

# Evaluating Usability of IRAIVI Pregnancy Prediction Model Using System Usability Scale

Rahul Shrivastava<sup>1\*</sup>, Manmohan Singhal<sup>2</sup>, Ashish Joshi<sup>3</sup>, Nivedita Mishra<sup>4</sup>, Neeraj Kumar Kashyap<sup>5</sup>, Rohit Shrivastava<sup>6</sup>, Eshwar Sai Tipirisetty<sup>7</sup>, Avisha Soni<sup>8</sup>

<sup>1</sup>School of Pharmaceutical & Population Health Informatics, Faculty of Pharmacy, DIT University, Dehradun; Ministry of Health & Family Welfare, Nirman Bhawan, New Delhi, India

<sup>2</sup>School of Pharmaceutical & Population Health Informatics, Faculty of Pharmacy, DIT University, Dehradun, India

<sup>3</sup>The University of Memphis, The University of Texas Health Science Center at Houston, Memphis, Tennessee, USA

<sup>4</sup>Sambodhi Research & Communication Pvt Ltd, Noida, India

<sup>5</sup>The INCLEN Trust International, New Delhi, India

<sup>6</sup>Ekdanta Dental Care, Bhopal, India

<sup>7</sup>Indian Institute of Public Health, Gandhinagar, India

<sup>8</sup>Indian Institute of Public Health, Hyderabad, India

DOI: 10.55489/njcm.141220233423

## ABSTRACT

**Background:** Barriers to utilize maternal healthcare services amongst pregnant women at community level differs substantial variation among slums and urban areas which is essential to recognize and resolve these issues within framework of district-level policy planning. IRAIVI pregnancy prediction model is developed to enhance optimal utilization of maternal healthcare services. The study was conducted to evaluate usability of IRAIVI using System Usability Scale amongst healthcare workforce.

**Methods:** This model was developed with a set of predictors based on data collected during baseline and follow up visits from currently pregnant women as per study protocol. For evaluating efficiency of this model, System Usability Scale (SUS) was adopted and shared with 25 randomly selected experts in the field of public health. The questionnaire was shared via google form and responses recorded on Likert's scale were then analysed using SPSS.

**Results:** IRAIVI model's usability assessed using SUS found to be user-friendly, best learning curve, adaptable to new system and highly acceptance by healthcare workforce. The SUS score averaging at 84 demonstrates favourable usability of the model.

**Conclusion:** This model has capability to accentuate maternal health services which in-turn can contribute for better ANC services in resource-constrained settings. Additionally, future opportunities can be explored through field studies in different settings.

**Keywords:** IRAIVI, SUS, usability, urban slum, prediction, model, citizen centric

## ARTICLE INFO

**Financial Support:** None declared

**Conflict of Interest:** None declared

**Received:** 03-10-2023, **Accepted:** 07-11-2023, **Published:** 01-12-2023

**\*Correspondence:** Dr. Rahul Shrivastava (Email: rahul.shrivastavamph14@gmail.com)

**How to cite this article:** Shrivastava R, Singhal M, Joshi A, Mishra N, Neeraj Kumar NK, Shrivastava R, Tipirisetty ES, Soni A. Evaluating Usability of IRAIVI Pregnancy Prediction Model Using System Usability Scale. Natl J Community Med 2023;14(12):827-833. DOI: 10.55489/njcm.141220233423

**Copy Right:** The Authors retain the copyrights of this article, with first publication rights granted to Medsci Publications.

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Share Alike (CC BY-SA) 4.0 License, which allows others to remix, adapt, and build upon the work commercially, as long as appropriate credit is given, and the new creations are licensed under the identical terms.

www.njcmindia.com | pISSN09763325 | eISSN22296816 | Published by Medsci Publications

## INTRODUCTION

Pregnant women in urban slums often face challenges due to dynamic settings and limited resources. The barriers to utilize the maternal healthcare service at confined level differs significantly between slum located in rural and urban areas that need to be known, and addressed at district level health policy planning.<sup>1</sup> This has led to increased maternal mortality and complications during the pregnancy followed by neonatal deaths.<sup>2</sup>

A machine learning (ML) based prediction model is a new field in the population health informatics where it can be used to predict future health outcomes. Machine learning algorithms are currently used widely in clinical subjects to develop models which can predict clinical-based probable events.<sup>3</sup> The majority of models were developed and assessed using historical data, whereas only a limited number underwent evaluation within a clinical setup but none of them reported to be used for urban slum communities.

Prediction model named as IRAIVI (Improved maternal health outcomes, Reduced maternal mortality, Accessible maternal health services, Informed and empowered mothers, Vocal and engaged communities, and Impactful interventions) was developed by a group of public health researcher to address these challenges by predicting desirable pregnancy outcomes (normal birth outcome without any complications) and facilitating effective antenatal care services uptake in urban slums of South Delhi as per the study protocol.<sup>4</sup> The model is developed by analysing different sets of datapoint patterns from the data fed (baseline and follow up) from the study.

Machine learning algorithm is an important component of predictive analytics, using the real-time as well as historical data to forecast any activity/behaviour and various spatiotemporal trends in correlation with other associated factors. This has great advantage in healthcare in predicting disease outcomes, prognosis, drug sales, prevalence of diseases, etcetera.<sup>5</sup>

As per literature available on related subject, the prediction models are categorised in many ways. Further, for research purpose, multiple variants of model were combined as per requirement to get the desired results.<sup>6</sup>

Once the datapoints are collected from the community and gathered in the system, it is time to select the right model for prediction. Furthermore, linear regression is among the most basic predictive models, relying on a pair of correlated variables: one acting as an independent variable and the other as a dependent variable. After development of any model, it needs evaluation for further adjustment for validating its accuracy.

The System Usability Scale (SUS) is a cost-effective yet efficient method for assessing a product's usability which has not limited till websites/webpages,

mobile phones, 2-way audio response systems, OTG apps, and more. It provides an easy-to-understand score from 1 (strongly disagree) to score 5 (strongly agree).

There is dearth of data on any tool available for usability assessment of artificial intelligence- based prediction model on healthy pregnancy outcomes. The objective of this research is to evaluate the usability of IRAIVI prediction model using System Usability Scale (SUS) amongst healthcare workforce before collecting the data for its validation on field.

## METHODOLOGY

The developed model predicting the desirable birth outcomes (normal birth outcome without any complication) was developed with a set of predictors based on the data collected during baseline and follow up visits from the target population (pregnant women aged between 18-44 years) of urban slums in South Delhi following the study protocol published in JMIR Research Protocol.<sup>4</sup> Though the original sample size was 225, during the data cleaning and curation process, 202 datapoints were used to develop this model.

Predictors like exposure to social media on maternal healthcare program, family planning practices, child order, weight and height of pregnant women, and any morbidity associated during pregnancy – diabetes mellitus, hypertension<sup>7</sup>, thyroidism<sup>8,9</sup> were key variables which were used to predict the pregnancy outcomes from pregnant women. The algorithm choice was based on data complexity and prediction task. An ensemble of K-Nearest Neighbours Algorithm (kNN) and Decision Tree using OR logic was selected due to its robust performance.

In evaluating the performance of the model statistically, F1 score employed as the evaluation metric, taking into account both precision and recall, offering a well-rounded assessment of the model's capacity to accurately categorize positive and negative cases. This metric is well-suited for situations with class imbalance, as it accounts for false positives and false negatives, offering a comprehensive evaluation of the model's predictive power (Table-1).

Overall, the described approach, incorporating binary tree knowledge assessment and ensemble models with varied classifiers, feature reduction, class imbalance handling, and F1 score evaluation, ensures a comprehensive and accurate prediction of birth outcomes and assessment of pregnant women's knowledge about Maternal health Services and programmes (Figure-1).

On piloting the IRAIVI prediction model on field, the features provide knowledge suggestions, predicts the mode of delivery, and notes the potential for complications and low birth weight in a group of 40 participants when validated.

**Table 1: Model parameters for different birth outcome predictors. The definition of each model is available in the Machine Learning algorithm subsection**

	Mode of delivery	Birth weight	Delivery complications
<b>Predictors</b>	1) Weight of pregnant women 2) Family planning adopted? 3) Gestational period 4) Mother educated after school? 5) First child?	1) Weight of pregnant women 2) BMI 3) Gestational period	1) Gestational period 2) Mode of delivery 3) First child? 4) Comorbidities present?
<b>Dimensionality reduction for KNN using PCA</b>	Yes	No	No
<b>Dimensionality reduction for decision tree using PCA</b>	Yes	Yes	No
<b>No. of neighbours</b>	5	7	3
<b>Class weight</b>	1:1	1:1.25	1:1
<b>Maximum depth</b>	7	5	3
<b>Minimum samples per leaf</b>	7	10	5
<b>Ensemble selection method</b>	OR logic	OR logic	AND logic
<b>Outcome variable</b>	1: Vaginal pregnancy 0: Caesarean pregnancy	1: Normal weight 0: Underweight	1: No delivery complications 0: Delivery complications
<b>Class distribution</b>	138:64	170:30	157:45
<b>Class imbalance addressed using smote?</b>	Yes	Yes	Yes

KNN= k-nearest neighbours' algorithm is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point; PCA=Principal Component Analysis for reducing redundancy information and extracting essential features; BMI=Body Mass Index; OR=Odds Ratio

**Figure 1: IRAIVI Pregnancy Prediction Model**

The birth outcome predicted from the model versus the actual with empirical comparison based on the number of ANC visits to their nearest healthcare centres is found to be using correlation coefficient (Phi coefficient) between the number of ANC visits and the predicted vs. actual delivery outcomes was approximately  $-0.1565$  that indicates a weak negative correlation between the two variables. However, accuracy for each group found to be (a) Less than 3 ANC visit as 83.3%; at least 4 ANC visit as 100%; and more than 4 ANC visits as 75% based on the follow-

ing data (Table-2). The overall accuracy was considered as weighted average of these group accuracies based on the number of cases in each group, hence, the overall Accuracy found to be 88.57% suggestive of high level of accuracy in this specific context and future public health research.

Further to evaluated the usability of this model, System Usability Scale (SUS) questionnaire was adopted and shared with 25 randomly selected subject experts (Public Health Specialists, Paediatricians, Gynaecologist & Obstetrician, Public Health Nurses, An-

ganwadi Workers) from known contact. SUS offers a valid, efficient, and dependable method for assessing the usability of this model. It serves as a widely used survey questionnaire for evaluating any system's usability. In the context of our study, it functions as a model incorporating a set of Likert-scale queries designed to gauge users' perceptions about the system's usability with pregnancy outcome predications using ten questions extracted directly from the SUS survey (Figure-2), along with their original statements. Each of these questions represents a statement to which stakeholders express their agreement level using a Likert scale, typically ranging from

"Strongly Disagree" to "Strongly Agree." These questionnaires were shared via google form and the responses were then analysed using SPSS.

**Table 2: Comparing results from Predicted vs Actual Birth Outcomes**

No. of ANC visits	Predicted		Actual	
	Normal	Caesarean	Normal	Caesarean
< 3 visits	12	1	10	3
At least 4 visits	15	2	15	2
> 4 visits	8	2	6	4

Sr. No.	SUS Questions for prediction model	Strongly disagree					Strongly agree				
		1	2	3	4	5	1	2	3	4	5
Q1	I think that I would like to use this system (model) frequently	1	2	3	4	5					
Q2	I found the system (model) unnecessarily complex	1	2	3	4	5					
Q3	I thought the system (model) was easy to use	1	2	3	4	5					
Q4	I think that I would need the support of a technical person to be able to use this system (model)	1	2	3	4	5					
Q5	I found the various functions in this system (model) were well integrated	1	2	3	4	5					
Q6	I thought there was too much inconsistency in this system (model)	1	2	3	4	5					
Q7	I would imagine that most people would learn to use this system (model) very quickly	1	2	3	4	5					
Q8	I found the system (model) very cumbersome to use	1	2	3	4	5					
Q9	I felt very confident using the system (model)	1	2	3	4	5					
Q10	I needed to learn a lot of things before I could get going with this system (model)	1	2	3	4	5					

Source: © Digital Equipment Corporation, 1986.

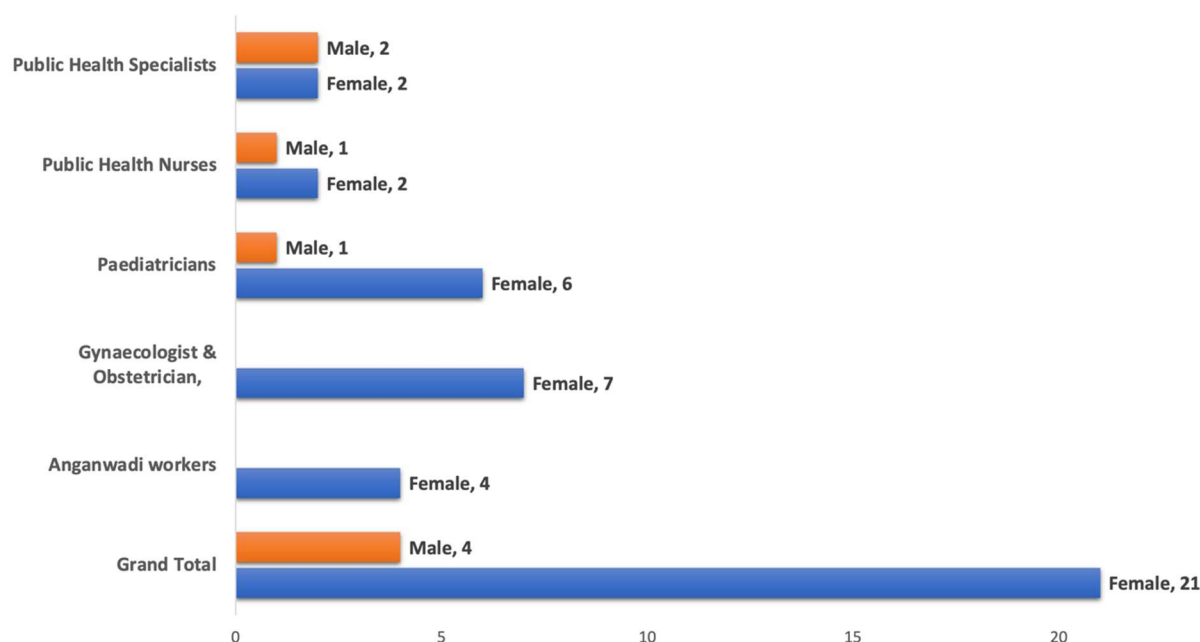
SUS=System Usability Score

**Figure 2: System Usability Scale used for evaluating IRAIVI.**

## RESULTS

This section showcases the analysis of the System Usability Scale (SUS) data in evaluation of usability of the IRAIVI pregnancy prediction model with healthcare workforce as mentioned in methodology (Figure-2). The participants reported as Public Health Specialists (3 females, 2 males), Paediatricians (5 females, 4 males), Gynaecologists & Obstetricians (4 females) Anganwadi Workers (4 females) and Public Health Nurses (2 females, 2 males) respectively. In Figure-3, the data indicates the following distribution across professions and genders out of 25 (21 females and 4 males).

The table-3 indicates that the questions number 1, 3, 5, and 7 received relatively high average scores, indicating that users generally found the system appealing, easy to use, well-integrated, and quick to learn. Questions number 2, 4, 6, and 8 received lower average scores, suggesting that users perceived the system as less complex, requiring less technical support, less cumbersome to use. Questions number 9 and 10 received an average score below 4 and above 2 respectively, suggesting that users were somewhat confident using the model and users felt they needed to learn quite a bit before getting comfortable with the system.



**Figure 3: Frequency distribution of respondents for SUS**

The average scores for each question in the SUS assessment scores reflect user's perceptions of the evaluated model's usability. Questions number Q1, Q3, and Q5 received high average scores (4.44, 4.72, and 4.44 respectively), indicating positive user perceptions regarding the system's frequent use, ease of use, and integration of functions. Whereas question numbers Q2, Q4, Q6, and Q8 obtained notably low average scores (1.76, 1.56, 1.48, and 1.16 respectively). This suggests that users found the system to be less complex, requiring less technical support, exhibiting consistencies, and less cumbersome to use. Question number Q7 achieved a high average score of 4.8, indicating that majority of users believed that they would quickly learn how to use this model. However, question numbers Q9 and Q10 received an average score of 3.84 and 2.68 respectively, indicating moderate user confidence in using the model, suggesting that participants felt a moderate need for learning before becoming comfortable with the model. Thus, usability assessment highlighted user acceptance and ease of use. The overall raw SUS score calculated by summing all scores from the respondents and multiplying each by 2.5. The final average SUS score is 84.

**Table 3: Average score for each SUS question**

Questions	Average scores
Q1	4.44
Q2	1.76
Q3	4.72
Q4	1.56
Q5	4.44
Q6	1.48
Q7	4.8
Q8	1.16
Q9	3.84
Q10	2.68
<b>SUS score</b>	<b>84</b>

## DISCUSSION

The IRAIVI prediction model's usability was assessed using the System Usability Scale (SUS). Healthcare workforce evaluated the model's user-friendliness, learning curve, adaptability to new system and its acceptance. The average SUS score of 84 indicated good usability of the model.

The overall SUS score of IRAIVI model considering the responses of all 25 participants, represents a good turn of usability.<sup>10,11</sup> While assessing, the influence of demographics on the SUS rates, the gender of participants and their familiarity with predictive tools showed significant influence. Females were having higher SUS rate (84%) in comparison to male participants (16%). Public health specialists (16%), paediatricians (12%), gynaecologist & obstetrician (28%), public health nurses, (12%) and Anganwadi workers (16%) found to be responding to the SUS questionnaire from public health cadre.

The chosen citizen centric ensemble model demonstrated the highest F1 score (84%), accuracy (75%), and precision (85%), indicating accurate birth complication prediction. Decision tree analysis revealed meaningful relationships between input features and outcomes, highlighting factors affecting birth complications. Due to the reduced sample size, the model was validated using the leave-one-out cross-validation technique<sup>12</sup>. In this validation approach, each data point was sequentially left out as a test set while the rest of the data was used for training. This process was repeated for each data point, ensuring that all data points must use for both training and testing. Leave-one-out cross-validation is particularly suitable for small datasets, allowing for a more reliable assessment of model performance.<sup>12</sup>

Initially such system was coined as a "rapid and basic

usability scale," that has demonstrated its speed without any drawbacks. However, in current scenario, it is probable that the SUS will remain a widely used method for assessing perceived usability of users in the coming years. Whenever researchers and/or professionals require a gauge for perceived usability for any system, it is highly advisable to contemplate employing the SUS.<sup>13,14</sup>

An application of precision techniques to predict progression to medical conditions (diabetes mellitus) provides preliminary improvement in prediction.<sup>6,15</sup> Understanding the mechanics of predictive modelling will help in troubleshooting and improving the performance of the model.<sup>16,17</sup>

In this model, it enables predicting the probability by forecasting the future birth outcomes by facilitate WHAT-IF analysis. This has used the primary data to identify the trends/patterns of the desired outcomes and use predictors to predict the outcomes. Even, there is also no such study that has investigated the depth of technicalities of these results that could be presented to healthcare professionals.<sup>8</sup> Extra caution was followed as the feature engineering is a crucial aspect of machine learning, where careful selection and transformation of features can significantly impact the model's performance. Hence, properly chosen features contributed to the model's ability to generalize well to new data and make accurate predictions to some extent. On the other hand, irrelevant or noisy features leading to suboptimal performance or even overfitting were avoided, where the model memorizes the training data but fails to generalize to unseen data.

In summary, features are the independent variables that represent various aspects of the data in a prediction model. They serve as the feed to the model and help it understand patterns, associations and relationships in the data to predict or classify new instances as per desired objectives. The choice and engineering of features were essential for building robust and high-performing machine learning models that can make accurate predictions and generalizations in real-world scenarios with respect to urban slum settings.

The overall SUS score of 84 signifies the aggregated assessment of the model's usability. While the system demonstrates several positive aspects to be celebrated, areas requiring enhancement are also evident. These insights serve as a valuable guide for refining the model's usability, ensuring a more streamlined and user-friendly experience. Hence scoring by SUS indicates that the IRAIVI prediction model has a reasonably good level of usability.

Limitation associated with this study is that being a prospective cross-sectional study conducted among pregnant women, the developed model had resulted a good acceptability and usability based on above results. However, due to small sample size, the results cannot be generalised in another geography/group.

## CONCLUSION

The IRAIVI prediction model offers a robust solution for predicting desirable pregnancy outcomes in urban slums. By combining machine learning methods and ensuring user-friendliness, this model has the potential to enhance maternal health services and contribute to better antenatal care in resource-constrained settings. Further analysis and interpretation of usability assessment results can provide insights for model improvement. Additionally, the model's implementation and impact on maternal health services uptake should be explored through field studies.

## REFERENCES

1. Khan Z, Mehnaz S, Siddiqui AR, Ansari A, Khalil S, Sachdeva S. All Slums are Not Equal: Maternal Health Conditions Among Two Urban Slum Dwellers. *Indian J Community Med.* 2012 Jan;37(1):50-6. doi: 10.4103/0970-0218.94027. PMID: 22529541; PMCID: PMC3326809.
2. Institute of Medicine (US) Committee on Improving Birth Outcomes; Bale JR, Stoll BJ, Lucas AO, editors. *Improving Birth Outcomes: Meeting the Challenge in the Developing World.* Washington (DC): National Academies Press (US); 2003. 2, Reducing Maternal Mortality and Morbidity. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK222105/> (Last accessed on 31 August 2023)
3. Sun H, Depraetere K, Meesseman L, Cabanillas Silva P, Szymanowsky R, Fliegenschmidt J, Hulde N, von Dossow V, Vanbiervliet M, De Baerdemaeker J, Roccaro-Waldmeyer DM, Stieg J, Domínguez Hidalgo M, Dahlweid FM. Machine Learning-Based Prediction Models for Different Clinical Risks in Different Hospitals: Evaluation of Live Performance. *J Med Internet Res.* 2022 Jun 7;24(6):e34295. doi: 10.2196/34295. PMID: 35502887; PMCID: PMC9214618.
4. Shrivastava R, Singhal M, Gupta M, Joshi A. Development of an Artificial Intelligence-Guided Citizen-Centric Predictive Model for the Uptake of Maternal Health Services Among Pregnant Women Living in Urban Slum Settings in India: Protocol for a Cross-sectional Study With a Mixed Methods Design. *JMIR Res Protoc.* 2023 Jan 27;12:e35452. doi: 10.2196/35452. PMID: 36705968; PMCID: PMC9919485.
5. Slobbe LCJ, Füssenich K, Wong A, Boshuizen HC, Nielen MMJ, Polder JJ, Feenstra TL, van Oers HAM. Estimating disease prevalence from drug utilization data using the Random Forest algorithm. *Eur J Public Health.* 2019 Aug 1;29(4):615-621. doi: 10.1093/eurpub/cky270. PMID: 30608539; PMCID: PMC6660107.
6. Jacobsen LM, Larsson HE, Tamura RN, Vehik K, Clasen J, Sosenko J, Hagopian WA, She JX, Steck AK, Rewers M, Simell O, Toppari J, Veijola R, Ziegler AG, Krischer JP, Akolkar B, Haller MJ; TEDDY Study Group. Predicting progression to type 1 diabetes from ages 3 to 6 in islet autoantibody positive TEDDY children. *Pediatr Diabetes.* 2019 May;20(3):263-270. doi: 10.1111/pedi.12812. Epub 2019 Jan 29. PMID: 30628751; PMCID: PMC6456374.
7. AlKaabi LA, Ahmed LS, Al Attiyah MF, Abdel-Rahman ME. Predicting hypertension using machine learning: Findings from Qatar Biobank Study. *PLoS One.* 2020 Oct 16;15(10):e0240370. doi: 10.1371/journal.pone.0240370.
8. Weng SF, Vaz L, Qureshi N, Kai J. Prediction of premature all-cause mortality: A prospective general population cohort study comparing machine-learning and standard epidemiological approaches. *PLoS One.* 2019 Mar 27;14(3):e0214365. doi:

- 10.1371/journal.pone.0214365. PMID: 30917171; PMCID: PMC6436798.
9. Chaganti R, Rustam F, De La Torre Díez I, Mazón JLV, Rodríguez CL, Ashraf I. Thyroid Disease Prediction Using Selective Features and Machine Learning Techniques. *Cancers (Basel)*. 2022 Aug 13;14(16):3914. doi: 10.3390/cancers14163914. PMID: 36010907; PMCID: PMC9405591.
  10. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. *J Usability Stud*. 2009;4(3):114–23. <https://uxpajournal.org/determining-what-individual-sus-scores-mean-adding-an-adjective-rating-scale/>
  11. Kortum PT, Bangor A. Usability ratings for everyday products measured with the system usability scale. *Int J Hum-Comput Interact*. 2013 Jan;29(2):67–76. doi: 10.1080/10447318.2012.681221
  12. Geroldinger, A., Lusa, L., Nold, M. *et al.* Leave-one-out cross-validation, penalization, and differential bias of some prediction model performance measures—a simulation study. *Diagn Progn Res* 7, 9 (2023). <https://doi.org/10.1186/s41512-023-00146-0>
  13. James R. Lewis (2018) The System Usability Scale: Past, Present, and Future, *International Journal of Human-Computer Interaction*, 34:7, 577-590, DOI: 10.1080/10447318.2018.1455307
  14. Khalifa M, Magrabi F, Gallego Luxan B. Evaluating the Impact of the Grading and Assessment of Predictive Tools Framework on Clinicians and Health Care Professionals' Decisions in Selecting Clinical Predictive Tools: Randomized Controlled Trial. *J Med Internet Res*. 2020 Jul 9;22(7):e15770. doi: 10.2196/15770. PMID: 32673228; PMCID: PMC7381257.
  15. Batko K, Ślęzak A. The use of Big Data Analytics in healthcare. *J Big Data*. 2022;9(1):3. doi: 10.1186/s40537-021-00553-4. Epub 2022 Jan 6. PMID: 35013701; PMCID: PMC8733917.
  16. Usability.Gov System Usability Scale (SUS). 2020. Available online: <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html> (last accessed on 30 August 2023).
  17. Wichmann RM, Fagundes TP, de Oliveira TA, Batista AFdM, Chiavegatto Filho ADP (2022) Physician preference for receiving machine learning predictive results: A cross-sectional multicentric study. *PLOS ONE* 17(12):e0278397. <https://doi.org/10.1371/journal.pone.0278397>