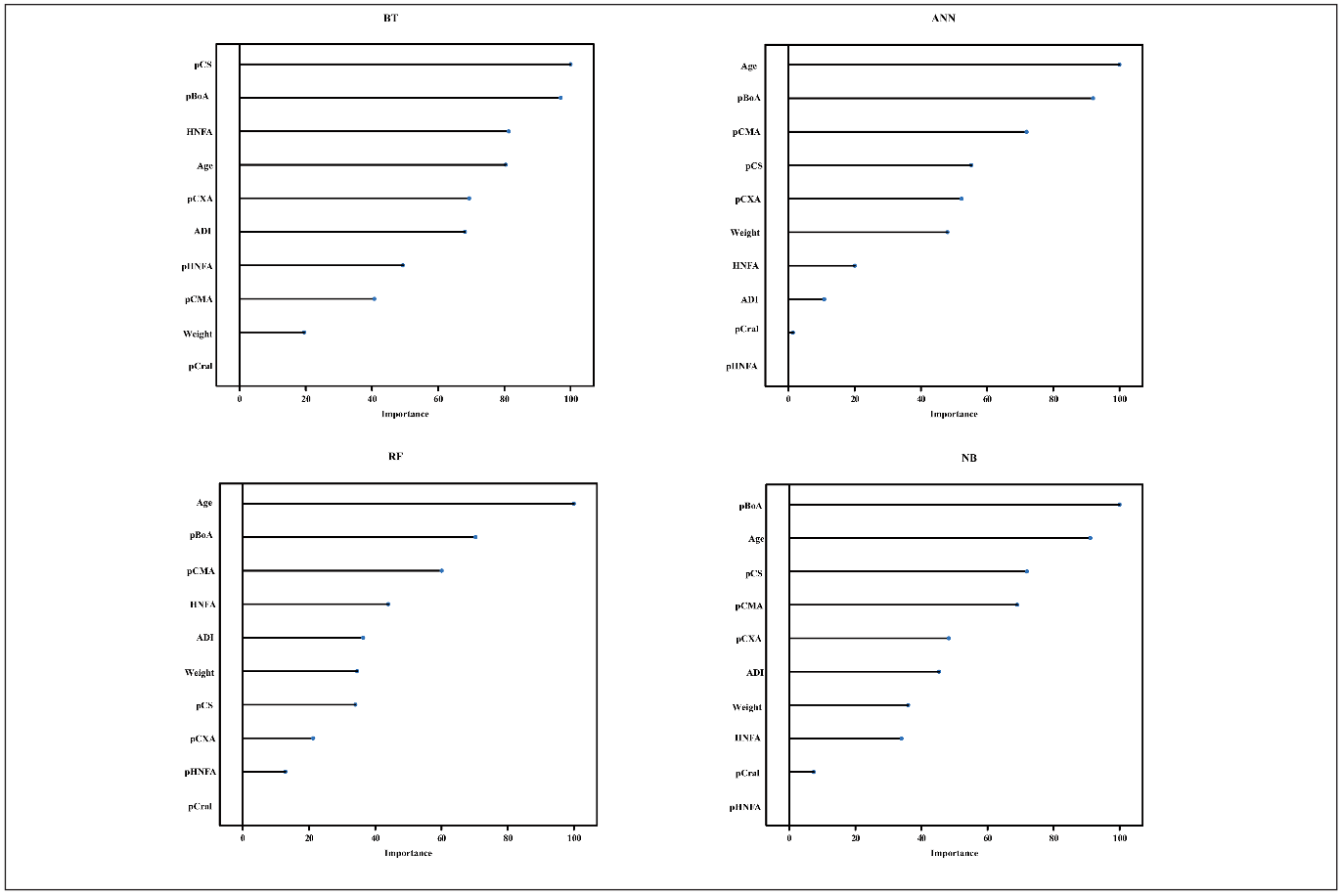

Figure 4: Variable importance in four different machine learning models. **BT:** Bagged tree, **ANN:** Artificial neural network, **RF:** Random forest, **NB:** Naive bayes, **pCS:** Postoperative clivus slope, **pCMA:** Postoperative cervico-medullary angle, **ADI:** Atlanto-dental interval, **pHNFA:** Postoperative head-neck flexion angle, **HNFA:** Head-neck flexion angle, **pCXA:** Postoperative clivo-axial angle, **pCraI:** Postoperative cranial incidence, **pBoA:** Postoperative boogaard's angle.

and LR models showed good calibration. By developing the predictive model, the current study fills in a gap in the existing literature regarding the exploration of BI prognosis through demographic characteristics and radiographic parameters. As the indications for BI surgery expand, a better prediction of its prognosis is vital; consequently, these findings add interest in the machine learning approach for doctors and researchers.

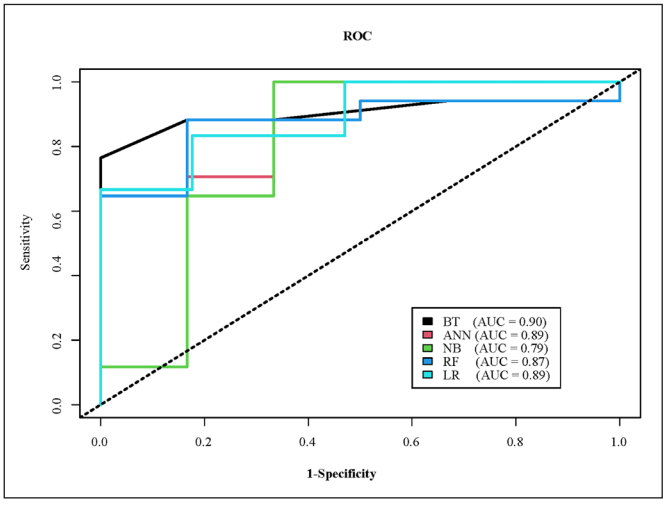

Figure 5: The area under the curve of the receiver-operating characteristic curve by machine learning models and logistic model in the validation cohort. **BT:** Bagged tree, **ANN:** Artificial neural network, **NB:** Naive bayes, **RF:** Random forest, **LR:** Logistic regression.

By the variable selection of RF, postoperative CMA, HNFA, ADI, postoperative CXA, age, postoperative CS, postoperative CI, weight, postoperative HNFA, and postoperative BoA were identified as strong predictors for BI prognosis. In accordance with the comprehensive ranking of the predictor importance by the four different machine learning algorithms, irrespective of age, postoperative BoA was the most important sagittal parameter, followed by postoperative CS, postoperative CMA, and postoperative CXA. Postoperative CXA, a representative of the occipitocervical relationship, was suggested previously by Wang et al. to be a criterion for objectively judging the degree of BI correction (30). Also, Xu and Gong argued that BI was present when the CXA was less than 140° (31), and Ma et al. proposed that the combined diagnosis of CXA and CPA would improve the accuracy of BI diagnosis (20). These findings all suggest the significance of CXA in the diagnosis of BI. Postoperative CMA is also commonly used as a criterion for whether or not the spinal cord is compressed. Junlong et al. showed that CMA could be used to assess the degree of spinal cord compression at the craniocervical junction, and a normal postoperative CMA indicated complete surgical decompression (15).

To our knowledge, this is the first prediction model for BI prognosis with clinical data generated by conventional LR and machine-learning algorithms. Machine learning algorithms

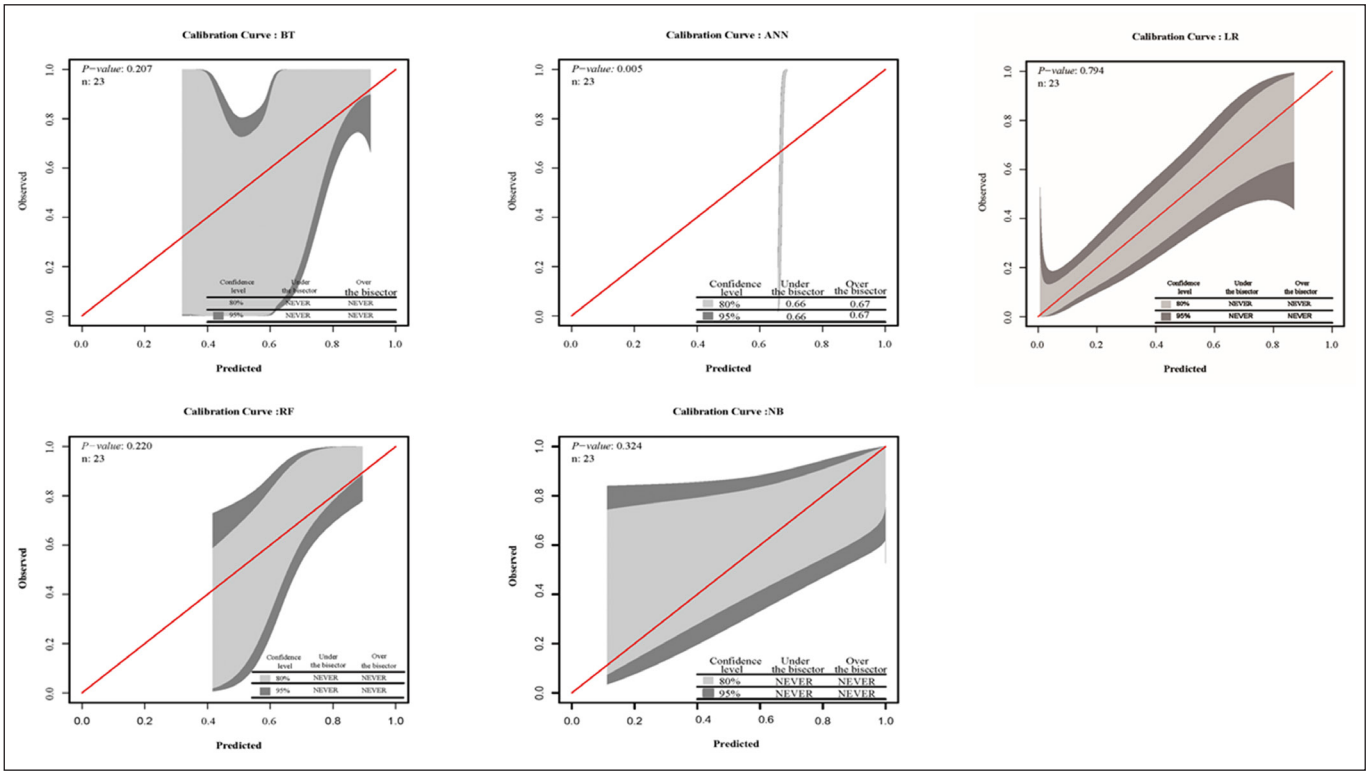

Figure 6: The calibration curve of the machine learning models and conventional logistic model was generated by showing the relationships between observed and predicted basilar invagination prognosis. **BT:** Bagged tree, **ANN:** Artificial neural network, **RF:** Random forest, **NB:** Naive bayes.

do not have a significant advantage over traditional logistic models in predicting BI prognosis. The BT model had the largest AUC and the highest sensitivity, while the specificity, PPV, and accuracy of LR were the highest, so the advantage of models based on machine learning algorithm was not obvious compared to the traditional LR model.

As a model that is easily available in all standard statistical software packages, the conventional LR models are adept at dealing with linear relationships between variables and outcomes. Nevertheless, confounding factors and collinearity will be adverse to the performance of this model, resulting in biased results. In contrast, the machine learning techniques used in the present study did not require a unique variable distribution and a priori specification; thus, they can manage the intricate relationship between the different variables. In the prediction of BI prognosis, variables such as radiographic parameters mainly act additively, with intricate nonlinearities and interactions adequately pronounced to make machine learning play a role. However, it has been generally acknowledged that the weakness of machine learning algorithms is their "black box" nature, which may reduce their interpretability and make them more difficult for clinical researchers to accept. Nevertheless, even more exciting, efforts in visualizing predictor variable importance continue (1). Although, in the current study, the predictive performances of machine learning and traditional models are identical, our research will not prevent the use of machine learning because there is no permanent "best" prediction model, and the choice of the best method relies on the clinical context. Overall, the prediction models in our study provide reference for predicting BI prognosis based on machine learning methods and conventional models. Doctors may benefit from the strategy of altering certain angles during the operation and initiating early prevention and treatment measures in patients with poor prognoses.

There were several limitations in this study. First, as a retrospective study performed at a single hospital, inevitably, this study had residual confounding, and the results cannot be easily applied to other settings. Second, owing to the complicated and heterogeneous nature of BI, variables such as race and vital signs including heart rate and blood pressure, which were not available in our study, might have a role in the outcomes of BI. Nevertheless, 63 variables were considered in this paper, and, based on that, we believe that, so far, this study presents the available best evidence for predicting the prognosis of BI. Third, it is noteworthy that our sample size was relatively small, alleviating the accuracy of model fitting. However, the receiver-operating characteristic curve of the BT model for this study still had a power of 0.90, accompanied by an accuracy of 78.26. In the future, prospective studies involving more institutions, additional variables, and larger sample sizes are needed to draw conclusions with higher levels of evidence. Fourth, owing to a lack of external verification, the accuracy of the prediction models might be affected, leading to overly optimistic estimates. Fifth, the complexity of machine learning algorithms may hinder their application in clinical practice. Nevertheless, owing to the fact that artificial intelligence algorithms are being applied in other medical fields (27), we expect that this will also be widely true in neuroscience soon.

CONCLUSION

In conclusion, the prediction of BI prognosis based on clinical manifestation and radiographic parameters is of value for patients suffering from this disease. This study identified some predictors of BI prognosis through machine learning; however, there is still room for further development and verification of machine learning algorithms in BI prognosis prediction.

AUTHORSHIP CONTRIBUTION

Study conception and design: LP, WZ, WL

Data collection: LP

Analysis and interpretation of results: LP, CP

Draft manuscript preparation: LP, FY

Critical revision of the article: CC, WL

Other (study supervision, fundings, materials, etc...): PW, JZ

All authors (LP, CP, FY, WZ, CC, PW, JZ, WL) reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Blanchet L, Vitale R, Stavropoulos G, Vorstenbosch RV, Smolinska A: Constructing bi-plots for random forest: tutorial. Analytica Chimica Acta 1131:146-155, 2020
2. Botelho RV, Ferreira EDZ: Angular craniometry in craniocervical junction malformation. Neurosurg Rev 36:603-610, 2013
3. Botelho RV, Ferreira JA, Ferreira EDZ: Basilar invagination: A craniocervical kyphosis. World Neurosurg 117:e180-e186, 2018
4. Chandra PS, Goyal N, Chauhan A, Ansari A, Sharma BS, Garg A: The severity of basilar invagination and atlantoaxial dislocation correlates with sagittal joint inclination, coronal joint inclination, and craniocervical tilt: A description of new indexes for the craniovertebral junction. Neurosurgery 10 Suppl 4:621-629; discussion 629-630, 2014
5. Dumitru D: Prediction of recurrent events in breast cancer using the Naive Bayesian classifcation. Annals of the University of Craiova 36:92-96, 2009
6. Fernandez-Delgado M, Cernadas E, Barro S, Amorim D: Do we need hundreds of classifiers to solve real world classification problems? J Machine Learning Research 15:3133-3181, 2014
7. Friedman Classification and Regression Trees. Wadsworth International Group
8. Goecks J, Jalili V, Heiser LM, Gray JW: How machine learning will transform biomedicine. Cell 181:92-101, 2020
9. Goel A: Treatment of basilar invagination by atlantoaxial joint distraction and direct lateral mass fixation. J Neurosurg Spine 1:281-286, 2004
10. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H: eDoctor: Machine learning and the future of medicine. J Intern Med 284(6):603-619, 2018

11. Hernandez-Suarez DF, Kim Y, Villablanca P, Gupta T, Wiley J, Nieves-Rodriguez BG, Rodriguez-Maldonado J, Feliu Maldonado R, da Luz Sant'Ana I, Sanina C, Cox-Alomar P, Ramakrishna H, Lopez-Candales A, O'Neill WW, Pinto DS, Latib A, Roche-Lima A: Machine learning prediction models for in-hospital mortality after transcatheter aortic valve replacement. JACC Cardiovasc Interv 12:1328-1338, 2019
12. Hirabayashi K, Miyakawa J, Satomi K, Maruyama T, Wakano K: Operative results and postoperative progression of ossification among patients with ossification of cervical posterior longitudinal ligament. Spine 6:354, 1981
13. Inoue T, Ando K, Kobayashi K, Nakashima H, Ito K, Katayama Y, Machino M, Kanbara S, Ito S, Yamaguchi H, Koshimizu H, Segi N, Kato F, Imagama S: Age-related changes in T1 and C7 slope and the correlation between them in more than 300 asymptomatic subjects. Spine (Phila Pa 1976) 46(8):E474-E481, 2021
14. Jian FZ, Chen Z, Wrede KH, Samii M, Ling F: Direct posterior reduction and fixation for the treatment of basilar invagination with atlantoaxial dislocation. Neurosurgery 66:678, 2010
15. Junlong Z, Zhimin P, Yiwei C, Haoqun Y, Zujue C, Quanfei L, Zhaoxun Z, Zhiyun L, Pingguo D, Jiangwei C: Postoperative cervical sagittal realignment improves patient-reported outcomes in chronic atlantoaxial anterior dislocation. Oper Neurosurg 15(6):643-650, 2018
16. Kleinbaum DG, Klein M: Logistic Regression (A Self-Learning Text). Springer, 2002
17. le Cessie S, van Houwelingen JC: Ridge estimators in logistic regression. J Royal Statistical Society Series C 41:191-201, 1992
18. Le Huec JC, Demezon H, Aunoble S: Sagittal parameters of global cervical balance using EOS imaging: Normative values from a prospective cohort of asymptomatic volunteers. Eur Spine J 24:63-71, 2015
19. Liaw A, Wiener M: Classification and regression by random forest. R News 2(3):18-22, 2002
20. Ma L, Guo L, Li X, Qin J, He W, Xiao X, Lu L, Xu Y, Wu Y: Clivopalate angle: A new diagnostic method for basilar invagination at magnetic resonance imaging. Eur Radiol 29(7):3450-3457, 2019
21. Miller DD, Brown EW: Artificial intelligence in medical practice: The question to the answer? Am J Med 131:129-133, 2018
22. Nattino G, Finazzi S, Bertolini G: A new calibration test and a reappraisal of the calibration belt for the assessment of prediction models based on dichotomous outcomes. Stat Med 33:2390-2407, 2014
23. Ojha VK, Abraham A, Snasel V: Metaheuristic design of feedforward neural networks: A review of two decades of research. Engineering Applications of Artificial Intelligence 60:97-116, 2017
24. Oshima Y, Takeshita K, Kato S, Doi T, Matsubayashi Y, Taniguchi Y, Nakajima K, Oguchi F, Okamoto N, Sakamoto R, Tanaka S: Comparison Between the Japanese Orthopaedic Association (JOA) Score and Patient-Reported JOA (PRO-JOA) Score to Evaluate Surgical Outcomes of Degenerative Cervical Myelopathy. Global Spine J, 2020 (Online ahead of print)
25. Peng L, Cheng C, Zuo W, Wang P, Mao Z, Li W: Letter to editor: Clivopalate angle: A new diagnostic method for basilar invagination at magnetic resonance imaging. Eur Radiol, 2021
26. Protopsaltis TS, Ramchandran S, Tishelman JC, Smith JS, Neuman BJ, Jr GMM, Lafage R, Klineberg EO, Hamilton DK, Lafage V: The importance of C2 slope, a singular marker of cervical deformity, correlates with patient-reported outcomes. Spine (Phila Pa 1976) 45(3):184-192, 2020
27. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP: Machine learning in cardiovascular medicine: Are we there yet? Heart 104(14):1156-1164, 2018
28. Smoker WRK, Khanna G: Imaging the craniocervical junction. Childs Nerv Syst 24:1123-1145, 2008
29. Tian Y, Yang J, Lan M, Zou T: Construction and analysis of a joint diagnosis model of random forest and artificial neural network for heart failure. Aging 12:26221-26235, 2020
30. Wang C, Yan M, Zhou HT, Wang SL, Dang GT: Open reduction of irreducible atlantoaxial dislocation by transoral anterior atlantoaxial release and posterior internal fixation. Spine 31:E306, 2006
31. Xu S, Gong R: Clivodens angle: A new diagnostic method for basilar invagination at computed tomography. Spine (Phila Pa 1976) 41(17):1365-1371, 2016
32. Yoshimoto H, Ito M, Abumi K, Kotani Y, Shono Y, Takada T, Minami A: A retrospective radiographic analysis of subaxial sagittal alignment after posterior C1-C2 fusion. Spine (Phila Pa 1976) 29:175-181, 2004
33. Zaharchuk G, Gong E, Wintermark M, Rubin D, Langlotz CP: Deep learning in neuroradiology. AJNR Am J Neuroradiol 39(10):1776-1784, 2018
34. Zhu W, Sha S, Liu Z, Li Y, Xu L, Zhang W, Qiu Y, Zhu Z: Influence of the occipital orientation on cervical sagittal alignment: A prospective radiographic study on 335 normal subjects. Sci Rep 8(1):15336, 2018