European Review for Medical and Pharmacological Sciences 2024; 28: 4666-4670

# Visual analogue scale foot and ankle *vs.* short-form 36 quality of life scores: artificial intelligence using machine learning analysis with an external validation

C. ANGTHONG[1], P. RAJBHANDARI[2], W. ANGTHONG[3]

[1]Faculty of Medicine, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
[2]Department of Orthopedics, R.I.P.A.S. Hospital, Bandar Seri Begawan, Brunei
[3]Department of Radiology, Faculty of Medicine, Thammasat University, Pathum Thani, Thailand

**Abstract. – OBJECTIVE:** We aimed to utilize artificial intelligence (AI) *via* machine learning (ML) to analyze the relationship between visual analogue scale foot and ankle (VASFA) and short-form 36 (SF-36) quality of life scores and determine AI's performance over the aforementioned analysis.

**MATERIALS AND METHODS:** We collected data from our registry of 819 data units or rows of datasets of foot and ankle patients with VASFA, SF-36 scores, and other demographic data. They were prepared and verified to be a proper input for building ML models using a web-based algorithm platform. After the first ML model was developed using random forest regression, the SF-36 percentage value was set as an endpoint. We developed a second ML model to evaluate it against the current algorithm. This new model employed a gradient-boosting regressor, where we omitted a key parameter, SF_Total, to correct the overfitting. We performed an external validation based on an unseen dataset from 42 data units of patients.

**RESULTS:** Internal validity showed an excellent relationship among the VASFA, SF-36 total score, and overall SF-36 percent values at a correlation coefficient (R2 score) of 1.000 based on the random forest regression model of ML (first model: 28XJ). The VASFA percent value of the total score (0=worst; 100=best) demonstrated the dynamic changes in the three zones of the score levels; these were unsatisfactory: ≤ 57.25; borderline: 57.26-80.99; satisfactory: ≥ 81 and could impact the levels of overall SF-36 percent value. A second ML model (model FK13) showed an R2 score of 0.977, which was a great performance. External validation showed no significant difference between the predicted and actual values, with a two-tailed *p*-value of 0.2136.

**CONCLUSIONS:** Our ML models predicted excellent relationships among VASFA, with or without SF-36 total score and overall SF-36 percentage values, with evidence from external validation.

*Key Words:*
Artificial intelligence, Foot, Ankle, Visual analog scale foot and ankle, Short-form 36.

## Introduction

The visual analogue scale for foot and ankle (VASFA) is a well-known and reliable score for measuring functional and other effects of foot and ankle disabilities[1]. It has been extensively validated in several studies[1,2] for its correlation with short-form 36 (SF-36) quality of life scores. VASFA is used to evaluate the functional status of patients with foot and ankle pathologies, including those undergoing total ankle replacement surgery[1-5]. However, there were some limitations in delineating the detailed impact of each subscale of the VASFA on the SF-36. With the rise of present-day Artificial Intelligence (AI) technology, it can effectively analyze various conditions in orthopedics and trauma fields due to its superior performance[6,7]. Currently, little is known about the role of AI assistance in the analysis of correlations between functional outcomes and health-related quality of life (QoL) scores in foot and ankle patients. The present study aimed to utilize AI to analyze the in-depth relationship between the VASFA and SF-36 quality of life scores, including data drivers or predictors and internal validity. This was also used to determine AI performance in the aforementioned analyses.

## Materials and Methods

We collected 819 data units or rows of datasets in all patients from our hospital database registry



*Corresponding Author:* Chayanin Angthong, MD; e-mail: chayanin.an@kmitl.ac.th

of foot and ankle patients who were diagnosed and treated by a single fellowship-trained foot and ankle surgeon. Each data unit or row of the dataset consisted of the VASFA with its subscales, SF-36 scores, and other demographic data as multiple items in Table I. All data units or rows were prepared and verified to be used as a disciplined dataset and a proper input for building machine learning (ML) models using a web-based algorithm platform on ObviouslyAI (available at: https://www.obviously.ai/).

ML automated the selection of an appropriate method to build a model for the analysis and prediction of the in-depth relationship between VASFA and SF-36 scores, including data drivers or predictors. The SF-36 percentage was set as the endpoint. Another variable with both categorical and continuous data was the drivers. As mentioned earlier, ML selected the random forest regression model to analyze the dataset. Results, including predictions, were obtained following complete analysis. We also created another ML model that employed a gradient-boosting regressor, where we omitted a key parameter, SF_Total, to correct the overfitting. Subsequently, we performed external validation based on an unseen dataset from 42 data units of foot and ankle patients under the care of the same foot and ankle surgeon. We used the zones of interest from the ML model to predict SF-36 percent values from each zone and then compared those values with the actual values from the original data unit using the *t*-test for analysis on GraphPad (La Jolla, CA, USA).

The institutional Ethics Committee of the Faculty of Medicine, Thammasat University, approved this study on December 9$^{th}$, 2016, with approval number MTU-OT-9-CR079-079/53. This approval included permission for data collection and analysis, including nonlinear correlation analysis, with no prohibition of AI use.

### *Statistical Analysis*

Internal validity was used to assess the relationship among parameters based on the random forest regression model or gradient-boosting regressor of the ML. Regarding the impact of drivers for in-depth details, our models 28XJ and FK13 were used to assess the relationships between VASFA and SF-36 percent values as a point-to-point analysis and the results. Regarding the external validation based on an unseen dataset from 42 data units of foot and ankle patients, a comparison was made between the predicted and actual values using GraphPad by Dotmatics (La Jolla, CA, USA). A $p$-value < 0.05 was considered as the significant difference.

## Results

Internal validity showed an excellent relationship between VASFA, SF-36 total score, and overall SF-36 percent values at a correlation coefficient ($R^2$ score) of 1.000 based on the random forest regression model of the ML (first model: 28XJ). We created a second model, model FK13, which used a gradient-boosting regressor that showed an $R^2$ score of 0.977, which was a great performance. In this method, we omitted a key parameter, SF_Total, to correct the overfitting in our study.

Regarding the impact of drivers for in-depth details, our model 28XJ revealed in-depth relationships between VASFA and SF-36 percent values (Figure 1). The VASFA percentage value of the total score (0=worst; 100=best) demonstrated interesting in-depth details in terms of dynamic changes in the three score level zones as a nonlinear curve pattern. There were three zones of interest, namely unsatisfactory: $\leq$ 57.25; borderline: 57.26-80.99; satisfactory: $\geq$ 81, which could impact the levels of overall SF-36 percent value.

Another important driver was the patient's sex, which had a different impact on the SF-36 percent value. The female sex negatively affected the SF-36 percent value by about -3.87; however, the male sex affected the same value in the range of -3.21–25.12.

The external validation showed no significant difference between the predicted and actual values, with a two-tailed $p$-value of 0.2136. This analysis confirmed the substantial validity of model 28XJ, which could predict the value of SF-36 using the VASFA score with good performance.

## Discussion

We highlighted the interesting role of AI using ML in demonstrating the in-depth relationship between VASFA and SF-36 scores. ML can identify the effects of factors not previously documented in foot and ankle patients, such as different zones of interest from the VASFA that uniquely influence SF-36 scores, as well as varying impacts of sex on the same score. The relationship between VASFA and SF-36 in these zones of interest has not been previously reported. Our ML model,

**Table I.** Dataset for the machine learning models building.

| Study number | Side | Period of symptom | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Item 6 | Item 7 | Item 8 | Item 9 | Item 10 | Item 11 | Item 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 738845 | RT | > 4 weeks | 65 | 30 | 30 | 20 | 40 | 40 | 0 | 50 | 50 | 50 | 0 | 30 |
| 956016 | LT | > 1 weeks | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| 956016 | LT | > 4 weeks | 40 | 70 | 70 | 70 | 60 | 60 | 100 | 60 | 50 | 40 | 80 | 50 |
| 1181597 | RT | > 4 weeks | 40 | 70 | 40 | 70 | 40 | 40 | 100 | 70 | 80 | 40 | 30 | 70 |
| 716681 | LT | 1 week | 60 | 100 | 100 | 0 | 40 | 100 | 100 | 40 | 50 | 70 | 100 | 70 |
| 716681 | LT | > 4 weeks | 70 | 100 | 100 | 50 | 70 | 100 | 100 | 30 | 100 | 100 | 100 | 100 |
| 716681 | LT | 1 year | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 716681 | LT | 1.1 year | 100 | 60 | 90 | 90 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 812569 | LT | > 1-2 weeks | 100 | 100 | 90 | 100 | 100 | 100 | 100 | 100 | 90 | 100 | 100 | 100 |
| 770975 | RT | > 1 week | 40 | 30 | 20 | 20 | 80 | 30 | 0 | 60 | 70 | 50 | 30 | 60 |
| 1154873 | LT | 1 week | 50 | 50 | 0 | 0 | 0 | 100 | 100 | 50 | 0 | 0 | 0 | 50 |
| 1154873 | LT | 1.9 year | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 1177799 | RT | > 1 weeks | 30 | 30 | 30 | 0 | 0 | 100 | 0 | 50 | 50 | 50 | 0 | 0 |
| 1177799 | RT | > 4 weeks | 0 | 80 | 80 | 50 | 50 | 50 | 100 | 0 | 0 | 0 | 50 | 0 |
| 1177799 | RT | > 8 weeks | 80 | 100 | 100 | 50 | 50 | 80 | 100 | 60 | 0 | 0 | 100 | 100 |
| 1112475 | RT | > 4 weeks | 80 | 20 | 20 | 20 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| 1112475 | RT | > 8 weeks | 100 | 100 | 100 | 100 | 100 | 20 | 100 | 100 | 100 | 100 | 50 | 100 |
| 1112475 | RT | > 6 months | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 1112475 | RT | 2 years | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 748784 | LT | > 2 weeks | 60 | 100 | 100 | 100 | 0 | 0 | 100 | 70 | 50 | 80 | 80 | 0 |
| 1120042 | RT>LT | > 4 weeks | 60 | 100 | 100 | 100 | 100 | 100 | 0 | 50 | 100 | 100 | 100 | 50 |
| 732433 | RT | > 3 weeks | 70 | 50 | 50 | 50 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| 732433 | RT | > 2 weeks | 70 | 50 | 50 | 90 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| 732433 | RT | > 8 weeks | 100 | 95 | 95 | 90 | 80 | 100 | 100 | 100 | 100 | 100 | 100 | 95 |

Item - Visual Analogue Scale Foot and Ankle score in each item, RT – right, LT – left.

| Driver Values<br>When driver is:<br>VASFA_Percent | Driver Outcomes<br>Impact on:<br>SF_Percent |
|---|---|
| 19.25 to 24.0 | 21.22% |
| 43.0 to 47.75 | 20.03% |
| 5.0 to 9.75 | 18.07% |
| 85.75 to 90.5 | 17.61% |
| 52.5 to 57.25 | 17.34% |
| 14.5 to 19.25 | 13.73% |
| 81.0 to 85.75 | 12.45% |
| 71.5 to 76.25 | 10.13% |
| 57.25 to 62.0 | 8.89% |
| 66.75 to 71.5 | 7.46% |
| 47.75 to 52.5 | 6.08% |
| 9.75 to 14.5 | 5.58% |
| 76.25 to 81.0 | 4.88% |
| 62.0 to 66.75 | 2.59% |

**Figure 1.** The in-depth relationships between VASFA and SF-36 percent values as a point-to-point analysis.

28XJ, outperformed the general capabilities of linear correlation analysis by providing more detailed insights through its nonlinear correlation analysis. Saarinen et al[3] reported the minimal important change (MIC) in VASFA using an anchor-based predictive method. In contrast, model 28XJ surpassed manual analysis by uncovering nonlinear relationships between VASFA and SF-36 percent values within the identified zones of interest. The results are shown and plotted in a meticulous table and nonlinear regression curves, respectively. The zones of interest curve indicated that VASFA did not correlate with SF-36 in a linear trend but showed a complex nonlinear pattern. In addition, it delineated an unknown zone of interest as the borderline zone, which was the zone of inconsistent trends of values, such as upward and downward trends. This was highly beneficial for foot and ankle surgeons, as well as other healthcare professionals, in using this zone of interest curve to predict the treatment outcomes related to patients' health-related quality of life. By using VASFA as a key indicator, they could better anticipate whether patients would fall into unsatisfactory or satisfactory outcome zones. In addition, our external validation ensured that our model 28XJ could be used in a real-world setting to predict the SF-36. This method is consistent with a previous study[8] on ML's validity to predict treatment outcomes.

Several previous studies[9-11] have proposed the benefits of AI for outcome prediction under various conditions. An interesting study[12] previously demonstrated that machine learning methods could predict functional outcomes with treatment goal setting at the beginning of treatment. Another study[13] showed that machine learning can accurately predict short-term adverse outcomes following open reduction and internal fixation of ankle fracture. Our ML was consistent with the aforementioned studies in predicting the quality-of-life outcomes following foot and ankle disabilities using patient parameters, especially VASFA and sex.

Our study had some limitations. First of all, our first model (model 28XJ) had potential overfitting. However, we created a second model, FK13, with another algorithm and dropped off a significant factor (SF_Total) to compare the performance to the first model and correct overfitting, respectively. The results of model FK13 showed great performance that could be evidence to support our overall ML platform for further research and broader application. Secondly, we performed external validation using a limited number of data units.

## Conclusions

Our ML platform (models 28XJ and FK13), based on a web-based ML algorithm platform, predicted an excellent relationship among VASFA, with or without SF-36 total score, and overall SF-36 percentage values supported by external validation. It outperformed manual investigation in terms of in-depth correlation analysis, especially nonlinear predictions, as the zones of interest between the VASFA and SF-36 scores. This ML platform needs further validation studies with a larger scale of data to augment the models' robustness for broader applications.

### Conflict of Interest

The authors declare they have no conflict of interest.

### Ethics Approval

The institutional Ethics Committee of the Faculty of Medicine, Thammasat University approved this study to collect and analyze the data from the patients (approval number: MTU-OT-9-CR079-079/53) on December 9th, 2016.

## Informed Consent

All study participants provided informed consent for data collection in this study.

## AI Disclosure

We disclose that no AI or assisted technologies were used in the drafting of our manuscript (abstract, text, and table). We used an AI platform (ObviouslyAI, available at: https://www.obviously.ai/) to analyze our dataset.

## Funding

None.

## Authors' Contributions

C.A.: study design, data collection, statistical analysis, data interpretation, and manuscript preparation. P.R.: manuscript preparation and literature search. W.A.: manuscript preparation.

## ORCID ID

Chayanin Angthong: 0000-0002-1104-8945
Prasit Rajbhandari: 0000-0002-3954-2435
Wirana Angthong: 0009-0002-9339-563X

## Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

## References

1) Angthong C, Chernchujit B, Suntharapa T, Harnroongroj T. Visual analogue scale foot and ankle: validity and reliability of Thai version of the new outcome score in subjective form. J Med Assoc Thai 2011; 94: 952-957.
2) Richter M, Zech S, Geerling J, Frink M, Knobloch K, Krettek C. A new foot and ankle outcome score: Questionnaire based, subjective, visual-analogue-scale, validated and computerized. Foot Ankle Surg 2006; 12: 191-199.
3) Saarinen AJ, Uimonen MM, Sandelin H, Toom A, Richter M, Repo JP. Minimal important change for the visual analogue scale foot and ankle (VAS-FA). Foot Ankle Surg 2021; 27: 196-200.
4) Angthong C. Health-related quality of life and functional outcomes in ankle arthritis patients based on treating with and without total ankle replacement surgery. J Med Assoc Thai 2015; 98 Suppl 2: S131-S137.
5) Stüber J, Zech S, Bay R, Qazzaz A, Richter M. Normative data of the Visual Analogue Scale Foot and Ankle (VAS FA) for pathological conditions. Foot Ankle Surg 2011; 17: 166-172.
6) Pongsakonpruttikul N, Angthong C, Kittichai V, Chuwongin S, Puengpipattrakul P, Thongpat P, Boonsang S, Tongloy T. Artificial intelligence assistance in radiographic detection and classification of knee osteoarthritis and its severity: a cross-sectional diagnostic study. Eur Rev Med Pharmacol Sci 2022; 26: 1549-1558.
7) Angthong C, Rungrattanawilai N, Pundee C. Artificial intelligence assistance in deciding management strategies for polytrauma and trauma patients. Pol Przegl Chir 2023; 96: 114-117.
8) Martin RK, Wastvedt S, Pareek A, Persson A, Visnes H, Fenstad AN, Moatshe G, Wolfson J, Lind M, Engebretsen L. Machine learning algorithm to predict anterior cruciate ligament revision demonstrates external validity. Knee Surg Sports Traumatol Arthrosc 2022; 30: 368-375.
9) Shamout F, Zhu T, Clifton DA. Machine learning for clinical outcome prediction. IEEE Rev Biomed Eng 2021; 14: 116-126.
10) Tunthanathip T, Oearsakul T, Chin J Traumatol. Application of machine learning to predict the outcome of pediatric traumatic brain injury. Chin J Traumatol 2021; 24: 350-355.
11) Christie SA, Conroy AS, Callcut RA, Hubbard AE, Cohen MJ. Dynamic multi-outcome prediction after injury: Applying adaptive machine learning for precision medicine in trauma. PLoS One 2019; 14: e0213836.
12) Kato C, Uemura O, Sato Y, Tsuji T. Functional outcome prediction after spinal cord injury using ensemble machine learning. Arch Phys Med Rehabil 2024; 105: 95-100.
13) Merrill RK, Ferrandino RM, Hoffman R, Shaffer GW, Ndu A. Machine learning accurately predicts short-term outcomes following open reduction and internal fixation of ankle fractures. J Foot Ankle Surg 2019; 58: 410-416.