Evaluating Usability of IRAIVI Pregnancy Prediction Model Using System Usability Scale

Rahul Shrivastava¹, Mannmohan Singhal², Ashish Joshi³, Nivedita Mishra⁴, Neeraj Kumar Kashyap⁵, Rohit Shrivastava⁶, Eshwar Sai Tipirisetty⁷, Avisha Soni⁸

¹School of Pharmaceutical & Population Health Informatics, Faculty of Pharmacy, DIT University, Dehradun; Ministry of Health & Family Welfare, Nirman Bhawan, New Delhi, India
²School of Pharmaceutical & Population Health Informatics, Faculty of Pharmacy, DIT University, Dehradun, India
³The University of Memphis, The University of Texas Health Science Center at Houston, Memphis, Tennessee, USA
⁴Sambodhi Research & Communication Pvt Ltd, Noida, India
⁵The INCLEN Trust International, New Delhi, India
⁶Ekdanta Dental Care, Bhopal, India
⁷Indian Institute of Public Health, Gandhinagar, India
⁸Indian Institute of Public Health, Hyderabad, India

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)


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.
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. This has led to increased maternal mortality and complications during the pregnancy followed by neonatal deaths.

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. 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. 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.

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.

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. 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, thyroidism 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

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Mode of delivery</th>
<th>Birth weight</th>
<th>Delivery complications</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1) Weight of pregnant women</td>
<td>1) Weight of pregnant women</td>
<td>1) Gestational period</td>
</tr>
<tr>
<td></td>
<td>2) Family planning adopted?</td>
<td>2) Gestational period</td>
<td>2) Mode of delivery</td>
</tr>
<tr>
<td></td>
<td>3) Gestational period</td>
<td>2) BMI</td>
<td>3) First child?</td>
</tr>
<tr>
<td></td>
<td>4) Mother educated after school?</td>
<td>3) Gestational period</td>
<td>4) Comorbidities present?</td>
</tr>
<tr>
<td></td>
<td>5) First child?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Dimensionality reduction for KNN using PCA: Yes
- Dimensionality reduction for decision tree using PCA: Yes
- No. of neighbours: 5
- Class weight: 1:1
- Maximum depth: 7
- Minimum samples per leaf: 7
- Ensemble selection method: OR logic
- Outcome variable: 1: Vaginal pregnancy
- Class distribution: 138:64
- Class imbalance addressed using smote?: 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 following 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

<table>
<thead>
<tr>
<th>No. of ANC visits</th>
<th>Predicted Normal</th>
<th>Predicted Caesarean</th>
<th>Actual Normal</th>
<th>Actual Caesarean</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 3 visits</td>
<td>12</td>
<td>1</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>At least 4 visits</td>
<td>15</td>
<td>2</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>&gt; 4 visits</td>
<td>8</td>
<td>2</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

**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) and 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.

---

[Source: © Digital Equipment Corporation, 1986.]
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

<table>
<thead>
<tr>
<th>Questions</th>
<th>Average scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>4.44</td>
</tr>
<tr>
<td>Q2</td>
<td>1.76</td>
</tr>
<tr>
<td>Q3</td>
<td>4.72</td>
</tr>
<tr>
<td>Q4</td>
<td>1.56</td>
</tr>
<tr>
<td>Q5</td>
<td>4.44</td>
</tr>
<tr>
<td>Q6</td>
<td>1.48</td>
</tr>
<tr>
<td>Q7</td>
<td>4.8</td>
</tr>
<tr>
<td>Q8</td>
<td>1.16</td>
</tr>
<tr>
<td>Q9</td>
<td>3.84</td>
</tr>
<tr>
<td>Q10</td>
<td>2.68</td>
</tr>
<tr>
<td><strong>SUS score</strong></td>
<td><strong>84</strong></td>
</tr>
</tbody>
</table>

### DISCUSSION

The IRAVII 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 IRAVII model considering the responses of all 25 participants, represents a good turn of usability. 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. 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. Initially such system was coined as a "rapid and basic
usability scale," that has demonstrated its speed without any drawbacks. However, in current scenarios, 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. \textsuperscript{13,14}

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

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.\textsuperscript{8} 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.

\section*{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.

\section*{REFERENCES}


