ICSSR Sponsored
ISSN: 2319-9997

Journal of Nehru Gram Bharati University, 2024; Vol. 13 (2):19-27

Proposed Model: Verification and Validation for disease prediction using data mining technique

Arvind K. Shukla*, Rohit Ramesh and Chandrakant K. Singh
Department Computer Application
Nehru Gram Bharati (Deemed to be University),
Prayagraj, India, Email: arvind.ecimt@gmail.com

Received: 25.10.2024 *Revised: 22.11.2024* *Accepted: 18.12.2024*

Abstract

Verification & Validation approaches are helpful in a development process that is model-based because they improve model comprehension and allow for the evaluation of model features that are provided implicitly inside the model. The objective of V&V is to evaluate models and explore alternative modeling methodologies. During verification, the developer identifies aspects crucial to the development process, while validation is mostly an activity conducted by the developer to demonstrate model attributes to the client. Verification is closely connected to testing and encompasses automated and semi-automatic (static) analytical methods like as model-checking and theorem proving

Keywords: *Model, Mining techniques, AHP Concept*

Introduction

This has led to substantial efforts in the development and implementation of validation and verification (V&V) techniques for models. Models generally facilitate the elucidation and strategizing of software development. Furthermore, they enable the implementation of systems that assess quality and identify deficiencies in conceptual concepts. However, since models are used so extensively throughout the process of developing software, there is a risk of errors being introduced into the models themselves. Validation and verification (V&V) techniques for models are crucial in identifying or mitigating the emergence of such defects.

1.1 AHP Validation Concept

The Analytical Hierarchy Process is a thorough decision-making process that organizes issues hierarchically and considers both quantitative and qualitative elements. The Analytical Hierarchy Process is recognized as one of the most effective approaches for deriving conclusions from many elements. Establishing a framework for prioritizing solutions is the first task. The last stage is assigning a

financial value to each level of the hierarchy and constructing a comparative matrix for like items. The first step of the decision-making process involves structuring the problem and desired outcome of the choice into a hierarchical relationship with the pertinent decision parts. Decision-making involves evaluating your possibilities, which are represented by the choices available to you. A Diagrammatic Representation of a Hierarchical Tree In the second stage of the procedure, in order to conduct out pair comparison, a questionnaire has to be produced and delivered to the respondents (who may be managers, experts, users, and so on) in order to collect their input. This feedback will be used to evaluate the effectiveness of the pair. It is important to note that each decision maker entered their desired amount for each member, and after that, individual judgements (of each respondent) were converted into group judgements (for each one of the pair comparisons) using their geometrical average.

This was done in order to ensure that the results were accurate. This is a crucial aspect of the conducted inquiry. A score of one indicates that the two characteristics are equivalent or possess the same degree of importance throughout this scale. This scale ranges from one to nine as it progresses from point to point. Conversely, the presence of the number nine in a paired matrix suggests that one component significantly outweighs the other in importance.

This Paper will focus on verifying and validating the given model and algorithms via the selection of diverse algorithm quantifications. The categorization of algorithms must be determined depending on the prognosis of the condition. Given that the issue pertains to the generation of several criteria, researchers use the Analytical Hierarchy Process (AHP) to validate both the model and their findings. The selection of a single characteristic from among many available options, the distribution of resources, and the production of projections are some examples of effective applications of AHP.

We need to break down the choice into its component parts so that we may arrive at a conclusion in an orderly fashion and provide a list of priorities as a result.

I. Clearly articulate the challenge at hand and specify the information that needs to be gathered.

II. Organize the decision hierarchy such that it descends from the highest level, which should include the decision's overarching purpose, down through the intermediate levels, which should contain the criteria on which future aspects should be based, and finally to the lowest level (which usually is a set of the alternatives).

III. Construct a collection of matrices for doing pairwise comparisons. A comparison is made between each element in the level directly below it and each element in the level above it for each element in the level above it.

IV. Apply the weights that were assigned to the priorities after doing the comparisons to the priorities that are on the level immediately below. Do this for each component. Sum together the weights for each component in the level below to determine its global priority. Continue weighing and adding until you determine the lowest-priority choice.

Since the selection of elements is a process that is tied to the construction, the application of AHP is a logical choice. The benefit of using this decision-support tool is that it allows you to reach a final position that is based on a paired pair assessment of both the criteria and the possibilities that the researcher has supplied. In addition, the AHP method is chosen since the rationale behind it is sound and can be easily comprehended, and the procedure of calculating its results is not too complicated.

The authors are eager to conduct an investigation into the sensitivity of alternative assessments with reference to the comparison of various algorithms. The construction element selection process is observed, and then improvements may be made because of this monitoring. The monitoring is done by altering only one element in a paired matrix.

1.2 Model Verification

The Analytic Hierarchy Process is a well recognized technique for determining the weight of criteria in decision-making (Gomez-Ruiz, 2010). This strategy has been used in several research projects and has been integrated with other techniques to address decision-making difficulties. The Analytic Hierarchy Process (AHP) technique does, however, have certain limitations when it comes to dealing with situations that include a high number of criteria and options (Yi-Chung, 2006). The decision-making process that will be used to choose the effective parameters (Precision, Recall, and Accuracy) using the AHP method will involve the problem being decomposed into machine algorithms. Consequently, several pairwise comparisons must be conducted to address the model reliance seen in tables 1.1 to

1.7 The Analytics Hierarchy Process (AHP) approach often fails to provide consistent comparative data, especially in scenarios including several support vector machines (SVM), tree networks (kNN), and k-nearest neighbors (kNN) (Accuracy, RMSE, and MSE). Table 1.1 presents the three models developed for assessing the consistency and eigenvalues of the model.

Step 1: Defined Model

Table 1.1 bjectives

SVM	**Model Value**
TREE	Model Value
Knn	Model Value

Table 1.2 Define Weight

	SVM	**kNN**	**Tree**
SVM	1	2.5	6.3
kNN	0.4	1	3.6
Tree	0.158730	0.277777	1

Table 1.3 Calculate Normalized Metrics

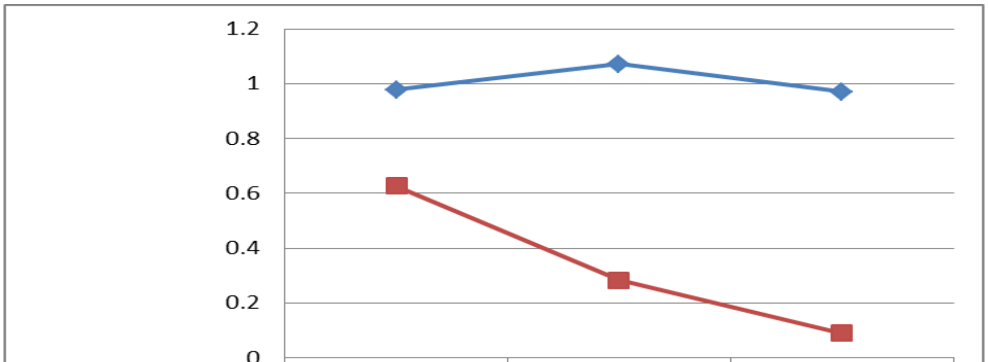
	SVM	**kNN**	**Tree**
SVM	0.6415	0.6618	0.5780
kNN	0.2566	0.2647	0.33.3
Tree	0.1018	0.0735	0.0917

Eigen Vector	**Priority Vector**
0.9775	0.6271
1.0724	0.2839
0.9705	0.0890
Principle Eigen Value (3.0203)	0.0102
	0.0175 and 0.0102= CI

In order to evaluate the models, we applied AHP rating scales for the criteria (see fig 1.1). Since SVM, Tree and kNN require maximization and range from 0 to 1(in this case from 0.6415 to 0.6618), we used the rating scale in table 1.3. As seen in Table 1.3, the kNN resulted the best model with an ideal priority Vector 1.0724, mainly because it was the highest evaluated in the two metrics: 0.9775 and 0.9705 in(Table 1.5). In table1.3, it is possible to see the average consistency index obtained from the output of the models in the test phase.

The consistency index (CI) can be calculated, using the following formula:

$$CI = \left[\frac{\lambda_{\max} - n}{n-1} \right] \quad \text{(Eq 1)}$$



	SVM	kNN	Tree
Eigen Vector	0.9775	1.0724	0.9705
Priority Vector	0.6271	0.2839	0.089

Fig. 1.1: Comparative Study

Where denotes the value that has the greatest impact on the matrix's variance. We calculate the mean of all data points, supposing this represents the maximum potential variation; thereafter, we determine the confidence interval and critical ratio to ensure consistency. We may ascertain the consistency of the judgments by using a statistic referred to as the consistency ratio (CR). Table 1.3 illustrates that the consistency ratio (CR) is calculated by dividing the confidence interval (CI) by the random index (RI), as shown in Equation (1).Table 1.3 allows for the assignment of weights derived from the outputs of the models in Chapter 4, which are compared to the inconsistency of the alternatives based on Accuracy, MSE, and RMSE. This may be calculated using a tree model that effectively contributes to accurate predictions.

Step 2: Defined Model Alternatives

After identifying the criteria and sub-criteria using the literature review methodology described in Table 1, The calculation phases of the Analytic Hierarchy Process (AHP) are used to establish a hierarchy of the consumer choice model pertinent to the online purchasing context. A comprehensive literature study resulted in the formulation of the following criteria and sub-criteria (Table 1). Equation (1) was used to establish homogeneity and then applied to Table 2 after data collection from field experts to ascertain weights for each criterion and sub-criterion shown in Table 1.4. The CR being less than 0.1 indicates that there are no concerns about the consistency of the data set. Equation

(1) was used to assess the consistency of the sub-criteria, with the resulting consistency ratio shown in Table 1.4.

Table 1. 4 Alternatives

Accuracy	Model Value
MSE	Model Value
RMSE	Model Value

Table 1.5 defines choice criteria as C = {Cj| j = 1, 2... n}. A (n×n) evaluation matrix may summarise data from pairwise comparison of n sub-criteria. This matrix's aij I j= 1, 2... n)elements indicate the criterion's weight quotient. A square matrix and a reciprocal matrix may be used to illustrate this pair-wise comparison. In this matrix, where aij = 1/aji, we would have (n n) matrices if all of the experts participated (see table 1.5).

Table 1. 5 Assign Weight

	Accurcy	**MSE**	**RMSE**
Accuracy	1	3.9	5.6
MSE	0.2564	1	3.2
RMSE	0.1785	0.3125	1
	1.435	5.213	9.800

In Table 1.6, the Accuracy resulted in the best value with an ideal priority Vector 0.6722, mainly because it was the highest evaluated in the two metrics: 0.6969 and 0.7482 (Table 1.6). Nonetheless, there are three alternatives for model that also stand out, and the RMSE came a close second with a score of 0.9645 (Table 1.6). Table 1.6 displays the average consistency index derived from the output of the alternatives throughout the testing phase. Due to the curse of dimensionality, the AHP may be used to evaluate possibilities among many models and substantiate the model's correctness. A novel strategy has been implemented to address the primary alternatives, with the specifics of this method detailed in Table 1.6, including the consistency index for validating the stage computations.

Table 1. 6 Normalised Metrics

	Accuracy	**MSE**	**RMSE**
Accuracy	0.6969	0.7482	0.5714
MSE	0.1787	0.1918	0.3265
RMSE	0.1244	0.0600	0.1020
Eigen Vector		**Priority**	
0.9645		0.6722	
1.2111		0.2324	
0.9357		0.0955	
Eigen Value 3.114		0.0557	

In table 1.7, we have calculated to overall priority of each criteria respect to model weight. We are observed that SVM values of very effective in this research work. Table 1.8 are given the finalize metrics in the form of accuracy RMSE and MSE context. Accuracy value is the maximum effective constraints which provides verifying and validated of research (see fig 1.2).

Table 1. 7 Calculate overall priority

	Tree (0.2839)	**kNN (0.089)**	**SVM (0.6271)**
Accuracy	0.6722	0.653	0.600
MSE	0.2324	0.251	0.200
RMSE	0.0955	0.096	0.200

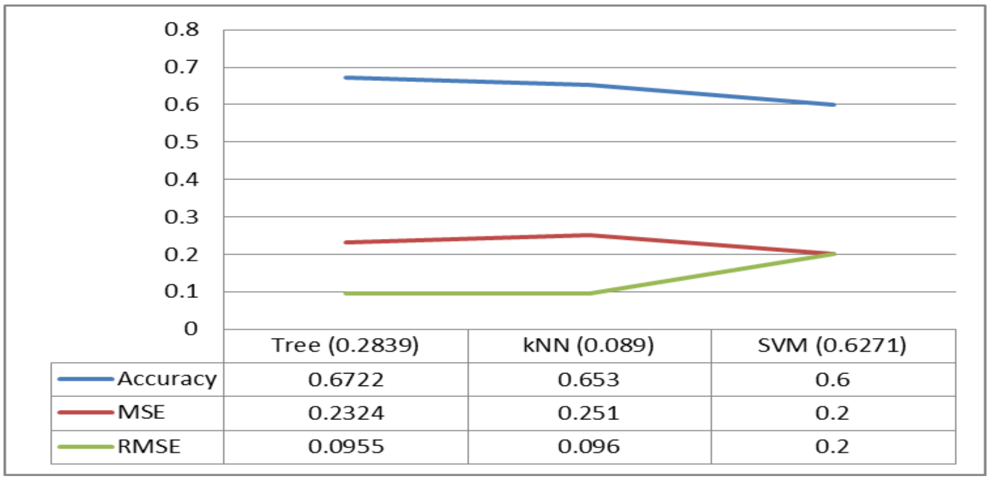


Fig 1.2: Evaluation Result of table 1.7

The table 1.8 presents the verbal evaluations that each of the three options received for each of the covering criteria. When we compare the findings from the pair-wise comparison approach known as a relative model to these findings from the ratings model as presented in Table 1.7, we find that the priority for the first two choices are quite near to one another. The latter two are a little bit distinct from one another. This should not come as a surprise.

The comparative model approach, which evaluates many solutions against various criteria, yields the most precise findings. The ratings method enables the evaluation of several possibilities rapidly, yielding findings that are adequately similar (see to figure 1.3).

Table 1. 8 Finalize Metrics

Accuracy	**0.6252**
MSE	0.2137
RMSE	0.1611
Highest Priority = Highest Score	

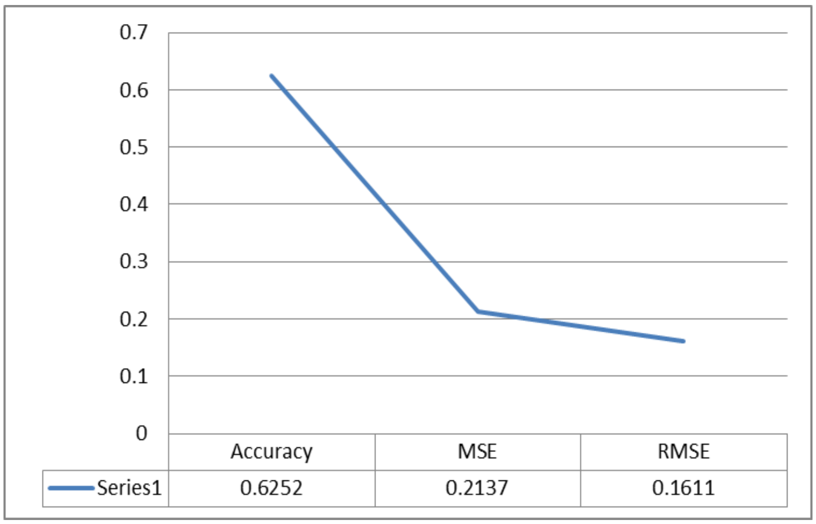


Fig 1.3: Finalize Impact

Conclusion

The models used in this work to rectify the inconsistency of the AHP pairwise comparison matrix are Tree, kNN, and SVM. These methodologies use machine learning instruments and are identified by their corresponding acronyms. Initially, simulations including training, validation, and testing are conducted to assess both methodologies. The SVM approach exhibits behavior similar to that of Trees
regarding CR reduction; yet, it demonstrates a superior accuracy rate in predicting previously unknown inputs provided to the network. Moreover, it has the advantage of a markedly accelerated convergence rate relative to the training pace of Tree. The ultimate weight established from the AHP tests is shown in Table 1.8. Upon comparison with the three models of the original input components. The subsequent observations and key points are as follows:

•An evaluation study of the Analytic Hierarchy Process (AHP) and MLP approach is also made while extracting the weights of criteria for Models and their criteria and alternatives.

•In this evaluation, we have first done to predict the disease through MLP algorithms with specific accuracy and prevention of breast cancer at an early stage. In the second stage, we accepted the AHP process to illustrates the model dependency variable such as precision, recall and accuracy.

• In table 1.8, We have found the index values of model accuracy in MSE, RMSE and accuracy and show that SVM is the most important model for disease prediction.

• One section of the Clarity concept has been completed, including the calculations for MSE, RMSE, and correctness according to the testability of table 1.5 (see to figure 1.3).

• In table 1.1, illustrates the weight according to the model, and we found that SVM is gained heights weight and supports table 1.3 results.

•In any observations, accuracy value is most important than other measurable value, and here accuracy is gained the highest score in table 1.8. Therefore, we can say that SVM is most important to other methods.

References:

A. K. Singh et al. (2020) Evaluating the Performance of Data Mining Models for Disease Prediction"

H. Abdi and L.J. Williams (2010). "Principal Component Analysis." Wiley Interdisciplinary Reviews: Computational Statistics.

J. Han et al. (2019) "Verification and Validation of Predictive Models in Data Mining"

J. Han, M. Kamber, and J. Pei (2012). Data Mining: Concepts and Techniques. Elsevier.

K. K. Singh et al. (2019) Data Mining Techniques for Disease Diagnosis and Prediction"

K. P. Murphy (2012). Machine Learning: A Probabilistic Perspective. MIT Press. Offers insights into machine learning algorithms that can be applied for disease prediction.

M. A. Wazed et al. (2019) Data Mining for Disease Diagnosis and Prediction: A Review"

R. C. T. Lee, A. L. M. McDonald (2019). "A systematic review of machine learning in cardiovascular disease prediction." Journal of Cardiovascular Medicine.

S. S. F. Al-Masri and A. M. A. Al-Shalabi (2020). "Data mining approaches for predicting diseases: A systematic review." International Journal of Data Mining and Knowledge Management Process.

S. S. Rao et al. (2020) A Survey on Data Mining Techniques for Disease Prediction"