

RESEARCH

Open Access



Machine learning models for predicting tibial intramedullary nail length

Sercan Capkin^{1*}, Ali Ihsan Kilic¹, Hakan Cici², Mehmet Akdemir³ and Mert Kahraman Marasli⁴

Abstract

Background Tibial intramedullary nailing (IMN) represents a standard treatment for fractures of the tibial shaft. Nevertheless, accurately predicting the appropriate nail length prior to surgery remains a challenging endeavour. Conventional techniques frequently depend on data obtained intraoperatively, which may prolong surgical time and elevate radiation exposure. This study employs anthropometric measurements to evaluate and contrast the efficacy of machine learning (ML) models in predicting tibial IMN length.

Methods A retrospective analysis was conducted on 163 patients who had undergone tibial IMN. Anthropometric data were collected, including the subject's height, shoe size, olecranon-to-5th metacarpal distance (OM), and tibial tuberosity-to-medial malleolus distance (TTMM). Four ML models, namely linear regression, random forest, decision tree, and XGBoost, were employed for the purpose of predicting tibial IMN length. The performance of the models was evaluated using the mean squared error (MSE) and the R-squared values.

Results The linear regression model demonstrated superior performance compared to the random forest, decision tree, and XGBoost models, with an R-squared value of 0.89, an MSE of 117.53, and a root mean squared error (RMSE) of 10.84 mm. The strongest correlation with IMN length was demonstrated by TTMM ($r=0.911$), followed by height ($r=0.899$) and OM ($r=0.811$). Furthermore, TTMM provided the greatest contribution to prediction accuracy, thereby supporting its use as a reliable predictor in clinical settings. The correlation between shoe size and the dependent variable was weaker ($r=0.823$), and the inclusion of shoe size in the model negatively impacted the prediction accuracy. Despite their ability to handle non-linear relationships, the random forest and XGBoost models yielded higher MSE values, indicating limited improvement over linear regression. These findings underscore the linear nature of the relationship between anthropometric variables and IMN length, with linear regression offering the most reliable predictions.

Conclusion Combining anthropometric measurements with ML models, particularly linear regression, effectively predicts IMN length. This approach can streamline preoperative planning by reducing intraoperative measurements and minimizing surgery time and radiation exposure. Further validation with larger datasets is necessary to confirm these findings across diverse populations.

Keywords Tibial intramedullary nail, Machine learning, Anthropometric measurements, Preoperative planning, Linear regression

*Correspondence:
Sercan Capkin
sercancapkn@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Tibial intramedullary nailing (IMN) is a well-established method for the treatment of tibial shaft fractures, offering several key advantages. These include minimal invasiveness, stable fixation, and early mobilization [1]. Nevertheless, accurately determining the optimal nail length remains a significant challenge in the context of IMN surgery [2]. It has been assumed that traditional intraoperative techniques, such as the use of radiographic rulers and guidewires, are reliable for the selection of the appropriate nail length. However, these methods have been shown to increase both operative time and radiation exposure [3]. An incorrect selection of nail length can result in complications such as malalignment, joint irritation and the necessity for revision surgery. It is therefore imperative to develop reliable preoperative methods for predicting the optimal tibial IMN length [4].

Recent studies have indicated a strong correlation between the precise measurement of IMN length and a range of body parameters, including patient height, shoe size, and specific lower limb dimensions [2, 5–8]. A number of studies have identified notable correlations between these measurements and the selection of an appropriate IMN length, underscoring the value of tailored measurement techniques [2, 5–9]. Nevertheless, the accurate prediction of nail length presents a significant challenge due to the inherent variability in patient anatomy. This complexity makes the standardization of procedures a challenging endeavour and necessitates the utilisation of bespoke approaches to achieve optimal surgical outcomes.

The advent of ML has led to a transformation in the field of clinical decision-making in orthopaedics. The application of ML algorithms, including those such as linear regression, random forests, decision trees, and XGBoost, has demonstrated considerable potential in the prediction of outcomes through the analysis of large datasets and the identification of complex patterns [10, 11]. These models may prove significant in predicting the length of IMN based on patient-specific measurements, potentially offering a data-driven and personalised approach to surgical planning. The simplicity and interpretability of linear regression models are among their most valuable attributes. In contrast, more complex models, such as random forests and XGBoost, demonstrate robust performance by capturing non-linear relationships and reducing overfitting, particularly in the context of smaller datasets [10, 11].

This study aims to evaluate and compare the performance of four ML models—linear regression, random forest, decision tree, and XGBoost—in predicting tibial IMN length based on anthropometric measurements. We hypothesize that ML models utilizing preoperative anthropometric measurements offer a reliable method

for predicting tibial IMN length, potentially reducing reliance on intraoperative estimations. This research not only explores the potential of ML to enhance surgical planning but also addresses a significant gap in the existing literature by providing a comparative analysis of the predictive capabilities of different models. A better understanding of these relationships may improve preoperative planning and help reduce complications associated with inaccurate nail length selection, ultimately improving patient outcomes.

Materials and methods

Study design

This retrospective study analyzed data from patients who underwent interlocking tibial IMN treatment for tibial shaft fractures between 1 January 2018 and 30 August 2024. The data were gathered from two distinct medical centers: İzmir Bakırçay University Training and Research Hospital and İzmir Democracy University Training and Research Hospital. It should be noted that the study was approved by the Ethics Committee of İzmir Bakırçay University on 2024 (Approval No: 1806/1786) and was conducted by the principles outlined in the Declaration of Helsinki.

Study population

The inclusion criteria were as follows: patients must be at least 18 years of age, have a confirmed diagnosis of tibial shaft fracture, have undergone treatment with an interlocking IMN, and possess complete clinical and radiological records following the surgery. Conversely, the exclusion criteria comprised patients younger than 18 years, those treated with non-interlocking nails, patients with incomplete data sets, inadequate follow-up periods, additional pathologies affecting the tibial anatomy (such as contractures or joint deformities), and those who had undergone revision surgeries or other procedures on the tibia.

The initial cohort comprised 376 patients who had undergone IMN. Following the application of the aforementioned inclusion and exclusion criteria, 163 patients were deemed eligible for inclusion in the final analysis, as illustrated in Fig. 1. This rigorous selection process was implemented in order to guarantee a uniform study population and to establish reliable data for subsequent analysis.

Data collection

Anthropometric measurements were obtained from hospital medical records and in-person evaluations using standardized measurement tools. The following measurements were collected for each patient:

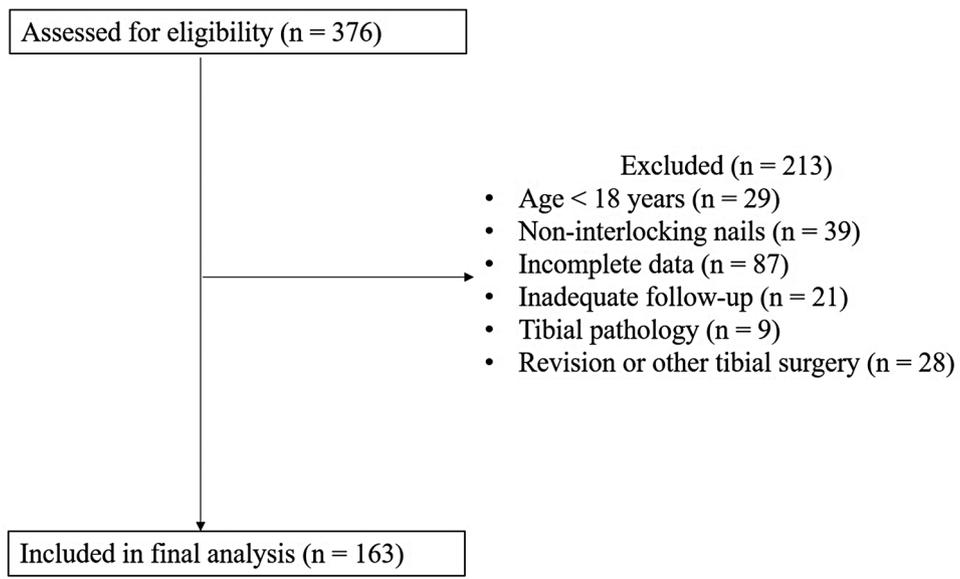


Fig. 1 Flowchart illustrating patient selection and exclusion criteria



Fig. 2 Measurement of olecranon-to-5th metacarpal distance (OM)

- **Height (cm):** Obtained from medical records or measured during hospital visits using a stadiometer.
- **Shoe size (EU):** Recorded in European sizes, either from hospital files or self-reported by patients during clinical visits.
- **Olecranon-to-5th metacarpal distance (OM, mm):** This distance was measured with the patient seated,

arm fully extended, and hand clenched in a fist position using a flexible tape (Fig. 2).

- **Tibial tuberosity-to-medial malleolus distance (TTMM, mm):** Measured from the unaffected leg with the patient seated and fully extended leg using a flexible tape (Fig. 3).
- **Tibial intramedullary nail length (mm):** The length of the tibial IMN used during surgery, retrieved from surgical records.

All measurements were taken during face-to-face evaluations to ensure consistency and reliability.

Machine learning models

To predict tibial IMN length, various ML models were developed and evaluated, using the collected anthropometric measurements as predictors and tibial IMN length as the dependent variable. The following models were applied:

- **Linear Regression:** A simple and interpretable model that assesses the linear relationship between the predictors and the dependent variable.
- **Random Forest:** An ensemble model that constructs multiple decision trees and averages their predictions to improve accuracy and reduce overfitting. Random forests are remarkably robust against overfitting in smaller datasets.
- **Decision Tree:** This model predicts the dependent variable by learning decision rules from the features, which helps it understand non-linear relationships between variables.



Fig. 3 Measurement of tibial tuberosity-to-medial malleolus distance (TTMM)

- **XGBoost:** A gradient boosting algorithm that sequentially builds models, with each iteration learning from the errors of the previous one. XGBoost is known for its high accuracy, especially in regression tasks.

The dataset was partitioned into a training set (80%) and an independent holdout test set (20%) for external validation, using the `train_test_split` function from scikit-learn. Prior to modelling, all continuous variables were normalized using the `StandardScaler` function in order to ensure that the features were on the same scale and to optimize model performance. The scaler was fitted exclusively on the training data and subsequently applied to the test data to prevent data leakage and ensure unbiased performance evaluation. The models were trained on the training set, and their performance on the test set was primarily evaluated using mean squared error (MSE), which penalizes larger errors more severely and is widely used for regression tasks. In addition, root mean squared error (RMSE) was calculated to enhance interpretability, as it is expressed in the same unit (millimeters, mm) as the target variable. RMSE provides a clinically meaningful estimation of the average prediction error and complements MSE by offering a more intuitive understanding of model performance. In addition to the holdout test set

Table 1 Descriptive statistics of the study variables

Variable	Mean \pm SD	Min	Max
Tibial IMN Length (mm)	338.65 \pm 33.93	280	420
Height (cm)	172.52 \pm 10.26	154	198
Shoe Size (EU)	40.70 \pm 2.12	36	47
TTMM (mm)	356.40 \pm 35.85	280	442
OM (mm)	384.40 \pm 75.88	251	568

Table 2 Pearson correlation coefficients for tibial IMN length and independent variables

Variable	Pearson Correlation (<i>r</i>)	<i>p</i> -value
Height (cm)	0.899	< 0.001
Shoe Size (EU)	0.823	< 0.001
TTMM (mm)	0.911	< 0.001
OM (mm)	0.811	< 0.001

evaluation, 10-fold cross-validation was conducted to improve the robustness and generalizability of the model performance assessment. No hyperparameter tuning was performed for any of the ML models; all models were trained using their default parameters as implemented in the scikit-learn and XGBoost libraries. This approach was chosen to ensure comparability and minimize overfitting, especially given the moderate sample size.

Statistical analysis

All statistical analyses were conducted using Python (version 3.12.6) with the SciPy (v1.14.1) and statsmodels (v0.14.3) libraries. The normality of residuals from all ML models was assessed using the Shapiro–Wilk test. A *p*-value < 0.05 was considered statistically significant. Additionally, 95% confidence intervals (CIs) for the regression coefficients in the linear regression model were calculated to evaluate the precision and reliability of its estimates. These statistical procedures were implemented to ensure the robustness of model assumptions and enhance the interpretability of the findings.

Results

Descriptive statistics

The study included 163 patients who underwent IMN for tibial shaft fractures. Table 1 summarizes descriptive statistics for the study variables.

Correlation analysis

Pearson correlation analysis revealed significant positive associations between tibial IMN length and all independent variables. The strongest correlation was found between TTMM and IMN length ($r=0.911$), followed by height ($r=0.899$), shoe size ($r=0.823$), and OM ($r=0.811$). All correlations were statistically significant ($p < 0.001$). The correlation coefficients for these variables are presented in Table 2.

Linear regression model results

A multiple linear regression model was developed to estimate IMN length based on anthropometric parameters. The model explained 89.45% of the variance in IMN length ($R^2 = 0.8945$), indicating a strong predictive relationship between the predictors and the outcome variable. The MSE was 117.53, representing the average squared difference between predicted and actual values. The RMSE was 10.84 mm, providing a clinically interpretable measure of prediction error. Regression coefficients for each variable are presented in Table 3.

The regression analysis identified height, TTMM, and OM as significant positive predictors of IMN length, whereas shoe size exhibited a negative but statistically non-significant association. The high R^2 value supports the utility of anthropometric measurements for accurate preoperative estimation of IMN length. To assess the precision and statistical reliability of the model estimates, 95% CIs were computed and are presented in Table 4.

Predictors such as height, TTMM, and OM demonstrated statistically significant effects, as their CIs did not include zero. This finding supports the robustness of the regression model and underscores the strong association between these anthropometric parameters and IMN length.

For comprehensive evaluation of model performance and statistical inference, regression coefficients (Table 3) were estimated using the *scikit-learn* library, while 95% CIs (Table 4) were derived using the *statsmodels* package. The differences between the coefficients reported in the two tables are primarily due to the standardization of predictor variables within the modeling pipeline. Despite this methodological difference, both approaches consistently identified the same significant predictors. This dual-method strategy was deliberately employed to ensure both accurate model training and rigorous statistical interpretation.

Residual analysis of linear regression model

Residual analysis was conducted to evaluate the accuracy and distributional assumptions of the linear regression model. The minimum residual (i.e., the difference between actual and predicted IMN lengths) was -33.63 mm, and the maximum residual was 11.58 mm. The mean residual was -3.56 mm, indicating that the model tended to slightly underestimate actual values on average. The standard deviation (SD) of the residuals was 10.24 mm, reflecting an acceptable level of prediction variability. To assess the normality of residuals, the Shapiro–Wilk test was performed. The test statistic was 0.9934 with a p-value of 0.667 , indicating no significant deviation from normality ($p > 0.05$). This finding supports the assumption of normally distributed residuals, which

Table 3 Regression coefficients for IMN length

Predictor	Coef- ficient (β)	Std. Error	t-value	p-value	Interpretation
Intercept	340.31	1.165	292.10	0.000	-
Height (cm)	14.33	4.992	2.870	0.005	Positive impact
Shoe Size (EU)	-5.37	3.132	-1.715	0.089	Negative impact
TTMM (mm)	8.33	1.875	4.440	0.000	Positive impact
OM (mm)	15.11	4.293	3.519	0.001	Positive impact

Table 4 95% CIs for regression coefficients

Predictor	Coefficient (β)	95% CI
Height (cm)	1.1043	[0.267, 1.941]
Shoe Size (EU)	-1.6455	[-4.096, 0.805]
TTMM (mm)	0.4562	[0.248, 0.665]
OM (mm)	0.1146	[0.073, 0.157]

is a key prerequisite for the validity of linear regression inference.

Residual analysis of other machine learning models

Residual analyses were also conducted for the Random Forest, Decision Tree, and XGBoost models to evaluate their predictive accuracy and error distributions. The mean residuals and SDs were 2.24 ± 12.83 mm for Random Forest, 1.21 ± 15.52 mm for Decision Tree, and 2.70 ± 14.41 mm for XGBoost. The Shapiro–Wilk test was used to assess the normality of residuals. The residuals of the XGBoost model did not significantly deviate from normality ($p = 0.5380$), supporting the assumption of normally distributed errors. In contrast, the residuals of the Random Forest ($p = 0.0187$) and Decision Tree ($p < 0.001$) models significantly deviated from normality, suggesting potential violations of distributional assumptions in these models.

Model evaluation

The F-statistic for the linear regression model was 248.0, with a p-value of $5.45e-67$, indicating that the model as a whole was statistically significant and exhibited a good overall fit. A comparative summary of performance metrics for all ML models, including R^2 , MSE, and RMSE, is provided in Table 5. These metrics demonstrate that the linear regression model achieved the highest predictive accuracy, with an R^2 of 0.8945 and the lowest RMSE (10.84 mm), followed by Random Forest (RMSE = 13.00 mm), XGBoost (14.63 mm), and Decision Tree (15.57 mm). Lower MSE and RMSE values indicate better model performance, and RMSE is expressed in mm to enhance clinical interpretability.

Table 5 Comparative performance metrics of ML models for predicting tibial IMN length

Model	R ²	MSE	RMSE (mm)
Linear Regression	0.8945	117.53	10.84
Random Forest	0.8482	169.01	13.00
Decision Tree	0.7820	242.42	15.57
XGBoost	0.8070	213.99	14.63

Note: R² indicates the proportion of variance in tibial IMN length explained by the model. MSE and RMSE reflect prediction error, with lower values indicating better performance. RMSE is reported in mm to enhance clinical interpretability

Random forest results

The Random Forest regression model yielded an MSE of 169.01 and an R² value of 0.8482. The RMSE was 13.00 mm, indicating a moderate level of average prediction error. Although Random Forest algorithms are generally effective in capturing complex, nonlinear relationships, the model demonstrated lower predictive accuracy than linear regression in this context, as evidenced by its higher MSE and lower R². These findings suggest that the simpler linear regression model outperformed Random Forest in estimating IMN length within this dataset.

Decision tree results

The Decision Tree regression model yielded an MSE of 242.42 and an R² value of 0.782. The RMSE was 15.57 mm, indicating a relatively higher average prediction error. These results suggest that the Decision Tree model underperformed compared to both the linear regression and Random Forest models, as evidenced by its higher MSE and lower R². Although decision trees are useful for capturing complex relationships, in this study, the simpler linear regression and Random Forest models provided more accurate predictions of IMN length.

XGBoost results

The XGBoost regression model yielded an MSE of 213.99 and an R² value of 0.807, indicating better performance than the Decision Tree model but lower predictive accuracy compared to the linear regression and Random Forest models. The RMSE was 14.63 mm, reflecting a moderately high level of average prediction error. Despite its advanced capabilities in handling complex data structures, the XGBoost model was less effective in this study than simpler models such as linear regression when predicting IMN length.

Cross-validation results

To improve the reliability of model evaluation, 10-fold cross-validation was performed on the full dataset. The mean R² scores and corresponding SDs across folds were 0.784 ± 0.144 for the linear regression model, 0.796 ± 0.104 for the Random Forest, 0.661 ± 0.175 for the Decision Tree, and 0.748 ± 0.133 for the XGBoost model.

These findings support the consistency of model performance across different data subsets. Among all models, the Random Forest Regressor achieved the highest average R² score, suggesting superior generalization ability, followed closely by linear regression and XGBoost.

Discussion

In this study, we aimed to predict the length of tibial IMN using anthropometric measurements by applying and comparing various ML models, including linear regression, random forest, decision tree, and XGBoost. The findings of this study offer invaluable insights into the manner in which these models predict IMN length based on patient-specific characteristics.

This study makes a pioneering contribution to the literature in that it calculates the size of the IMN based on anthropometric data. There is a paucity of studies in which artificial intelligence (AI) and ML models have been employed to ascertain implant components in procedures such as knee, hip, and shoulder arthroplasty, which represent some of the most prevalent surgical practices in orthopaedics [12–18]. A systematic review by Salman et al. [14] demonstrated that AI models are highly accurate in estimating total knee arthroplasty component dimensions. The accuracy rates for AI models in estimating femoral component dimensions ranged from 88.3 to 99.7%, while the accuracy rates for tibial component dimensions ranged from 90 to 99.9%. Furthermore, the deviations were limited to just one size. Similarly, the present study achieved high accuracy in predicting tibial IMN length using anthropometric data, with the model accounting for 89.45% of the variability in IMN length (R² = 0.8945). This illustrates a robust correlation between the input variables and the length of the intramedullary nail. As the number of studies that employ ML for the purpose of implant dimension prediction increases, it is anticipated that the accuracy and reliability of these models will improve, thereby further enhancing their utility in the context of preoperative planning. The integration of AI into surgical workflows has the potential to significantly enhance clinical decision-making, minimise human error and, ultimately, improve patient outcomes.

The linear regression model demonstrated superior performance compared to more complex models, such as random forest and XGBoost, as evidenced by its lower MSE and higher R² values. These results indicate that the relationship between the independent variables (height, shoe size, TTMM, and OM) and IMN length is predominantly linear. Consequently, simpler models such as linear regression are more appropriate for this task. It is noteworthy that the model explained approximately 89.45% of the variance in IMN length, which serves to emphasise the effectiveness of anthropometric measurements in preoperative planning for tibial fractures.

Furthermore, the RMSE of the linear regression model was 10.84 mm, indicating that, on average, the predicted IMN length deviated from the actual value by approximately 1 cm. This unit-consistent metric enhances the clinical interpretability of the model and supports its potential applicability in real-world surgical planning. Clinically, an average deviation of approximately 10 mm is generally considered acceptable for tibial IMN length prediction, as intramedullary nails are commonly manufactured in 20 mm increments. Therefore, an RMSE of 10.84 mm falls within an acceptable error margin and is unlikely to result in clinically significant implant selection errors. Combined with the high predictive accuracy of the model ($R^2 = 0.8945$), these findings confirm that the model meets both statistical and clinical thresholds of adequacy for preoperative use.

In our study, complex ML models such as Random Forest and XGBoost—typically known for their superior performance on larger datasets—may have been adversely affected by the limited sample size. Acknowledging this potential issue, we implemented several methodological precautions to minimize the risk of overfitting, including 10-fold cross-validation, holdout validation, and default hyperparameters without any tuning. Nevertheless, our findings revealed overfitting, particularly in the Random Forest model, which exhibited a high cross-validation R^2 score but a higher RMSE and significant deviation from normality in residuals. These observations suggest that more straightforward and more interpretable models, such as linear regression, may offer more reliable and clinically applicable results in small datasets.

The ML models employed in this study enabled the prediction of tibial IMN length using only external anthropometric parameters such as height, TTMM, OM, and shoe size, without the need for radiological imaging. This approach stands in contrast to conventional preoperative planning techniques, which rely heavily on radiographic evaluations. For instance, Keltz et al. [3] reported that digital templating based on contralateral leg radiographs required post hoc adjustment of the selected nail length in approximately 28% of cases, underscoring the variability and limitations inherent in imaging-based methods.

In comparison, our ML-based linear regression model achieved superior predictive performance ($R^2 = 0.8945$, RMSE=10.84 mm). In addition to outperforming radiographic techniques, it also exceeded the accuracy of traditional anthropometry-based estimation methods. Albay and Kaygusuz [2], for example, developed sex-specific formulas using single anthropometric measures such as the knee-to-ankle joint line distance (JJ), tibial tuberosity to medial malleolus (TM), tibial tuberosity to ankle joint (TA), and OM, with reported R^2 values of 0.8284 in males (JJ) and 0.8735 in females (TM). Unlike these single-variable approaches, our model integrated

multiple anthropometric predictors, thereby improving both estimation accuracy and clinical utility.

Further supporting these findings, Galbraith et al. [4] compared various radiographic and anthropometric methods for IMN length estimation and reported the highest accuracy for AP scanograms (100%) and intraoperative ruler/guidewire measurements (94%). In contrast, they observed particularly low accuracy rates for isolated anthropometric parameters such as body height (13%) and TM (38%). These results highlight the limitations of single-measure techniques in preoperative planning. Similarly, Issac et al. [5] reported that even the most accurate traditional anthropometric method—TA plus 11 mm—yielded 81% accuracy, with other techniques performing notably lower. These approaches often rely on fixed offsets and assume linearity, which can limit their adaptability. While these radiographic methods demonstrate excellent accuracy under ideal conditions, they require specialized equipment, trained personnel, and may not always be feasible in time-sensitive or resource-limited settings. In contrast, our ML-based model offers a practical, accessible alternative that maintains a high level of predictive performance using only easily obtainable anthropometric inputs.

In addition to these quantitative findings, it is also important to consider the potential practical value of the proposed model. Although intraoperative fluoroscopic measurements and preoperative imaging are established standard practices for determining IMN length, ML models can offer valuable assistance in certain clinical scenarios. In pre-hospital settings, emergencies, or facilities with limited imaging capabilities, a rapid and non-invasive estimation method may support early surgical planning and reduce reliance on intraoperative decision-making. Our model is not intended to replace conventional techniques, but rather to complement them by enhancing efficiency and accuracy when standard methods are unavailable or impractical.

The results demonstrated a robust correlation between IMN length and TTMM, height, and OM. TTMM demonstrated the highest correlation with IMN length ($r=0.911$), followed by height ($r=0.899$) and OM ($r=0.811$), which is consistent with the existing literature that emphasises the significance of lower limb dimensions in selecting appropriate IMN length [2, 5, 6, 9]. However, shoe size exhibited a weaker correlation ($r=0.823$) and negatively impacted the linear regression model, indicating that it may not be a reliable predictor when used in conjunction with other anthropometric variables. Despite its lack of statistical significance in the multivariable model ($p=0.089$), shoe size was deliberately retained due to its acceptable correlation with IMN length and its contribution to the overall generalizability of the model. Its inclusion did not lead

to multicollinearity or degrade model performance, as reflected by the high R^2 (0.8945) and clinically acceptable RMSE (10.84 mm). Moreover, all anthropometric predictors—including height, shoe size, TTMM, and OM—were selected a priori based on clinical relevance and prior literature, and evaluated collectively to preserve the integrity of the multivariable design. Accordingly, excluding shoe size could have reduced model robustness, particularly in diverse patient subgroups.

In a study conducted by Jain et al. [19], five anthropometric measurements were evaluated in 100 patients, including the distance from the knee joint to the ankle joint (K-A), TTMM, OM, thigh length, and leg length. The results of their study were in accordance with our findings, with the strongest correlations observed for TTMM and OM. Furthermore, they emphasised that OM exhibited the highest accuracy due to its ease of palpation, whereas TTMM was more susceptible to inter-observer variability. Similarly, Hegde et al. [7] reported strong correlations between OM and IMN length ($r=0.966$), indicating that OM may be a reliable alternative in cases where TTMM measurements are complex, such as in obese patients or those with bilateral tibial fractures.

Blair [20] additionally investigated the potential of OM as a predictor for IMN length, developing a predictive formula based on OM measurements. His study corroborated the reliability of OM, particularly in instances where TTMM was impractical. The findings of our study corroborate those of Blair, indicating that the incorporation of OM into ML models enhances the accuracy of preoperative planning by reducing the impact of human error.

In a study conducted by Sharma et al. [9], the usefulness of arm length (AL) as a new anthropometric measure for estimating tibial IMN length was assessed. The researchers observed that AL exhibited a slight average discrepancy and the most precise agreement limits when compared to conventional lower limb measurements, including TTMM and knee joint line to medial malleolus (KJL-MM). Although AL demonstrated a robust correlation with tibial nail length, the study substantiated that TTMM remained the most dependable predictor ($r=0.911$) when integrated with a multitude of anthropometric measurements.

One of the significant strengths of our study is the use of multiple ML models to predict tibial IMN length based on anthropometric measurements. By incorporating a variety of variables, such as height, shoe size, TTMM, and OM, we were able to provide a comprehensive analysis of the factors influencing IMN length. Additionally, using ML models allowed us to compare different approaches and select the most accurate model, which in our case was linear regression, explaining approximately

89.45% of the variance in IMN length. The comparatively lower performance of more complex models such as Random Forest and XGBoost was interpreted as a reflection of the predominantly linear structure of the dataset, rather than a limitation of feature selection or data robustness. All anthropometric variables were selected based on prior clinical evidence and literature support. While the study revealed strong correlations between anthropometric variables and IMN length, it does not aim to establish causality. Rather, these associations were leveraged for their predictive value within the context of preoperative planning, where even proportion-based relationships may provide clinically meaningful guidance. However, there are some limitations to consider:

1. While robust, our dataset was limited to a specific patient population undergoing tibial IMN, which may impact the generalizability of the findings. The model's clinical applicability is also confined to a single surgical context. However, this approach may serve as a foundational framework for future studies, as similar ML-based models could be adapted for implant sizing in other orthopedic procedures, such as femoral or humeral nailing or joint arthroplasty.
2. The variability in anthropometric measurements, particularly shoe size, introduced some inconsistencies in prediction accuracy, suggesting that certain variables may need more reliable predictors across different patient groups.
3. Although ML models offer significant advantages, further validation with larger datasets and diverse populations would be beneficial to confirm the applicability of our findings in broader clinical settings. Additionally, the absence of external validation limits the ability to assess the model's performance in independent patient cohorts.

Conclusion

In the present study, we demonstrated the potential of ML models in predicting the length of the tibial IMN based on patient-specific measurements. The advantages of utilising data-driven techniques to enhance the precision of preoperative planning were elucidated through a comparative analysis of diverse ML methodologies. Our findings demonstrated that even relatively simple models, such as linear regression, can achieve high levels of predictive accuracy. The integration of ML models represents an efficient means of enhancing clinical decision-making and improving patient outcomes. Further research is required to investigate the broader applications of ML in orthopaedics, with the aim of validating and refining predictive models for routine clinical use.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12891-025-08657-1>.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

S. Capkin designed the study, performed data analysis, interpreted the data, and was a major contributor to writing the manuscript. A.I. Kilic assisted in manuscript preparation, data analysis, and study design. Hakan Cici contributed to data analysis, study design, and manuscript preparation. Mehmet Akdemir and Mert Kahraman Maraslı critically revised the manuscript, contributed to study design and data interpretation. All authors read and approved the final manuscript.

Funding

The authors declare that no funding was received for conducting this study.

Data availability

The complete source code and a sample anonymized dataset used to develop the ML models are provided as a supplementary ZIP file titled "imn_ml_model_package.zip." This archive includes the Python script, dataset, and a README file with detailed instructions, enabling full reproducibility of the analyses. Further data may be available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of [Bakircay University Editoril Board] on [2024] (Approval No: 1806/1786). All procedures were conducted in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments. Informed consent was obtained from all individual participants included in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Faculty of Medicine, Department of Orthopaedics and Traumatology, Izmir Bakircay University, Izmir 36665, Turkey

²Faculty of Medicine, Department of Orthopaedics and Traumatology, Izmir Democracy University, Izmir, Turkey

³Department of Orthopaedics and Traumatology, Izmir Ekol Hospital, Izmir, Turkey

⁴Department of Orthopaedics and Traumatology, Gebze Fatih State Hospital, Kocaeli, Turkey

Received: 25 October 2024 / Accepted: 14 April 2025

Published online: 21 April 2025

References

- Cazzato G, Saccomanno MF, Noia G, Masci G, Peruzzi M, Marinangeli M, et al. Intramedullary nailing of tibial shaft fractures in the semi-extended position using a Suprapatellar approach: A retrospective case series. *Injury*. 2018;49(Suppl 3):61–4. <https://doi.org/10.1016/j.injury.2018.10.003>.
- Albay C, Kaygusuz MA. Formulae derived from anthropometric measurements to estimate ideal tibial nail length. *Acta Ortop Bras*. 2021;29(2):76–80. <https://doi.org/10.1590/1413-785220212902244108>.
- Keltz E, Dreyfuss D, Ginesin E, Ghayeb N, Hous N, Yavnai N, Norman D, Stahl I. Preoperative evaluation of intramedullary tibial nail Measurements-A review of the literature and a new technique using contralateral radiographs and digital planning. *J Am Acad Orthop Surg Glob Res Rev*. 2019;3(3):e015. <https://doi.org/10.5435/JAAOSGlobal-D-19-00015>.
- Galbraith JG, O'Leary DP, Dailey HL, Kennedy TE, Mitra A, Harty JA. Preoperative Estimation of tibial nail length: because size does matter. *Injury*. 2012;43(11):1962–8. <https://doi.org/10.1016/j.injury.2012.07.190>.
- Issac RT, Gopalan H, Abraham M, John C, Issac SM, Jacob D. Preoperative determination of tibial nail length: an anthropometric study. *Chin J Traumatol*. 2016;19(3):151–5. <https://doi.org/10.1016/j.cjtee.2016.03.003>.
- Venkateswaran B, Warner RM, Hunt N, Shaw DL, Tulwa N, Deacon P. An easy and accurate preoperative method for determining tibial nail lengths. *Injury*. 2003;34(10):752–5. [https://doi.org/10.1016/s0020-1383\(02\)00370-4](https://doi.org/10.1016/s0020-1383(02)00370-4).
- Hegde A, Mohammed N, Ahmed NR. Correlation between tibial nail length and olecranon to 5th metacarpal head measurement: an anthropometric study. *Chin J Traumatol*. 2019;22(6):361–3. <https://doi.org/10.1016/j.cjtee.2019.07.002>.
- Şahin R, Şahin S, Kazdal C, Balık MS. Can the length of the tibia nail be predicted correctly before the operation according to the patient's height and shoe size?? *Cureus*. 2024;16(1):e52653. <https://doi.org/10.7759/cureus.52653>.
- Sharma A, Sinha S, Gupta S, Gupta A, Narang A, Sharma P, Kanjia RK. Evaluation of arm length as a new upper limb anthropometric method for preoperative Estimation of tibial intramedullary nail length. *Strategies Trauma Limb Reconstr*. 2021;16(1):20–6. <https://doi.org/10.5005/jp-journals-10080-1520>.
- Martin RK, Ley C, Pareek A, Groll A, Tischer T, Seil R. Artificial intelligence and machine learning: an introduction for orthopaedic surgeons. *Knee Surg Sports Traumatol Arthrosc*. 2022;30(2):361–4. <https://doi.org/10.1007/s00167-021-06741-2>.
- Groot OQ, Ogink PT, Lans A, Twining PK, Kapoor ND, DiGiovanni W, et al. Machine learning prediction models in orthopedic surgery: A systematic review in transparent reporting. *J Orthop Res*. 2022;40(2):475–83. <https://doi.org/10.1002/jor.25036>.
- Boubekri A, Murphy M, Scheidt M, Shivdasani K, Anderson J, Garbis N, et al. Artificial intelligence machine learning algorithms versus standard linear demographic analysis in predicting implant size of anatomic and reverse total shoulder arthroplasty. *J Am Acad Orthop Surg Glob Res Rev*. 2024;8(8):e2400182. <https://doi.org/10.5435/JAAOSGlobal-D-24-00182>.
- Zampogna B, Torre G, Zampoli A, Parisi F, Ferrini A, Shanmugasundaram S, et al. Can machine learning predict the accuracy of preoperative planning for total hip arthroplasty, basing on patient-related factors? An explorative investigation on supervised machine learning classification models. *J Clin Orthop Trauma*. 2024;53:102470. <https://doi.org/10.1016/j.jcot.2024.102470>.
- Salman LA, Khatkar H, Al-Ani A, Alzobi OZ, Abudalou A, Hatnoully AT, et al. Reliability of artificial intelligence in predicting total knee arthroplasty component sizes: a systematic review. *Eur J Orthop Surg Traumatol*. 2024;34(2):747–56. <https://doi.org/10.1007/s00590-023-03784-8>.
- Park KB, Kim MS, Yoon DK, Jeon YD. Clinical validation of a deep learning-based approach for preoperative decision-making in implant size for total knee arthroplasty. *J Orthop Surg Res*. 2024;19(1):637. <https://doi.org/10.1186/s13018-024-05128-6>.
- Kunze KN, Polce EM, Patel A, Courtney PM, Levine BR. Validation and performance of a machine-learning derived prediction guide for total knee arthroplasty component sizing. *Arch Orthop Trauma Surg*. 2021;141(12):2235–44. <https://doi.org/10.1007/s00402-021-04041-5>.
- Burge TA, Jones GG, Jordan CM, Jeffers JRT, Myant CW. A computational tool for automatic selection of total knee replacement implant size using X-ray images. *Front Bioeng Biotechnol*. 2022;10:971096. <https://doi.org/10.3389/fbioe.2022.971096>.
- Polce EM, Kunze KN, Paul KM, Levine BR. Machine learning predicts femoral and tibial implant size mismatch for total knee arthroplasty. *Arthroplast Today*. 2021;8:268–e772. <https://doi.org/10.1016/j.artd.2021.01.006>.
- Jain RK, Deshpande M, Bohra T, Jain N, Gulve M. The correlation of various anthropometric measurements with tibia interlocking nail length measured intra-operatively. *Orthopaedic Journal of M P Chapter [Internet]*. 2020;27(1). Available from: <https://ojmpc.com/index.php/ojmpc/article/view/111>

20. Blair S. Estimating tibial nail length using forearm referencing. *Injury*. 2005;36(1):160–2. <https://doi.org/10.1016/j.injury.2003.09.032>. <https://doi.org/10.1016/j.injury.2003.09.032>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.