

Development of a vaccine-hesitancy prediction instrument: Application of machine learning

Nour Hanah Othman^{1*}, Irraivan Elamvazuthi², Santhanathan Rajendram¹, Helvinder Kaur Balbir Singh³,
Muhammad Hadhrami Mohd Hussain⁴, Mohammed Tahir Ansari⁵, Karna Vishnu Vardhana Reddy⁶

¹Department of Pharmacy, Faculty of Pharmacy and Health Sciences, Royal College of Medicine Perak, Universiti Kuala Lumpur, Ipoh, Malaysia.

²Department of Electrical and Electronics Engineering, Institute of Health and Analytics, Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia.

³Department of Psychology, Faculty of Medicine, Royal College of Medicine Perak, Universiti Kuala Lumpur, Ipoh, Malaysia.

⁴Department of Clinical Medicine, Faculty of Medicine, Royal College of Medicine Perak, Universiti Kuala Lumpur, Ipoh, Malaysia.

⁵School of Pharmacy, Faculty of Science and Engineering, University of Nottingham Malaysia, Semenyih, Malaysia.

⁶Department of Electronics and Communication Engineering, Aditya University, Surampalem, India.

ARTICLE HISTORY

Received on: 25/10/2025

Accepted on: 21/02/2026

Available Online: 15/04/2026

Key words:

Vaccine hesitancy, childhood, vaccination, immunization, machine learning.

ABSTRACT

As childhood vaccination is vital for children to prevent them from vaccine preventable diseases, vaccine hesitancy (VH) is a phenomenon that can jeopardize this preventive mechanism. This study aims to develop an instrument to predict VH among parents towards childhood immunization by using machine learning (ML) algorithms. In this study, the approach of predicting VH was to focus on attitude, behavior and practice through the administration of a questionnaire which was verified by statistical analysis and ML algorithms. The researchers developed a 26-item instrument adapted from two other studies. Experts from three different fields reviewed the instrument for content validity. From the pilot study, a 13-item instrument was generated and has a Cronbach alpha value of 0.850 for reliability. The instrument was applied to 510 respondents who are parents attending the Obstetrics and Gynecology and Pediatric Clinics of the state referral hospital, and have children between the ages of 0 and 15 years old. The data collected was subjected to 10 ML algorithms. It was found that in terms of accuracy, the logistic regression with bagging method produced the best results with 99.02% for the hold-out method and 97.45% for the 10-fold cross-validation method. The results of our study show that there is potential of the instrument to anticipate parental VH in the local situation. The instrument's prospects can be further enhanced if its performance is validated against an objective parameter such as vaccination records.

1. INTRODUCTION

Childhood vaccination programs have played a vital role in reducing the global burden of infectious diseases. These programs typically begin at birth and continue through adolescence, protecting against a wide range of potentially life-threatening illnesses. Since 1974, vaccination has prevented 154 million deaths, including 146 million among children under

5 years old, of whom 101 million were infants under 1 year old [1].

Following the now-discredited 1998 paper authored by Dr. Andrew Wakefield on the link between vaccines and autism, the belief in vaccines began to decline, even after the paper was retracted 12 years later [2]. Despite the proven benefits of vaccination, some parents remain hesitant to immunize their children. Common concerns include vaccine safety, potential side effects, and misconceptions about the necessity of vaccines for diseases that are now rare in many developed countries [3]. However, public health experts emphasize that maintaining high vaccination rates is crucial for preserving herd immunity and preventing the resurgence of previously controlled diseases [4].

*Corresponding Author

Nour Hanah Othman, Department of Pharmacy, Faculty of Pharmacy and Health Sciences, Royal College of Medicine Perak, Universiti Kuala Lumpur, Ipoh, Malaysia. E-mail: nourhanah@unikl.edu.my

VH is the delay or refusal of vaccines even when vaccination services are available, which can be caused by false information, mistrust, or complacency [5]. VH can vary from general reluctance to specific concerns about vaccines and can be seen in different populations and contexts [5]. This is a common phenomenon and has been reported by more than 90% of countries globally [6]. In Malaysia, the National Immunization Program, which was developed in the 1950s under the Maternal and Child Health Programs of the Ministry of Health claimed a high immunization coverage of between 95% and 98% among children under the age of 2 [7]. The National Health Immunization Program also subsidizes vaccines for children aged <15 [8]. However, vaccination rates started to decline in the last 10 years, and from 2014 to 2018, there were outbreaks of measles [9,10] and diphtheria [11] in certain parts of the country. Data extracted from the 2016 National Health and Morbidity Survey showed that complete immunization coverage among children was 86.4% [12], which is unsatisfactory in achieving herd immunity according to the World Health Organization (WHO) standards.

Many tools have been developed to detect and measure VH. The most commonly used include the Parent Attitudes About Childhood Vaccines Survey (PACV) [13], and the Vaccine Hesitancy Scale (VHS), developed by the WHO Strategic Advisory Group of Experts (SAGE) Working Group on VH [14]. Many researchers have attempted to quantify resistance to vaccination using these tools. The results have often been used to relate the hesitancy of vaccinations to demographic characteristics, religious beliefs, behavior, attitudes, and parental knowledge. As shown in their studies, there are no common factors that are congruent towards predicting this phenomenon. The 3-year analysis from the WHO/UNICEF Joint Reporting Form data revealed that in two-thirds of the countries that conducted the reporting, the reasons for hesitancy in vaccination are based on opinions rather than assessments [6]. The WHO Expanded Program on Immunization targets to reach unvaccinated and under-vaccinated children and communities through innovative distribution and administration methods as outlined in the Regional Strategic Framework 2022–2026 [15]. Recognizing growing VH, proactive measures to identify barriers and formulate strategies on preemptive risk communication and community engagement to counter false narratives [16] need to be undertaken in the future.

In recent years, several approaches have been taken to utilize digital technology efficiently to predict VH. These include the use of mobile applications for vaccination [17] and Immunization Information Systems, which are electronic registry systems that have the potential to enhance vaccine uptake. The advent of big data, artificial intelligence (AI), and machine learning (ML) provides an opportunity for medical and health datasets to be analyzed and predictions to be made on the diagnosis and prognosis of diseases [18]. Several ML methods have been used to build disease prediction models for respiratory diseases [19], cancer [18], and diabetes [20] with high accuracy. Villavicencio *et al.* [21] applied several ML models using the Waikato Environment for Knowledge Analysis (WEKA) to analyze and predict the presence of COVID-19 in individuals. Lincoln *et al.* [22] assessed the willingness of

adults to take the COVID-19 vaccine with a comprehensive set of putative predictors. The predictive power was analyzed using an ML algorithm, and it was found that vaccine hesitancy (VA) could be predicted with high accuracy by a parsimonious ML algorithm.

Malaysia is experiencing an emergence of anti-vaccine sentiments among certain groups, and their apprehensions toward the safety and side effects of vaccines [23] may cause a spillover effect towards trust in childhood vaccines. Although the government plays a responsible role in providing free vaccines for childhood, it has been shown that there are parents who are still refusing vaccines for their children. Preventive measures must be taken to ensure that signs of parents' reluctance are detected early. The advancement of AI in healthcare provides policymakers with an alternative to manage VH. Therefore, this study aims to develop an instrument based on parents' opinions, practices, and attitudes to predict VH among Malaysian parents on childhood immunization by applying ML algorithms.

2. METHODOLOGY

2.1. Study design

The development of the instrument was divided into two phases. In Phase 1, a questionnaire was created by adopting items from prior research. Experts then reviewed the questionnaire to determine its content validity. The questionnaire was also translated into the national language. A cross-sectional pilot study was conducted to obtain the appropriate items and an instrument with internal consistency. In Phase 2, the instrument derived from the pilot study was self-administered to participants at a general hospital, and the data were analyzed using ML algorithms.

2.2. Adoption/development of item

Instruments, tools, questionnaires, and survey questions used in previous studies were searched through PUBMED, Google Scholar, EBSCO host, SCOPUS, and WOS using keywords such as "VH," "childhood vaccinations," or "immunizations," "questionnaire," "tools," "instrument," and "survey questions." Two scales, the PACV [13] and VHS [14] questionnaires, were identified for use. The PACV is a validated measure developed by Opel *et al.* [13] to measure VH in the United States (US), a high-income country. The language used in PACV is written in American English and designed for a US audience for self-administration, readable at the 6th-grade level. The VHS is a tool created by the SAGE Working Group on VH, convened by the WHO. The resulting survey includes a combination of core, Likert-scale, and open-ended questions, designed to uncover the scope, drivers, and nuances of VH. It was administered through the WHO-UNICEF Joint Reporting Forms in several regions in the Americas and the European and East, Central, and Southern African regions. Apart from high-income countries the tool can also be applied to middle- and low- and middle-income countries (LMIC). The authors recommended continuous monitoring, cultural adaptation, and validation of the tool in varied contexts. Both the PACV and VHS are combined to harmonize the tools from the perspective of different country settings, from high to LMIC. It is considered

that combining the two surveys would improve its adaptability in a newly industrialized nation like Malaysia and negate the effects of biases.

Twenty-six items were adopted and adapted from both the PACV and VHS, of which 12 questions were from PACV, consisting of 5-point Likert scale questions (5 questions), Yes, No, and Don't know questions (5 questions), and 11-point scale questions (2 questions). The other 14 questions are from the VHS, consisting of the dichotomous "Yes/No" and 5-point Likert scale questions/statements. The open-ended questions in the VHS were not adopted, as they require some knowledge of vaccine types and recalls, which would likely deter parents from answering them.

Malaysia's population consists of Malays, Chinese, Indians, and indigenous groups, each with its own primary language. Considering the significant cultural and language gap between the Malaysian population and the population in which both tools were developed and administered, validating the content of the instrument is crucial. This consideration applies to both the PACV and VHS. The content validation of the questionnaire was conducted through a non-face-to-face approach, whereby an online content validation form was sent to a three-member expert panel comprising a pediatrician, a primary care physician, and a social scientist. The panel was tasked with reviewing the relevance, essentiality, and clarity of the items for the Malaysian population. They were also encouraged to write their comments on the statements in the validation form. Clear instructions were given to the panel to facilitate the review process. To determine the relevance of the items [24], the Item Content Validity Index (I-CVI) was calculated. The I-CVI was determined from the scale-level content validity index based on the average method (S-CVI/Ave) and the scale-level content validity index based on the universal agreement method (S-CVI/UA).

The adapted questionnaire was translated into Malay (the country's national language) using both forward and backward translation to capture the intended meaning and maintain the local semantics and conceptual equivalence. An accredited translation agency performed the translation. A combination of both Malay- and English-worded items was used in the generated questionnaire.

2.3. Pilot study

The pilot study was conducted between April and August 2022. The sample size for the pilot study was determined by the number of items in the questionnaire. The minimum sample required for the pilot was based on the recommended range of 3:1 to 6:1 on the number of cases per variable (N/p) [25]. Hair Joseph [26] suggested that a minimum sample size is dependent upon the model complexity, and if there are five or fewer latent constructs with more than three measuring items each, the minimum sample size required is 100. Therefore, the sample size was calculated based on a 5:1 N/p ratio, and the minimum number required was 130.

Universal sampling was conducted to recruit participants from the staff of the institution. To ensure sufficient samples, participants were also recruited from a local private hospital, with the sampling frame being the number of

patients on a particular day. Respondents with children aged between 0 and 15 years who understand English or Malay were eligible for inclusion and invited to take part in the study. The questionnaire, participants' information leaflet (PIL), and informed consent form (ICF) were distributed to eligible staff through the heads of departments. Participants recruited at the hospital received the dual-language questionnaire after they understood the study and provided written consent.

Several preliminary steps were conducted to determine the suitability of analyzing the data using exploratory factor analysis (EFA). Items with communalities of 0.5 and above were considered as the criterion to undergo EFA, as the sample size was small [27]. Sampling adequacy was checked using the Kaiser–Meyer–Olkin (KMO) measured value, and data suitability was evaluated with Bartlett's test of sphericity. A *p*-value of <0.05 would show the appropriateness of conducting EFA on the data. Principal component analysis (PCA), using the Varimax rotation method, was performed to extract components.

2.4. ML phase

The questionnaire generated from the pilot study was applied to a sample of mothers/fathers at a public hospital. The data was collected at Hospital Raja Permaisuri Bainun (HRPB), Ipoh, between October and December 2023. It is a regional referral hospital for the state of Perak, situated in the capital city of Ipoh. HRPB is a 990-bed hospital offering multidisciplinary medical services, serving outpatients and inpatients from all districts in Perak. The sampling frame consisted of parents who attend the Pediatric and Obstetrics and Gynecology (O&G) Clinic of the hospital. The inclusion criteria of parents are those with children aged from 0 (at birth) to 15 years old. The questionnaire was distributed to parents (either father or mother) who visited the O&G Clinic, and those who brought their children to the Pediatric Clinic. The questionnaire was delivered using systematic random sampling. At every fifth interval, individuals were approached based on their queue number. Each participant was given the questionnaire, ICF, and PIL. Participants who agree to answer the questions in the instrument must also sign the ICF, which is witnessed by the data collector. After answering the questionnaire, the questionnaires were collected, and the data collector checked for their completion. A token of appreciation was given to the participants. During this phase, the targeted minimum number of questionnaires is 500.

2.5. ML algorithms

In Phase 2 of the study, ML algorithms were applied to evaluate and verify the instrument construct. The data obtained from both the Pediatric and O & G clinics were entered into an Excel sheet. The data were exported and analyzed using the WEKA version 3.8.6 software. In this study, we used supervised ML algorithms to build an accurate model. The algorithms used were Random Forest (RF), K-Nearest Neighbor (KNN), Logistic Regression (LR), RF + Bagging, KNN + Bagging, LR + Bagging, RF + AdaboostM1, KNN + AdaboostM1, LR + AdaboostM1, and Decision Tree (DT) + RF. The hold-out (80:20) and 10-fold cross-validation methods were used in the model development.

2.5.1. Random forest

RF is an ensemble learning technique that generates numerous DTs during the training phase and produces the class that represents the mode of the classes (for classification) or the average prediction (for regression) from the individual trees. It employs bagging (bootstrap aggregating) to create several subsets of the training data and incorporates randomness in feature selection during tree partitioning, thereby mitigating overfitting and enhancing generalization. Each tree is trained autonomously, and the ultimate choice is determined by majority voting among all trees in the ensemble.

2.5.2. K-nearest neighbors

KNN is a nonparametric, instance-based learning method characterized by its sluggish approach, utilized for both classification and regression tasks. It categorizes a data point according to the classifications of its neighbors. The procedure, given a value of k , determines the k training instances nearest to the test instance utilizing a distance measure like Euclidean or Manhattan distance and assigns the predominant class among them to the instance. Due to its lack of assumptions regarding the underlying data distribution, it exhibits significant flexibility; however, it is computationally demanding for extensive datasets.

2.5.3. Logistic regression

LR is a supervised learning approach predominantly employed for binary classification applications. It quantifies the likelihood that a specific input point is associated with a designated class through the logistic (sigmoid) function. The model evaluates the association between the dependent binary variable and one or more independent variables by utilizing a logistic function on a linear amalgamation of the features. Multinomial and ordinal extensions are employed for multiclass issues. It is very effective for linearly separable datasets.

2.5.4. Decision tree

A DT is a supervised learning technique employed for classification and regression applications. It iteratively divides the dataset into subgroups according to feature values, forming a tree-like decision model. At each node, the algorithm selects the feature and threshold that most effectively distinguishes the classes, employing metrics such as Gini impurity, entropy (information gain), or variance reduction (in regression). The procedure persists until a termination criterion is satisfied (e.g., maximum depth, minimum samples per leaf). It is simple to interpret, accommodates nonlinear data, and necessitates minimal preprocessing. However, it is susceptible to overfitting if inadequately trimmed or regularized.

2.5.5. Bagging

Bagging is an ensemble method aimed at enhancing the stability and precision of ML algorithms, especially high-variance models such as DTs. It operates by generating several bootstrapped datasets (random samples with replacement) from the original dataset, training

an individual model on each bootstrapped dataset, and consolidating predictions (e.g., majority voting for classification, averaging for regression). Bagging diminishes variance without augmenting bias, so effectively mitigating overfitting. RF is a widely utilized bagging technique that employs DTs as base learners, with arbitrary feature selection.

2.5.6. AdaBoostM1

AdaBoostM1 is an enhancement of the original AdaBoost algorithm designed for multiclass classification. It constructs an ensemble of weak learners, typically shallow DTs or stumps, in a sequential fashion. During each iteration, instances incorrectly categorized by the existing model are allocated increased weights, compelling the subsequent learner to concentrate more on them. A weighted majority vote is employed for the final forecast, with each model's vote weighted according to its accuracy. AdaBoost's efficacy resides in transforming weak learners into a robust ensemble by decreasing exponential loss and enhancing model bias resilience. Nonetheless, it is susceptible to noisy data and outliers because of its focus on misclassified instances.

2.5.7. Performance metrics

Accuracy is the proportion of total correct predictions (both true positives and true negatives) to the total number of predictions.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

True positive rate (TPR) is the proportion of actual positives correctly predicted.

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

False positive rate (FPR) is the proportion of actual negatives that were incorrectly classified as positive.

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}}$$

Precision is the proportion of predicted positives that are actually correct.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall is also referred to as TPR or sensitivity. High recall means most actual positives are captured, but possibly with more false positives (FP).

The F1 score is the harmonic mean of precision and recall, balancing both concerns.

$$\text{F1 - Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

2.6. Data processing and scoring

The original scores of the questionnaire were used for analysis in the pilot study phase. For the first five items, “Yes”, “Unsure”, and “No” were given a scale of “0,1 and 2”, or reversed, depending on the direction of the statements [13,28]. Likewise, a “1–5” and “1–10” scores were applied accordingly to the other questions. The scores assigned to each item are based on the direction of the statement, as in the original study and other studies that used the translated versions [13,28,29].

While in the ML phase, we collapsed the different response formats, 3-point and 5-point Likert scales, into “0” and “2”, and “0”, “1”, and “2”, respectively. We then added the scores of the individual items, giving a total score of 26. In line with previous studies [13,28,29], the total score was transformed into a 0–100 scale by applying a simple linear transformation. The generated score was dichotomized into “1”, for a score of ≥ 50 , indicating hesitancy, and “0” for a score of < 50 , indicating nonhesitancy. We used this binary scale for the ML model analysis. To confirm that the cut-off point of 50% is the optimum threshold level, we computed the TP, TN, FP, and false negative (FN) at four different thresholds: 40%, 45%, 50%, and 55%. We then calculated the sensitivity, specificity, accuracy, and Youden’s J Index and compared them across all four threshold levels.

3. RESULTS

3.1. Content validity

All three experts reviewed the content of the questionnaire and assigned scores for the relevance, essentiality, and clarity of each item, which range from 1 to 4, with scores 3 and 4 denoting acceptable levels. During the review, one panel commented that the meaning of one item in the questionnaire is ambiguous. Two expert panels disagreed with using the word “shots,” which is unusual in the context of the local setting. An example of such a statement is “Children get more shots than are good for them.” One item of the VHS duplicates an item of the PACV. Thus, only one was included. The questionnaire that was sent for translation had 25 items.

The I-CVI was calculated by dividing the number of experts giving a rating of 3 or 4 by the number of experts. Except for one expert, all experts assigned a score of 3 on all items, demonstrating agreement that each item was relevant. The S-CVI/Ave and the S-CVI/UA are calculated as follows:

$$S-CVI/Av = \frac{\text{Total-Content Validity Index Score}}{\text{No of Items}}$$

$$= 24.67/25=0.987$$

The S-CVI/UA is

$$\text{The S-CVI/UA} = \frac{\text{Total-Content Validity Index Score}}{\text{Universal Agreement}}$$

$$= 24/25 = 0.96$$

The calculation of I-CVI, UA, S-CVI/Av, and S-CVI/UA is shown in Table 1. We concluded that all the validity index scales meet satisfactory levels, as the minimum level for

Table 1. Relevance rating on instrument’s items by experts.

Item	Expert 1	Expert 2	Expert 3	Experts in agreement	I-CVI	UA
Q1	1	1	1	3	1	1
Q2	1	1	1	3	1	1
Q3	1	1	1	3	1	1
Q4	1	1	1	3	1	1
Q5	1	1	1	3	1	1
Q6	1	0	1	2	0.67	0
Q7	1	1	1	3	1	1
Q8	1	1	1	3	1	1
Q9	1	1	1	3	1	1
Q10	1	1	1	3	1	1
Q11	1	1	1	3	1	1
Q12	1	1	1	3	1	1
Q13	1	1	1	3	1	1
Q14	1	1	1	3	1	1
Q15	1	1	1	3	1	1
Q16	1	1	1	3	1	1
Q17	1	1	1	3	1	1
Q18	1	1	1	3	1	1
Q19	1	1	1	3	1	1
Q20	1	1	1	3	1	1
Q21	1	1	1	3	1	1
Q22	1	1	1	3	1	1
Q23	1	1	1	3	1	1
Q24	1	1	1	3	1	1
Q25	1	1	1	3	1	1
				S-CVI/ (Av)	*0.987	
				S-CVI/ (UA)		**0.960
Proportion relevance	1	0.96	1			
Av proportion of items judged as relevance across three experts				0.987		

Notes: I-CVI (Item) = (number of experts giving rating 3 or 4) / (number of experts)

*S-CVI/(Av) = mean (I-CVI across items)

**S-CVI/(UA) = number of items with I-CVI = 1 / (total number of items)

average scale content validity should be 0.80 and above for fewer than 4 raters [30].

Two members of the research team reviewed both the forward and backward translation of the 25-item instrument. Discrepancies between the versions were discussed to ensure that the concept, meaning, and semantics remained equivalent. The most appropriate translation was chosen for both the English and Malay versions, based on the best-translated

options. Two items were found to have similar meanings after translation: “I am concerned about serious adverse effects of vaccines” and “How concerned are you that your child may experience serious side effects from injecting a vaccine?” In the first, responses are rated on a Likert scale of 1–5, with 1 meaning “strongly disagree” and 5 “strongly agree,” while in the second, 1 indicates “not at all concerned” and 5 “very concerned.” The first statement addresses general concerns about the potential for serious vaccine side effects. In contrast, the second focuses on parents' worries that vaccination could cause such effects, influencing their acceptance or rejection of vaccines. As a result, the first item was excluded.

Neither the PACV nor the VHS addressed perceptions of natural immunity, where beliefs about “natural immunity” versus vaccination are common across many populations. From the literature review, it appears that in certain Malaysian populations, there is a greater reliance on traditional medicines or religious healing practices. These include the use of complementary or homeopathic medicines and concerns about the halal status of vaccines [8,31]. The influence of complementary medicines and religious beliefs on the use of halal vaccines may require further investigation beyond what current tools offer. Consequently, questions on these two areas were added. Although the three experts did not rate these items, their relevance was assessed through statistical analysis after the pilot study. After including these two concepts, the final instrument used for the pilot comprised 26 items.

3.2. Pilot study

132 participants were recruited for the study. Of these, 124 complete datasets were available for analysis after excluding eight participants who did not meet the inclusion criteria. The instrument achieved a KMO value of 0.887 with a significant p -value of <0.0001 for Bartlett's Sphericity test, as shown in Table 2. PCA was conducted, and the number of principal components

was determined using both the cumulative explained variance and the scree plot. Components explaining at least 60% of the total variance were retained. Additionally, the Kaiser criterion was applied, retaining components with eigenvalues greater than one. There were 15 items with loadings greater than 0.5, and three subdomains meeting this criterion were extracted [26]. Reliability tests were performed on these three subdomains. The first subdomain consisted of 10 items, with a Cronbach's alpha of 0.861. The second subdomain consisted of three items, with a Cronbach's alpha of 0.881, while the internal consistency value of subdomain 3, with two items, was 0.510. Including this subdomain in the final instrument resulted in a lower reliability value, with Cronbach's alpha of 0.618, compared to 0.850 if only subdomains 1 and 2 were used. We excluded subdomain 3, as for a construct to have adequate construct reliability, the value should be 0.70 or higher [26]. Table 3 presents the items, factor loadings, and subdomains that were extracted and included in the final instrument following the PCA. The first and second subdomains were labelled “attitudes and beliefs about vaccines,” and “safety and efficacy,” respectively.

For the instrument's score across the four thresholds examined (40%, 45%, 50%, and 55%), the optimal performance was observed at the 50% threshold in terms of sensitivity, specificity, and overall accuracy. The instrument achieved perfect sensitivity, specificity, and accuracy (all 1.0000), indicating complete discrimination between cases and noncases within this dataset. Evaluation of the optimal cut-off was further

Table 2. Value of KMO and Bartlett's test.

KMO and Bartlett's test		
KMO measure of sampling adequacy		0.887
Bartlett's test of sphericity	Approximate chi-square	2,199.988
	df	253
	Sig.	0.0000

Table 3. Items, factor loadings and components of instrument.

Rotated component matrix		Domain	
		1	2
Q14	Having my child vaccinated is important for the health of others in the community	0.919	
Q12	Childhood vaccines are important for my child's health	0.918	
Q15	All childhood vaccines offered by the government program in my community are beneficial	0.899	
Q3	Do you believe that vaccines can protect children from serious diseases	0.836	
Q13	Childhood vaccines are effective	0.866	
Q1	Have you ever delayed to get a vaccination for your child for reasons other than illness or allergy?	0.808	
Q2	Have you ever refused a vaccination for your child for reasons other than illness or allergy?	0.780	
Q5	If you had another baby today, would you want the baby to receive all the recommended vaccines?	0.760	
Q11	I am able to openly discuss my concerns about vaccinations with my child's doctor	0.593	
Q4	Do you think most parents like you have their children vaccinated with all the recommended vaccines	0.716	
Q24	How concerned are you that any of the vaccines your child receives may be unsafe?		0.906
Q23	How concerned are you that your child may experience serious side effects from injecting a vaccine?		0.884
Q25	How concerned are you that a vaccine might not prevent the disease?		0.847

Cronbach alpha values for Factor 1 = .875, Factor 2 = .732. Cronbach alpha value for 13-item instrument is .850.

Table 4. 13-item instrument scoring scale.

Item No	(Yes = 2, 1 = Unsure, No = 0)
1	Have you ever delayed getting a vaccination for your child for reasons other than illness or allergy?
2	Have you ever refused a vaccination for your child for reasons other than illness or allergy?
3	If you had another baby today, would you want the baby to receive all the recommended vaccines? (Yes = 0, 1 = Unsure, No = 2)
4	Do you think that most parents like you have their children vaccinated with all the recommended vaccines?
5	Do you believe that vaccines can protect children from serious diseases? Strongly agree = 0, Neither agree nor disagree = 1, Strongly disagree = 2
6	I am able to openly discuss my concerns about vaccinations with my child's doctor
7	Childhood vaccines are important for my child's health
8	Childhood vaccines are effective
9	Having my child vaccinated is important for the health of others in the community
10	All childhood vaccines offered by the government programme in my community are beneficial (Not at all concerned = 0, Not sure = 1, Very concerned = 2)
11	How concerned are you that your child may experience serious side effects from injecting a vaccine?
12	How concerned are you that any of the vaccines your child receives may be unsafe?
13	How concerned are you that a vaccine might not prevent the disease?

Table 5. Results of ML algorithm based on hold-out (80:20) method.

Algorithm	Accuracy	TP Rate	FP Rate	Precision	Recall	F-score	AUC
RF	97.06%	0.971	0.314	0.969	0.971	0.969	0.988
KNN	97.06%	0.971	0.158	0.973	0.971	0.972	0.906
LR	96.08%	0.961	0.002	0.976	0.961	0.965	0.996
RF + Bagging	97.06%	0.971	0.314	0.969	0.971	0.969	0.993
KNN + Bagging	97.06%	0.971	0.158	0.973	0.971	0.972	0.988
LR + Bagging	99.02%	0.99	0.001	0.992	0.99	0.991	1
RF + AdaBoostM1	97.06%	0.971	0.314	0.969	0.971	0.969	0.988
KNN + AdaBoostM1	97.06%	0.971	0.158	0.973	0.971	0.972	0.906
LR + AdaBoostM1	96.08%	0.961	0.002	0.976	0.961	0.965	0.996
DT + RF	96.08%	0.961	0.315	0.961	0.961	0.961	0.99

supported by Youden's *J*-index, which quantifies the maximum potential effectiveness of a diagnostic threshold. The *J*-index increased from 0.7170 at the 40% threshold to 0.9807 at 45%, indicating a marked improvement in the balance between sensitivity and specificity. The highest *J*-index was observed at the 50% threshold ($J = 1.0000$), demonstrating perfect discriminatory performance in this dataset.

3.3. ML algorithm

The final instrument is presented in Table 4 and was used during the primary data collection at HRPB, Ipoh. A total of 524 datasets were collected, but only 510 datasets were accepted for ML analysis. Fourteen datasets were excluded due to several missing values and violations of the inclusion criteria. The characteristics of the participants were analyzed. The parents' age ranges from 19 and 62 with a mean age of 35.4 (± 7.6) years. The youngest child was 1 month old while

the oldest was 15 years old, and a mean of 3.6 ($+3.7$) years. The respondents were predominantly Malays (78%) with others comprising only 21.5%. Most respondents were working with 61% being employed and 9.2% self-employed. The income level of the majority is less than RM5000 per month (80%), and out of these 35% of them earned <RM2000. The total VH score ranged from 0% to 96% while the binary scale of VH found 52 (10.2%) hesitant and 458 (89.8%) nonhesitant mothers or fathers.

The results of the ML model are presented in Tables 5 and 6, respectively. According to the findings, logistic regression with bagging (LR + Bagging) yielded the best results, with 99.02% accuracy for the hold-out method and 97.45% for the 10-fold cross-validation method. The performance metrics of several ML algorithms evaluated using the hold-out approach (80:20 train-test split) are displayed in Table 5. Among the tested models, the one that performed the best overall was LR

Table 6. Results based on 10-fold cross-validation method.

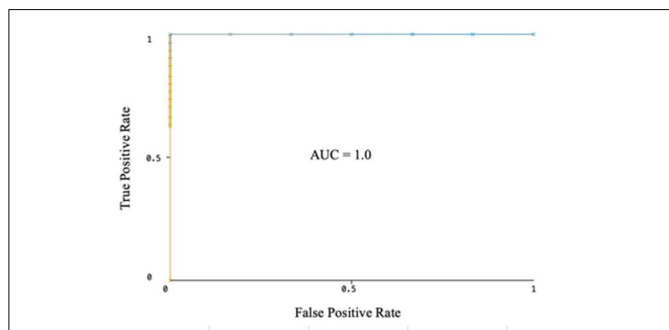
Algorithm	Accuracy	TP rate	FP rate	Precision	Recall	F-score	AUC
RF	96.67%	0.967	0.216	0.966	0.967	0.965	0.992
KNN	95.88%	0.959	0.217	0.957	0.959	0.958	0.871
LR	97.06%	0.971	0.15	0.97	0.971	0.97	0.978
RF + Bagging	96.67%	0.967	0.216	0.966	0.967	0.965	0.993
KNN + Bagging	95.88%	0.959	0.217	0.957	0.959	0.958	0.945
LR + Bagging	97.45%	0.975	0.052	0.977	0.975	0.975	0.992
RF + AdaboostM1	96.86%	0.969	0.2	0.968	0.969	0.968	0.993
KNN + AdaBoostM1	95.88%	0.959	0.217	0.957	0.959	0.958	0.871
LR + AdaBoostM1	97.06%	0.971	0.15	0.97	0.971	0.97	0.978
DT + RF	95.88%	0.959	0.201	0.958	0.959	0.958	0.99

+ bagging. Its accuracy of 99.02% is notably higher than that of other algorithms. Moreover, this model exhibited exceptional classification reliability, with the lowest FPR (0.001) and the highest area under the curve (AUC = 1.0). The F1-score (0.991), precision (0.992), and recall (0.990) all underscore its robustness and effectiveness in accurately predicting the class with the fewest possible errors.

A consistent accuracy of 97.06% was maintained by other models such as RF, KNN, and their ensemble variations, including Bagging and AdaBoostM1. In contrast to RF-based models, which had a larger FPR (0.314), KNN-based models exhibited somewhat higher precision (0.973) and lower FPRs (0.158). Strong individual candidates, LR and LR with AdaBoostM1, demonstrated excellent performance with very low FPRs (0.002), high AUC (0.996), and respectable accuracy (96.08%). The DT + RF achieved moderate results with 96.08% accuracy and an AUC of 0.990; however, its relatively high FPR (0.315) may limit its use in sensitive or real-time classification tasks. Based on the 80:20 hold-out validation, LR + Bagging is the most accurate, precise, and reliable algorithm, outperforming all other models in this study. It is particularly suitable for critical applications like security screening, where reducing FP and ensuring high classification accuracy are essential.

The evaluation results of various ML algorithms using the 10-fold cross-validation technique, which offers a more comprehensive assessment of model performance, are presented in Table 6. With a maximum accuracy of 97.45%, LR + Bagging was the best-performing model among those tested. It also achieved an excellent AUC of 0.992, a low FPR (0.052), and a significant TPR (0.975), along with high precision (0.977), recall (0.975), and F1-score (0.975). These results demonstrate its effectiveness and robustness in classification tasks. Standalone LR and its variation with AdaBoostM1 showed competitive performance with accuracy rates of 97.06% and 0.978 AUC, respectively, although with a slightly higher FPR of 0.15.

With high TP rates (~0.967 to 0.969) and strong AUC values (0.992–0.993), ensemble methods using RF, such as Bagging and AdaBoostM1, maintained consistent accuracy levels

**Figure 1.** ROC curve of LR + Bagging using 80:20 hold-out method.

of around 96.67% to 96.86%, indicating stable performance. In contrast to LR-based models, they were affected by comparatively higher FPRs (0.2–0.216). Conversely, the least successful methods in this evaluation were KNN and its ensemble variations, which consistently produced lower AUC values (0.871–0.945) and an accuracy of 95.88%. The DT + RF combination, with a moderate AUC of 0.990, performed similarly to KNN.

The comparison between the 10-fold cross-validation and hold-out (80:20) methods shows consistent overall rankings among algorithms, with LR + Bagging performing best in both cases. It achieved 99.02% accuracy and an AUC of 1.0 in the hold-out method, slightly higher than the 97.45% accuracy and AUC of 0.992 observed in cross-validation, highlighting its strong and stable performance.

Figure 1 illustrates the receiver operating characteristics (ROC) curve for LR + Bagging, utilizing the 80:20 hold-out method. The curve sharply rises toward the top-left corner, indicating a high TPR with a minimal FPR. This is consistent with the model's AUC of 1.0, reflecting perfect separability between classes in this specific train-test split. It confirms that the model performs exceptionally well when evaluated on a single, unseen data partition.

Figure 2 displays the ROC curve for LR + Bagging using 10-fold cross-validation, which also shows excellent performance, closely following the ideal curve. Although slightly below the hold-out ROC curve, it still reflects a high

level of accuracy and robustness across multiple data splits. The associated AUC of 0.992 demonstrates that the model maintains strong generalization ability and consistent classification performance, even when evaluated rigorously. The key hyperparameters of the ML algorithms are provided in Table 7.

3.4. Error and misclassification analysis

In 80:20 hold-out validation, across all models, the number of misclassified samples was ≤ 4 out of 102 ($\approx 3.9\%$), indicating high model reliability. LR+Bagging produced

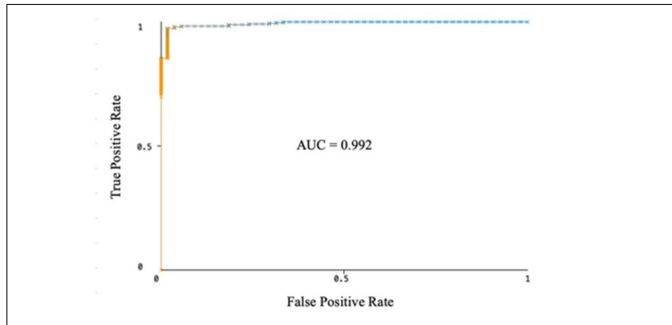


Figure 2. ROC curve of LR + Bagging using 10-fold cross-validation.

Table 7. Key hyperparameters.

Algorithm	Key hyperparameters (as implemented in WEKA 3.8.6)
RF	Number of tree: 100 Maximum depth: unlimited
KNN	Number of neighbors(<i>k</i>): 5 Distance metric: Euclidean distance Weighting: uniform Normalization: min-max normalization
LR	Regularization: L2 (ridge) Ridge parameter: 1.0E ⁻⁸ Max iterations: 1,000
J48 DT	Confidence factor for pruning: 0.25 Minimum instances per leaf: 2
Bagging (with RF/KNN/LR base)	Bag size percent: 100 Random seed: 1 Number of iterations: 10 Base learner: specified algorithm (RF, KNN, or LR)
AdaBoostM1 (with RF/KNN/LR base)	Number of iterations: 10 Weight threshold: 100 Random seed: 1 Base learner: specified algorithm (RF, KNN, or LR)
DT + RF hybrid	Base tree: J48 Number of RF trees: 50 Combination rule: majority voting

the fewest errors (1 FP 0 FN), corresponding to an accuracy of 99.02%. RF, KNN, and their ensemble variations each misclassified three samples, mainly near the class boundary, where respondents scored between 45 and 55 on the hesitancy scale. The confusion matrices for each classifier confirm that the majority of errors occurred among “borderline hesitant” respondents rather than clearly polarized classes. Confusion matrices for 80:20 hold-out validation are provided in Table 8.

In the 10-fold cross-validation setup, total misclassification rates ranged from 2.5% to 4.1%, consistent across folds. Most errors were FN (hesitant parents misclassified as nonhesitant). This outcome aligns with the dataset’s slightly higher feature overlap in neutral-attitude responses, which limits separability in probabilistic classifiers such as LR and KNN. Nevertheless, LR + Bagging consistently achieved the highest average accuracy (97.45%) and lowest FPR (0.052). Confusion matrices for 10-fold cross validation are provided in Table 9.

Table 8. Confusion matrices for 80:20 hold-out validation.

Classifier	For 80:20 hold-out validation		
		0	1
RF	0	95	1
	1	2	4
KNN	0	94	2
	1	1	5
LR	0	92	4
	1	0	6
RF + Bagging	0	95	1
	1	2	4
KNN + Bagging	0	94	2
	1	1	5
LR + Bagging	0	95	1
	1	0	6
RF + AdaboostM1	0	95	1
	1	2	4
KNN + AdaBoostM1	0	94	2
	1	1	5
LR + AdaBoostM1	0	92	4
	1	0	6
DT + RF	0	94	2
	1	2	4

Table 9. Confusion matrices for 10-fold cross validation.

Classifier	For 10-fold cross-validation		
		0	1
RF	0	452	4
	1	13	41
KNN	0	448	8
	1	13	41
LR	0	450	6
	1	9	45
RF + Bagging	0	452	4
	1	13	41
KNN + Bagging	0	448	8
	1	13	41
LR + Bagging	0	446	10
	1	3	51
RF + AdaboostM1	0	452	4
	1	12	42
KNN + AdaBoostM1	0	448	8
	1	13	41
LR + AdaBoostM1	0	450	6
	1	9	45
DT + RF	0	447	9
	1	12	42

4. DISCUSSION

The purpose of the study is to develop an instrument to predict VH in childhood immunization. This study employed a three-stage development process, including item relevance rating, multivariate analysis, and ML algorithms to create the questionnaire, which we named the Malaysian VH Instrument. While many instruments developed earlier have been validated and employed by many researchers, combining, translating, and adapting them for a different population requires the modified instrument to be reliable and consistent, making multiple layers of validation necessary. Furthermore, a crucial factor is to validate the instrument against an external objective criterion.

In this respect, this study has several limitations. First, in the pilot study, the adapted instrument was meant to be administered to a small sample and then analyzed for item-total correlations, missing data, and floor or ceiling effects. For the internal structure validation and to test whether the adapted items still reflect the intended constructs's dimensionality, this

typically uses EFA for which the recommended sample size is 200 or more. Due to time constraints, we are unable to collect another set of data to fulfill the requirement of conducting an EFA. PCA, which is not strictly a method of EFA, was used. With PCA, we summarized the variables into fewer composite indices and reduced the multicollinearity before the ML modelling was applied.

Second, our primary data collection site is a single tertiary referral hospital, which may not accurately reflect the state's population composition and may introduce bias towards a specific ethnic group in its representation. This was shown by the demographic characteristics of the sample, whereby more than 70% of the respondents were from one ethnic group, the Malays, whereas the Chinese make up around 44% of the local population. In our initial proposal, we planned to get respondents from a few private hospitals, but they were reluctant to allow the recruitment of their clients. During the data collection, we also planned for a systematic random sampling, but we were unsure whether this was followed through due to the lack of reports on the daily sampling frame and the number of respondents. Therefore, there could be a possibility of convenience rather than random sampling. This could also be one of the reasons for the majority of respondents to be from one ethnic group.

In the fields of public health and medicine, ML has become increasingly significant. ML relies on various algorithms to solve data problems and enables computers to learn without explicit programming. It utilizes many potential predictors or inputs that can enhance classification performance. However, in the medical and health sectors, collecting large datasets requires additional time and computational resources. Studies have shown that the average accuracy of classifiers trained on small datasets ranges from 62% to 99% [32]. Many studies that took the ML approach usually used national data from the registry or Electronic Health Records [33]. Our study has a limitation in obtaining a large dataset, as the process of conducting a survey is time-consuming. Nevertheless, the application of a multicountry survey makes it possible for a large participation, big dataset, and the prediction of vaccine willingness using ML with high accuracy [22]. This should be the approach in the future, whereby collaboration between researchers at different centers in the country would enable large data collection.

Even though our instrument achieves high accuracy with an algorithm using LR with bagging, which is 99% in the 80:20 hold-out method and 97% in the 10-fold cross-validation method, it lacks external validation to confirm its predictive power. At best, this instrument is an adaptation and combined use of PACV and VHS applied in the Malaysian context. It is still at the development stage, and the application of the ML models is to evaluate the instrument's internal classification performance. To further explore its potential usefulness, a subsequent phase of the study is necessary to test its correlation with a benchmark variable, such as actual vaccination records.

5. CONCLUSION

This study's objective is to develop an instrument to predict the reluctance of mothers/parents to vaccinate their children against vaccine-preventable diseases in childhood

based on attitude, opinion, and beliefs. Content analysis and multivariate data analysis, which included PCA, reliability, and consistency tests were employed. To strengthen the instrument's performance in the absence of external validation ten supervised ML algorithms were used. A comparative analysis was conducted by evaluating the models' performance by the hold-out (80:20) and 10-fold cross-validation methods through the WEKA ML software. The results show that in terms of accuracy, LR + Bagging produced the best results with 99.02% for the hold-out method and 97.45% for the 10-fold cross-validation method.

The instrument developed from this study needs to be further validated by examining the association between hesitancy levels and parents' actual behavior, such as their timeliness in vaccinating their children. Its potential usefulness can be further explored, given the limited studies on ML models using an opinion-based instrument.

6. ACKNOWLEDGMENT

We would like to thank the Director-General of Health Malaysia for his permission to publish this article. We would also like to thank all participants in the pilot and main data collection phase for their willingness to participate in the study. We also want to thank the hospital director, the clinical research centre, and the clinic staff of the Pediatric and O&G Clinics of HRPB for their permission and cooperation given for us to collect the data. We would also like to thank the management of Hospital Ar-Ridzuan for allowing us to obtain data from their patients. We thank everyone who has collaborated either directly or indirectly towards this study.

7. AUTHOR CONTRIBUTIONS

All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agree to be accountable for all aspects of the work. All the authors are eligible to be author as per the International Committee of Medical Journal Editors (ICMJE) requirements/guidelines.

8. FINANCIAL SUPPORT

The authors would like to acknowledge the Ministry of Higher Education for the funds provided in this project under the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2021/SKK06/UNIKL/02/3). We would also like to express our appreciation to Universiti Kuala Lumpur for other monetary assistance given during the study.

9. CONFLICTS OF INTEREST

The authors report no financial or any other conflicts of interest in this work.

10. ETHICAL APPROVAL

The study protocol was approved by the Medical Research Ethics Committee (MREC) of the Ministry of Health, Malaysia (NMRR ID: 22-01702-IOU [IIR]), and by the Royal

College of Medicine Perak, Universiti Kuala Lumpur (UniKL), Research Ethics Committee (Approval No.:REC/2021/016). All data collected was kept confidential, and no unique identifiable information was collected. Participants were only offered a little gift in exchange for their voluntary involvement in the study.

11. DATA AVAILABILITY

All tables and figures are generated from the analysis of the data collected and are included in the study. Any other data specifically requested will be made available.

12. PUBLISHER'S NOTE

All claims expressed in this article are solely those of the authors and do not necessarily represent those of the publisher, the editors and the reviewers. This journal remains neutral with regard to jurisdictional claims in published institutional affiliation.

13. USE OF ARTIFICIAL INTELLIGENCE (AI)-ASSISTED TECHNOLOGY

The authors declare that they have not used artificial intelligence (AI)-tools for writing and editing of the manuscript, and no images were manipulated using AI.

REFERENCES

- Shattock AJ, Johnson HC, Sim SY, Carter A, Lambach P, Hutubessy RCW, *et al.* Contribution of vaccination to improved survival and health: modelling 50 years of the Expanded Programme on Immunization. *Lancet*. 2024;403(10441):2307–16. doi: [https://doi.org/10.1016/S0140-6736\(24\)00850-X](https://doi.org/10.1016/S0140-6736(24)00850-X)
- Godlee F, Smith J, Marcovitch H. Wakefield's article linking MMR vaccine and autism was fraudulent. *BMJ*. 2011;342:7452. doi: <https://doi.org/10.1136/bmj.c7452>
- Geoghegan S, O'Callaghan KP, Offit PA. Vaccine safety: myths and misinformation. *Front Microbiol*. 2020;11:372. doi: <https://doi.org/10.3389/fmicb.2020.00372>
- Metcalf CJE, Ferrari M, Graham AL, Grenfell BT. Understanding herd immunity. *Trends Immunol*. 2015;36(12):753–5. doi: <https://doi.org/10.1016/j.it.2015.10.004>
- Macdonald NE. Vaccine hesitancy: definition, scope and determinants. *Vaccine*. 2015;33(34):4161–4. doi: <https://doi.org/10.1016/j.vaccine.2015.04.036>
- Lane S, Macdonald NE, Marti M, Dumolard L. Vaccine hesitancy around the globe: analysis of three years of WHO/UNICEF joint reporting form data-2015–2017. *Vaccine*. 2018;36(26):3861–7. doi: <https://doi.org/10.1016/j.vaccine.2018.03.063>
- Zulkifli SN, Minhat HS, Lim PY. The factors of primary childhood immunization uptake among the urban poor children in Malaysia. *Malaysian J Public Health Med*. 2022;22(3):53–62. doi: <https://doi.org/10.37268/mjphm/vol.22/no.3/art.1550>
- Chan HK, Soelar SA, Md Ali SM, Ahmad F, Abu Hassan MR. Trends in vaccination refusal in children under 2 years of age in Kedah, Malaysia: a 4-year review from 2013 to 2016. *Asia Pac J Public Health*. 2018;30(2):137–46. doi: <https://doi.org/10.1177/1010539517751312>
- Abd Rahman NA, Ismail WRW, Abd Rahman R, Yusof MP, Idris IB. Who would get measles in Petaling District? A trend analysis of measles outbreak from 2014-2018. *Malaysian J Med Health Sci*. 2020;16(3):67–72.
- Mat Daud MRH, Yaacob NA, Ibrahim MI, Wan Muhammad WAR. Five-year trend of measles and its associated factors in Pahang, Malaysia: a population-based study. *Int J Environ Res Public Health*. 2022;19(13):19. doi: <https://doi.org/10.3390/ijerph19138017>

11. Tok PSK, Jilani M, Misnar NF, Bidin NS, Rosli N, Toha HR. A diphtheria outbreak in Johor Bahru, Malaysia: public health investigation and response. *J Infect Dev Ctries.* 2022;16(7):1159–65. doi: <https://doi.org/10.3855/jdc.16076>
12. Lim KK, Chan YY, Noor Ani A, Rohani J, Siti Norfadhilah ZA, Santhi MR. Complete immunization coverage and its determinants among children in Malaysia: findings from the National Health and Morbidity Survey (NHMS) 2016. *Public Health.* 2017;153:52–7. doi: <https://doi.org/10.1016/j.puhe.2017.08.001>
13. Opel DJ, Mangione-Smith R, Taylor JA, Korfiatis C, Wiese C, Catz S, *et al.* Development of a survey to identify vaccine-hesitant parents: the parent attitudes about childhood vaccines survey. *Hum Vaccin.* 2011;7(4):419–25.
14. Larson HJ, Jarrett C, Schulz WS, Chaudhuri M, Zhou Y, Dube E, *et al.* Measuring vaccine hesitancy: the development of a survey tool. *Vaccine.* 2015;33(34):4165–75. doi: <https://doi.org/10.1016/j.vaccine.2015.04.037>
15. Wazed S. 50th anniversary of expanded programme on immunization: shaping the next 50 years in the WHO South-East Asia region. *Indian J Med Res.* 2024;160(34):259–61. doi: https://doi.org/10.25259/IJMR_1677_2024
16. World Health Organization. Communicating risk in public health emergencies: a WHO guideline for emergency risk communication (ERC) policy and practice. Geneva: World Health Organization; 2018.
17. Atkinson KM, Westeinde J, Ducharme R, Wilson SE, Deeks SL, Crowcroft N, *et al.* Can mobile technologies improve on-time vaccination? A study piloting maternal use of ImmunizeCA, a Pan-Canadian immunization app. *Hum Vaccines Immunother.* 2016;12(10):2654–61. doi: <https://doi.org/10.1080/21645515.2016.1194146>
18. Zhang B, Shi H, Wang H. Machine learning and AI in cancer prognosis, prediction and treatment selection: a critical approach. *J Multidiscip Healthc.* 2023;16:1779-1791. doi: <https://doi.org/10.2147/JMDH.S410301>
19. Mekov E, Miravittles M, Petkov R. Artificial intelligence and machine learning in respiratory medicine. *Expert Rev Respir Med.* 2020;14(6):559–64. doi: <https://doi.org/10.1080/17476348.2020.1743181>
20. Keshkar A, Ayareh N, Atighi F, Hamidi R, Yazdanpanahi P, Karimi A, *et al.* Artificial intelligence in diabetes management: revolutionizing the diagnosis of diabetes mellitus; a literature review. *Shiraz E Med J.* 2024;25(7) :e146903. doi: <https://doi.org/10.5812/semj-146903>.
21. Villavicencio CN, Macrohon JJE, Inbaraj XA, Jeng JH, Hsieh JG. Covid-19 prediction applying supervised machine learning algorithms with comparative analysis using Weka. *Algorithms.* 2021;14(7):7. doi: <https://doi.org/10.3390/a14070201>
22. Lincoln TM, Schlier B, Strakeljahn F, Gaudiano BA, So SH, Kingston J, *et al.* Taking a machine learning approach to optimize prediction of vaccine hesitancy in high income countries. *Scientific Rep.* 2022;12(1):1. doi: <https://doi.org/10.1038/s41598-022-05915-3>
23. Iskandar IM, Mohamed S. Malaysian newspapers' coverage of anti-vaxxers: implication on vaccine intakes. *Malaysian J Med Health Sci.* 2023;19(2):124–9. doi: <https://doi.org/10.47836/mjmhs.19.2.19>
24. Yusoff MSB. ABC of content validation and content validity index calculation. *Educ Med J.* 2019;11(2):49–54. doi: <https://doi.org/10.21315/eimj2019.11.2.6>
25. Cattell RB. The scientific use of factor analysis in behavioral and life sciences. 1st ed. New York, NY: Springer; 2012. doi: <https://doi.org/10.1007/978-1-4684-2262-7>
26. Hair Joseph F, Black William C, Babin Barry J, Anderson Rolph E. Multivariate data analysis. 8th ed. Northway, Andover, Hampshire: Cheriton House; 2019.
27. Hogarty KY, Hines CV, Kromrey JD, Ferron JM, Mumford KR. The quality of factor solutions in exploratory factor analysis: the influence of sample size, communalities, and overdetermination. *Educ Psychol Meas.* 2005;65(2):202–6. doi: <https://doi.org/10.1002/nur.20199>
28. Abd Halim H, Abdul-Razak S, Md Yasin M, Isa MR. Validation study of the parent attitudes about childhood vaccines (PACV) questionnaire: the Malay version. *Hum Vaccin Immunother.* 2020;16(5):1040–9. doi: <https://doi.org/10.1080/21645515.2019.1674112>
29. Mohd Azizi FS, Kew Y, Moy FM. Vaccine hesitancy among parents in a ethnic country, Malaysia [Internet]. *Vaccine Internet.* 2017;35(22):2955–61. doi: <https://doi.org/10.1016/j.vaccine.2017.04.010>
30. Polit DF, Beck CT, Owen SV. Focus on research methods: is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health.* 2007;30(4):459–67.
31. Ahmed A, Lee KS, Bukhsh A, Al-Worafi YM, Sarker MMR, Ming LC, *et al.* Outbreak of vaccine-preventable diseases in Muslim majority countries. *J Infect Public Health.* 2018;11(2):153–5. doi: <https://doi.org/10.1016/j.jiph.2017.09.007>
32. Althnian A, AlSaeed D, Al-Baity H, Samha A, Dris A Bin, Alzakari N, *et al.* Impact of dataset size on classification performance: an empirical evaluation in the medical domain. *Appl Sci (Switzerland).* 2021;11(2):1–8. doi: <https://doi.org/10.3390/app11020796>
33. Bell A, Rich A, Teng M, Oreskovic T, Bras NB, Mestrinho L, *et al.* Proactive advising: a machine learning driven approach to vaccine hesitancy. 2019 IEEE International Conference on Healthcare Informatics (ICHI), IEEE, 2019. 1–6 pp. doi: <https://doi.org/10.1109/ICHI.2019.8904616>

How to cite this article:

Othman NH, Elamvazuthi I, Rajendram S, Singh HKB, Hussain MHM, Ansari MT, Reddy KVV. Development of a vaccine-hesitancy prediction instrument: Application of machine learning. *J Appl Pharm Sci.* 2026;16(05):187-198. DOI: 10.7324/JAPS.2026.249217