

Evaluating Measles Incidence Rates Using Machine Learning and Time Series Methods in the Center of Iran; 1997-2020

Javad Nazari

Arak University of Medical Sciences

Parnia-Sadat Fathi

Arak University of Medical Sciences

Nahid Sharahi

Arak University of Medical Sciences

Majid Taheri

Arak University of Medical Sciences

Payam Amini

Ahvaz Jondishapour University of Medical Sciences

Amir Almasi-Hashiani (✉ amiralmasi2007@gmail.com)

Arak University of Medical Sciences <https://orcid.org/0000-0003-4434-561X>

Research article

Keywords: Measles, Machine learning, Time series, Infection

Posted Date: July 24th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-45999/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Iranian Journal of Public Health on April 19th, 2022. See the published version at <https://doi.org/10.18502/ijph.v51i4.9252>.

Abstract

Background: Measles is a feverish condition labeled among the most infectious viral illnesses in the globe. Despite the presence of a secure, accessible, affordable and efficient vaccine, measles continues to be a worldwide concern.

Methods: This study uses machine learning and time series methods to assess factors that placed people at a higher risk of measles. This historical cohort study contained the Measles incidence in Markazi Province, the center of Iran, from April 1997 to February 2020. Logistic regression, linear discriminant analysis, random forest, artificial neural network, bagging, support vector machine, and naïve Bayes were used to make the classification. Zero-inflated negative binomial regression for time series was utilized to assess development of measles over time.

Results: The prevalence of measles was 14.5% over the recent 24 years and a constant trend of almost zero cases was observed from 2002 to 2020. The order of independent variable importance were recent years, age, vaccination, rhinorrhea, male sex, contact with measles patients, cough, conjunctivitis, ethnic, and fever. Younger age, less probability of contact and no fever is associated with less odds of zero cases. Only 7 new cases were forecasted for the next two years. Bagging and random forest were the most accurate classification methods.

Conclusion: Even if the numbers of new cases are almost zero during the recent years, it has been showed that age and contact are responsible for non-occurrence of measles. October and May are prone to have new cases for 2021 and 2022.

Background

Measles is among the most infectious disorders of humans that may cause severe illness and adverse symptoms [1]. This disease is caused by measles virus and includes several symptoms such as fever (may be as high as 105°F [40.5°C]), malaise, cough, coryza (nasal mucous membrane inflammation), conjunctivitis, Koplik spots (enanthem, or a rash on the mucous membranes), and maculopapular rash (exanthema, or a skin rash) [1–3].

Measles can quickly spread by sick people's coughs and sneezes and can even transmit by close interaction with mouth or nasal secretions [2]. Studies have reported a significant higher basic reproduction number for measles in comparison to other spreading viruses such as Influenza [4]. Measles was an instance of the relationship between demographic factors and population patterns of the epidemics. This has been demonstrated that birth levels cause differences across multi-annual measles epidemic periods. Case-fatality levels tend to be significant in tropical areas such as Asia and sub-Saharan Africa and grow to 20–30% among disadvantaged groups like refugees [4].

Greater regulation in vaccines has decreased the rate of this infection during the recent years worldwide [5]. However, the United Nations (UN) has warned countries about the increase in the amount of measles

reports worldwide by 48.4 percent in 2019 which is also growing steadily due to inadequate monitoring of the vaccinations, development in anti-vaccination campaigns, economic and political issues surrounding health-care programs [4, 6]. Compared to 2016, the frequency of measles increased by 167 percent in 2018, where America and Africa had the most and least rate respectively [7]. The estimated annual measles occurrence rate in Iran was small, however in 2016, following a rise in the frequency of positive measles reports in 2015, the incidence rate of measles experienced a significant decrease in the next year [8]. As the Eastern Mediterranean region witnessed the largest increase in measles cases in 2019, Iran was granted a measles exclusion status in October 2019 [9]. Although the number of new cases has been diminished significantly in the center of Iran during the recent years, the central parts of Iran such as Markazi province – with the highest risk point in Iranian districts [10]- might find a considerable increase in the number of measles cases in a near future similar to what Brazil, Madagascar, Ukraine, Yemen, Philippines, and Venezuela experienced [6].

To evaluate the future condition of measles, we need to analyze the current data by appropriate statistical tools. Machine learning approaches are helpful to decide about the measles condition of an individual using the information on the demographic and clinical characteristics. However, different techniques result in varying prediction precision, based on the nature of the data and the method selection depends on its accuracy. Moreover, time series techniques are useful to assess the dynamical mechanisms of monthly measles incidence rate over years. Regarding the excess number of zero cases in the recent years, applying appropriate zero-inflated methods can help us to find the most important factors resulting in non-occurrence of measles [11–14].

Removal of measles is a public health concern since travel-related infections will be inevitable as long as the virus continues to circulate in any part of the globe. Therefore, the goal of this study is to classify factors that placed people at a higher risk of measles using different classification methods and assess their impact on the series of measles monthly incidence frequencies using a time series approaches.

Methods

Data

We used the dataset of a historical cohort study, conducted on the Measles incidence in Markazi Province, the center of Iran, from April 1997 to February 2020. The data were extracted from the database of the Vice-chancellor of Health Services, Arak University of Medical Sciences, Markazi, Iran.

The data contained the information about individuals' Measles test results (positive/ negative), gender (male/ female), age (year), location (urban/ rural), any contact with measles patients (yes/ no), ethnic (Iranian/ non-Iranian), and some clinical signal such as rhinorrhea (yes/ no), fever (yes/ no), conjunctivitis (yes/ no), cough (yes/ no), and history of vaccination (yes/ no).

The result of measles test was considered as the binary response variable and independent variables were utilized to classify the cases into two levels of outcome using classification approaches including

Logistic Regression (LR), Linear Discriminant Analysis (LDA), Random Forest (RF), Artificial Neural Network (ANN), Bagging, Support Vector Machine (SVM), and Naïve Bayes were used to make the classification.

Moreover, regarding the nature of monthly measles new cases over the study period, time series models were evaluated. Regarding the excess zeros in the series and the count type of response variable (measles frequency), Zero-Inflated Negative Binomial (ZINB) regression for time series was utilized. The use of negative binomial is due to the presence of overdispersion in the series of observations.

Statistical Analysis

Logistic Regression (LR)

The response variable (measles) follows a binomial distribution and the effect of different predictors on the outcome is assessed via a logit link function. The model formula is as follows:

$$\log\left(\frac{\pi}{1-\pi}\right) = \sum_{i=1}^k \beta_i X_i$$

In this model, π is the probability of measles, X 's are the covariates and B 's are the regression. Odds ratio is used to report the effect of each variable on the outcome [15].

Linear Discriminant Analysis (LDA)

LDA is analogous to LR, and refers the dependent variable to linear predictors and classifies the outcome based on the independent variables. LDA addresses the problem by the conditional likelihood of the factors given the outcome class. This method eliminates the dispersion among the same category cases and optimizes the dispersion between the categories [16]. Standardized coefficients are used to determine the most important independent variables.

Random Forest (RF)

Leo Breiman presented the method for the first time in which the regression tree and classification are combined. Random Forest is a technique in which powerful and quick computations are achieved over large datasets. The dataset is sampled in RF to shape the trees by substitution and at the nodes random sets of predictors are picked. The most significant predictors can be detected by mean decrease Gini and mean decrease accuracy tools. The key variables define the binary result such that the analysis is done with the utmost precision [17].

Artificial Neural Network (ANN)

The approach is focused on the role of the human brain. Multilayer perceptron (MLP) is the most frequently adopted approach of many forms in artificial neural networking. This approach involves input, output, and secret layers, where there are multiple nodes in each row. Through adding a degree of

nonlinearity, an activation mechanism converts the data within each layer into the next one. The input layer is composed of all risk factors that influence the outcome. measles as the binary outcome shows up throughout the output layer. To find the network's optimal results, dynamic nonlinear projection between input and output layers is carried out using the number of nodes [18, 19]. The normalized importance of independent variables is used to find more affecting factors.

Bagging

Bagging is a technique of machine learning which works by combining bootstrapping and aggregating. The number of B bootstrap samples in this method is selected from the training set. The noisy observations are reduced by bootstrapping, and even removed. Those sets must then supply the classifiers. Both sets should also have improved behavior for the classifiers relative to the original collection. This makes bagging strategy a valuable method for creating a stronger classifier when the training set poses noisy observations. This approach provides the importance of independent variables as the order of factors affecting the outcome [20].

Naïve Bayes

This method functions on the basis of the popular theorem of the Bayes and results in straightforward and quick classification. Using the Bayes theorem, the prior likelihood of contributing to each category of the outcome is conditioned on the predictor variables. At the final step, the subject will be assigned to the category with the highest posterior probability [21]. Naïve Bayes provides quality estimate of the attributes to conclude about the significant factors affecting measles [22].

Support Vector Machine (SVM)

The goal of the support vector machine method is to locate a hyperplane in an P-dimensional space (the number of attributes) that separately classifies the binary outcome. There are several different hyperplanes which could be selected to distinguish the two levels of the outcome. The aim is to seek a plane with the maximum range, that is to say the maximum gap between the dependent variable categories. Maximizing the gap from the margins offers sufficient clarification such that new observations can be identified with better trust [23].

Comparing the methods

The Several metrics of sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall accuracy were given to assess the discriminative quality of the computational models. We split the dataset into two training sets (70 percent of cases) and testing sets (30 percent of cases). For each model, we then validated the method 500 times and listed the assessment criterion as the average of the 500 iterations.

Zero-Inflated Negative Binomial (ZINB) regression for time series

This model assumes a negative binomial distribution for the observation with excess zero and fits a two-part model. The first part assesses the impact of predictors on the counts observations such as a usual negative binomial regression and outputs the estimate of coefficients where a logarithm link function is used. The second part of the model uses a logit link function to evaluate the impact of independent variables on non-occurrence of the outcome [24-26].

The independent variables were set as the proportion of a certain level to the sample space of the variables to assess the effect of independent variables categories on the series of measles incidence frequencies over the time. Regarding the variables shown in Table 5, sex (male proportion) indicates the percentage of male in the population of our study. ethnic (Iranian proportion) is the percentage of an Iranian case, location (urban proportion) is the percentage of living in urban areas, vaccination (yes proportion) is the percentage of vaccinated case, contact (yes proportion) is the percentage of contact, fever (yes proportion) is the percentage of fever, cough (yes proportion) is the percentage of cough, rhinorrhea (yes proportion) is the percentage of rhinorrhea, and conjunctivitis (yes proportion) is the percentage of conjunctivitis. Moreover for the purpose of ZINB time series model, age is categorized into six levels as (<1, 1-4, 4-9, 9-15, 15-19, >19) to make the convergence of maximizing the likelihood function possible and for the purpose of ease of interpretation [27].

Software

In this study, all analyses were implemented in R version 3.6.3 using the randomForest, 1071, rpart, CORElearn, ZIM, and rminer packages.

Results

Data description

The data included the information of 2114 cases in which 306 (14.5%) experienced Measles from 1997 to 2020. We used Measles as the binary response variable (Yes/ No). The majority (mode) of reports with positive and negative Measles was in 1999 and 2003 respectively and a constant trend of almost zero cases was observed from 2002 to 2020 (Figure 1).

Characteristics of patients as well as their unadjusted association with the outcome are shown in Table 1. The mean (standard deviation) of cases with and without Measles were 15.71 (7.18) and 12.55 (8.49) years respectively. Those cases with Measles were male (prevalence:15.5%, $p=0.129$), non-Iranian (prevalence:27.2%, $p<0.001$), in rural areas (prevalence:16.6%, $p=0.005$), non-vaccinated (prevalence:21.1%, $p<0.001$), in contact with measles patients (prevalence:18%, $p<0.001$), with fever (prevalence:18%, $p<0.001$), with cough (prevalence: 18.8%, $p<0.001$), with rhinorrhea (prevalence: 20.7%, $p<0.001$) and conjunctivitis (prevalence: 21%, $p<0.001$)

Table 1. The distribution of variables between the categories of measles and the results of Chi-square independent test

Variable	Measles		P-Value
	Mean (\pm SD) Or n (%)		
	No	Yes	
Time (Year)	1382.6 (\pm 4.7)	1378.34 (\pm 1.6)	<0.001
Age	12.55 (\pm 8.49)	15.71 (\pm 7.18)	<0.001
Sex			0.129
Female	787 (86.9)	119 (13.1)	
Male	1021 (84.5)	187 (15.5)	
Ethnic			<0.001
Non-Iranian	75 (72.8)	28 (27.2)	
Iranian	1733 (86.2)	278 (13.8)	
Location			0.005
Rural	888 (83.4)	177 (16.6)	
Urban	920 (87.7)	129 (12.3)	
Vaccination			<0.001
No	679 (78.9)	182 (21.1)	
Yes	1129 (90.1)	124 (9.9)	
Contact			<0.001
No	1064 (88.2)	143 (11.8)	
Yes	744 (82.0)	163 (18.0)	
Fever			<0.001
No	661 (92.3)	55 (7.7)	
Yes	1147 (82.0)	251 (18.0)	
Cough			<0.001
No	913 (90.2)	99 (9.8)	
Yes	895 (81.2)	207 (18.8)	
Rhinorrhea			<0.001
No	902 (92.8)	70 (7.2)	
Yes	906 (79.3)	236 (20.7)	
Conjunctivitis			<0.001
No	942 (92.5)	76 (7.5)	
Yes	866 (79.0)	230 (21.0)	

Performance of the models in predicting Measles

Table 2 shows the performance of seven classification approaches using sensitivity, specificity, positive predictive value, negative predictive value and accuracy obtained from 500 repetition of cross validation strategy. Except for SVM, LR, and NB, other four approaches demonstrated sensitivities higher than 0.50. Bagging and LR approaches sowed the highest specificity and PPV among the methods respectively. In contrast to other models, NB exposed the lowest NPV of 78%. Average on both train and test datasets, the

Bagging and RF methods resulted in higher accuracies and were introduced as the best classifiers among different approaches.

Table 2. The result of different classification methods classifying measles using independent variables followed by 500 repetition of cross validation

Method	Set	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Total accuracy
Bagging	Train	0.94 ± 0.01	0.98 ± 0.002	0.87 ± 0.01	0.99 ± 0.004	0.97 ± 0.002
	Test	0.57 ± 0.05	0.91 ± 0.01	0.50 ± 0.05	0.94 ± 0.01	0.87 ± 0.01
RF	Train	0.86 ± 0.02	0.94 ± 0.003	0.65 ± 0.02	0.98 ± 0.003	0.94 ± 0.003
	Test	0.70 ± 0.06	0.94 ± 0.009	0.50 ± 0.04	0.90 ± 0.01	0.90 ± 0.01
LR	Train	0.35 ± 0.03	0.97 ± 0.004	0.86 ± 0.02	0.84 ± 0.006	0.83 ± 0.008
	Test	0.37 ± 0.08	0.97 ± 0.02	0.93 ± 0.08	0.83 ± 0.02	0.82 ± 0.02
LDA	Train	0.65 ± 0.04	0.87 ± 0.005	0.16 ± 0.02	0.98 ± 0.004	0.86 ± 0.006
	Test	0.60 ± 0.12	0.87 ± 0.01	0.15 ± 0.02	0.98 ± 0.01	0.86 ± 0.01
ANN	Train	0.88 ± 0.02	0.66 ± 0.02	0.21 ± 0.05	0.98 ± 0.06	0.87 ± 0.004
	Test	0.86 ± 0.03	0.63 ± 0.02	0.17 ± 0.03	0.98 ± 0.05	0.85 ± 0.03
Naïve bayes	Train	0.34 ± 0.01	0.94 ± 0.005	0.68 ± 0.02	0.78 ± 0.008	0.76 ± 0.007
	Test	0.34 ± 0.03	0.94 ± 0.01	0.68 ± 0.04	0.78 ± 0.01	0.76 ± 0.01
SVM	Train	0.50 ± 0.02	0.54 ± 0.12	0.15 ± 0.04	0.86 ± 0.03	0.55 ± 0.12
	Test	0.48 ± 0.10	0.53 ± 0.10	0.15 ± 0.02	0.86 ± 0.02	0.52 ± 0.09

RF: Random Forest; LSSVM: Least-squares support-vector machine; LDA: Linear Discernment Analysis; NB: Naive Bayes; LR: Logistic Regression; ANN: Artificial Neural Network; SVM: support-vector machine

The association of Measles and independent variables

Regarding the results shown in table 3 and 4, the classification approaches found almost the same results. Recent years were associated with less number of new cases and time was the most significant variable predicting measles in Markazi province. Age was the second most important factor associated with measles. The unadjusted odds ratio showed that one-year increase in age is associated with 0.02 (95% confidence interval: 0.01-0.03) more likelihood of measles. Vaccination and rhinorrhea were the third and fourth most important affecting variables so that vaccinated cases were 0.08 (the LR adjusted 95% confidence interval for OR: 0.89-0.95) less prone to experience measles and those with rhinorrhea were 1.04 (the LR adjusted 95% confidence interval for OR: 0.99-1.09) less times in risk comparing to those without this sign. Adjusted for other variables, the methods revealed that male cases are more in risk than females so that odds of measles were almost 0.01 more among men. Moreover, any contact with measles patients increased the odds about 65 and 1% in unadjusted and adjusted perspectives respectively. The methods also showed that cough and conjunctivitis have same amount of importance for predicting measles. Ethnic and fever had less influence in comparison to other variables.

Table 3. The importance of independent variables resulted by perfumed methods

Independent Variable Importance	ANN Normalized Importance	Bagging Importance	SVM Importance	NB Quality estimate of the attribute	LDA Standardized Coefficients	RF Mean decrease accuracy
Rhinorrhea	11.5%	43.42	0.05	0.8%	0.08	9.69
Age	34%	139.95	0.22	0.9%	0.05	17.87
Fever	7.6%	18.76	0.05	-0.1%	0.71	13.11
Ethnic (Iran)	13.6%	21.45	0.22	-0.2%	-0.09	7.08
Gender (Male)	6.3%	35.11	0.02	0.5%	0.10	1.24
Conjunctivitis	7%	39.10	0.02	-0.7%	-0.10	10.58
Contact	7.8%	38.25	0.03	-0.6%	-0.09	7.02
Cough	9.5%	36.49	0.03	-0.6%	-0.44	11.04
Urban	12.1%	36.48	0.04	-0.6%	-0.13	4.89
Vaccine	13.1%	47.12	0.13	-0.8%	-0.31	20.08
Time (Year)	100%	169.08	0.19	-1.9%	-0.99	70.40

RF: Random Forest; LDA: Linear Discernment Analysis; CM: Core Model; ANN: Artificial Neural Network; BMI: Body Mass Index

Table 4. Logistic regression analyses for relationship between demographic/clinical factors and measles

Variable	Prevalence of Measles, Mean (\pm SD) Or n (%)	Unadjusted	Adjusted
		OR (95% CI)	OR (95% CI)
Time (Year)	1999.34 (\pm 1.6)	0.97 (0.96-0.98)	0.97 (0.96-0.98)
Age	15.70 (\pm 7.18)	1.02 (1.01-1.03)	1.00 (0.99-0.1.01)
Sex			
Female	119 (13.1%)	1	1
Male	187 (15.5%)	1.02 (0.99-1.05)	1.01 (0.98-1.04)
Ethnic			
Non-Iranian	28 (27.2%)	1	1
Iranian	278 (13.8%)	0.87 (0.82-0.93)	0.91 (0.85-0.97)
Location			
Urban	129 (12.3%)	1	1
Rural	177 (16.6%)	1.04 (1.01-1.07)	1.03 (1.01-1.06)
Vaccination			
No	182 (21.1%)	1	1
Yes	124 (9.9%)	0.89 (0.86-0.92)	0.92 (0.89-0.95)
Contact			
No	143 (11.8%)	1	1
Yes	163 (18.0%)	1.06 (1.03-1.09)	1.01 (0.99-1.03)
Fever			
No	55 (7.7%)	1	1
Yes	251 (18.0%)	1.11 (1.07-1.14)	1.21 (1.14-1.28)
Cough			
No	99 (9.8%)	1	1
Yes	207 (18.8%)	1.09 (1.06-1.12)	1.11 (1.06-1.16)
Rhinorrhea			
No	70 (7.2%)	1	1
Yes	236 (20.7%)	1.14 (1.11-1.17)	1.04 (0.99-1.09)
Conjunctivitis			
No	76 (7.5%)	1	1
Yes	230 (21.0%)	1.14 (1.11-1.18)	1.04 (0.99-1.09)

BMI: Body Mass Index; OR: Odds Ratio; CI: Confidence Interval

Based on the results in Table 5, age is an effective variable both on the frequency of measles and zero inflation part of the model. The frequency of measles patients increases by 2.41 individuals in average as one-level increase in age (95% CI: 1.02 – 5.77). The frequency of measles cases was 9, 38, 40, 111, 64, and 44 for the six levels of <1, 1-4, 4-9, 9-15, 15-19, and >19 respectively. In addition, one-level younger age is associated with 96% less odds of zero cases (OR:0.04, 95%CI: 0.01-0.23). One percent increase the probability of contact, fever and conjunctivitis is associated with 9.5 (95% CI: 6.42 – 14.06), 7.14 (95% CI:

4.83 – 10.57), and 6.47 (95% CI: 4.37 – 9.57) more measles cases respectively. Moreover, the odds ratio of zero measles cases is 9.17 (95% CI: 6.25 – 13.47) and 11.72 (95% CI: 7.92 – 17.34) as one percent decrease in the probability of contact and fever respectively. the negative binomial parameter was estimated as 0.92 in the model. Moreover, the time series model forecasted seven new cases who might be found during the next two years on May and October.

Table 5. The results of Zero-Inflated Negative Binomial time series regression assessing the impact of independent variables on the series of measles observations over the time

Variable	ZINB Regression		Zero-inflated part	
	Exponential (Beta)	95% CI	OR	95% CI
Age (Categorical)	2.41	1.02 - 5.77	0.04	0.01 - 0.23
Sex (Male proportion)	0.97	0.19 - 4.88	1.45	0.08 - 24.88
Ethnic (Iranian proportion)	0.19	0.03 - 1.05	0.81	0.02 - 26.78
Location (Urban proportion)	0.35	0.08 - 1.55	0.31	0.02 - 5.05
Vaccination (Yes proportion)	2.48	0.60 - 10.30	1.05	0.10 - 11.12
Contact (Yes proportion)	9.50	6.42 - 14.06	0.10	0.07 - 0.16
Fever (Yes proportion)	7.14	4.83 - 10.57	0.08	0.06 - 0.13
Cough (Yes proportion)	0.09	0.01 - 9.33	0.02	0.01 - 1.08
Rhinorrhea (Yes proportion)	0.10	0.01 - 1.60	0.42	0.04 - 5.12
Conjunctivitis (Yes proportion)	6.47	4.37 - 9.57	1.22	0.12 - 12.60

95% CI: 95% Confidence Interval; ZINB: Zero-inflated negative binomial

Discussion

Our data showed a significant change point in 2003 when the health policy-makers started one of the major immunization campaigns against measles and rubella in three phases including catch-up, keep-up, and follow-up. This operation yielded to an impressive reduction in measles incidence rate [9]. In other words, the large rate of immunization coverage in all Iranian cities and villages is among the most significant reasons for the reduced incidence of measles based on the WHO risk evaluation method [28]. However, The World health organization reported peaks in places with large total vaccination coverage over recent months, such as the United States of America, Thailand and Tunisia, as the infection has quickly spread by many groups of unvaccinated people [29]. Since vaccination coverage in many countries is suboptimal, measles seems to be spread across countries. To reach the elimination target, several governments have to bring incremental progress in the scope of their regular childhood immunization systems and to close immunity gaps between different age groups who have skipped out on vaccination opportunities [30].

The individuals between 9 and 19 years old had the higher rate of measles in our observations. While measles is typically a childhood illness, infection occurs in individuals of any age. Age might be confounded with vaccination so that unvaccinated, partly vaccinated, or weakened immunity cases in any age are in danger. Particularly, unvaccinated youths are at the greatest risk. Depending on the local immunization procedures, age-specific attack levels could be higher in vulnerable babies younger than 12 months, school-age children or young adults [31].

We found out that men are more in risk comparing to women. That might be due to lack of balanced job development between the sexes so that men are busier in social jobs which need more contact and communication. Southwest of Tehran, Markazi Province is flooded with Afghan immigrants, mostly men, where the plurality of cases come from rural locations and refugee communities with higher incidence areas [10]. At the other side, long-term political instability, religious conflicts, violence, and inadequate healthcare services have impacted Iran's neighboring countries and the emergence of almost 6 percent of Iran's population has placed increased strain at Iran's health care framework [32]. It has been reported that aspects like neighboring with the capital, situated on the key road to the western regions of the country, a substantial majority of conventional service and manufacturing systems that are ideal for workers contribute to Markazi province being able to handle a significant number of non-Iranian citizens [32]. This fact is a potential and considerable factor for increasing the risk of any contact with infected cases [33] which is responsible for the occurrence of measles based on zero-inflated analysis in our data. The rise in perceived vulnerability of higher-risk communities was largely attributed to inadequate standard of monitoring, focus due to weak execution of the strategy, and the inclusion of vulnerable population groups [10].

The results of our study was achieved using classification and time series methods. The manner under which the predictors influence the result is important for deciding the correct technique of classification, in addition to the distribution of the result groups. Therefore, conflict may be found in conducting classification methods in different fields. After all, it's recommended to replicate the cross validation phase to verify the findings. In our study, the cross validation technique was carried out over hundreds of repetitions to estimate the measles outcome. Machine learning and stochastic process approaches are widely used in different aspects such as random for minimizing false negatives of measles prediction model [34], naïve bayes for predicting immunize-able, logistic regression for assessing the increased measles seronegativity in adults with major depressive disorder diseases [35], artificial neural network for forecasting measles coverage using artificial neural network [36], zero-inflated models to investigate the transmission potential of modified measles during an outbreak in Japan [37], and finite-range time series of counts with an application on measles data [38].

Several limitations of this historical cohort study should be noted. For example, many of other potential predictors have not been recorded by the center and could have significant impact on the classification and forecasting procedures in our study.

Although forecasting analysis confirmed the continuing procedure of constant low number of new cases for the next two years in addition to this fact that the number of new measles cases had dramatically decreased over the recent decade in Iran which yielded this country to receive the elimination certificate, there are many European and Americas regions who have lost their certificate in 2018 and 2019 [9]. This is an alarm for the health policy makers to continue the restrictions on measles infection such as preventing contact with suspicious cases and strict oversight of immigration systems from neighbor countries which are the most deterministic factors to continue the measles elimination strategy.

Conclusion

Even if the numbers of new cases are almost zero during the recent years, it has been show that age and contact are responsible for non-occurrence of measles. October and May are prone to have new cases for 2021 and 2022.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Arak University of Medical Sciences. In this study, we used the exiting registered data in the Health deputy of Arak University of Medical Sciences.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and analyzed during the current study are available to be collected from the corresponding author on reasonable request.

Competing interests

All authors declared no conflict of interest.

Funding

This study was funded by Vice-chancellor of Research, Arak University of Medical Sciences, Markazi, Iran. The funder has no role in data analysis, interpretation and manuscript drafting.

Authors' contributions

Study conception and design: JN, PSF, NS, MT and AAH. Data collection, statistical expertise, analysis and interpretation of data: PA, AAH, JN, NS, PSF, MT. Manuscript preparation, supervision, administrative

support and critical revision of the paper: PA, AAH, JN, NS, PSF, MT. All authors read and approved the final manuscript.

Acknowledgements

The authors thank all health care providers who collect the data. We sincerely appreciate the scientific support of Vice-chancellor of Research, Arak University of Medical Sciences, Markazi, Iran.

References

1. World Health Organization: **Global measles and rubella strategic plan: 2012**. 2012.
2. Bester JC: **Measles and measles vaccination: a review**. *JAMA pediatrics* 2016, **170**(12):1209-1215.
3. Perkins A: **Measles: Resurgence of a once eliminated disease**. *Nursing made Incredibly Easy* 2019, **17**(5):26-31.
4. Congera P, Maraolo AE, Parente S, Moriello NS, Bianco V, Tosone G: **Measles in pregnant women: a systematic review of clinical outcomes and a meta-analysis of antibodies seroprevalence**. *Journal of Infection* 2019.
5. Li S, Ma C, Hao L, Su Q, An Z, Ma F, Xie S, Xu A, Zhang Y, Ding Z: **Demographic transition and the dynamics of measles in six provinces in China: A modeling study**. *PLoS medicine* 2017, **14**(4).
6. Dyer O: **Measles: alarming worldwide surge seriously threatens children, says UN**. In.: British Medical Journal Publishing Group; 2019.
7. Patel MK, Dumolard L, Nedelec Y, Sodha SV, Steulet C, Gacic-Dobo M, Kretsinger K, McFarland J, Rota PA, Goodson JL: **Progress Toward Regional Measles Elimination—Worldwide, 2000–2018**. *Morbidity and Mortality Weekly Report* 2019, **68**(48):1105.
8. Mohammadbeigi A, Zahraei SM, Sabouri A, Asgarian A, Afrashteh S, Ansari H: **The spatial analysis of annual measles incidence and transition threat assessment in Iran in 2016**. *Medical Journal of The Islamic Republic of Iran (MJIRI)* 2019, **33**(1):788-793.
9. Namaki S, Gouya MM, Zahraei SM, Khalili N, Sobhani H, Akbari ME: **The elimination of measles in Iran**. *The Lancet Global Health* 2020, **8**(2):e173-e174.
10. Mohammadbeigi A, Zahraei SM, Asgarian A, Afrashteh S, Mohammadsalehi N, Khazaei S, Ansari H: **Estimation of measles risk using the World Health Organization Measles Programmatic Risk Assessment Tool, Iran**. *Heliyon* 2018, **4**(11):e00886.
11. Shumway RH, Stoffer DS: **Time series analysis and its applications: with R examples**: Springer; 2017.
12. Molnar C: **Interpretable machine learning**: Lulu. com; 2019.
13. Alpaydin E: **Introduction to machine learning**: MIT press; 2020.
14. Shamma N, Mohammadpour M, Shirozhan M: **A time series model based on dependent zero inflated counting series**. *Computational Statistics* 2020:1-21.
15. Agresti A, Kateri M: **Categorical data analysis**: Springer; 2011.

16. Izenman AJ: **Linear discriminant analysis**. In: *Modern multivariate statistical techniques*. edn.: Springer; 2013: 237-280.
17. Cutler A, Cutler DR, Stevens JR: **Random forests**. In: *Ensemble machine learning*. edn.: Springer; 2012: 157-175.
18. Dreiseitl S, Ohno-Machado L: **Logistic regression and artificial neural network classification models: a methodology review**. *Journal of biomedical informatics* 2002, **35**(5-6):352-359.
19. Maroufizadeh S, Amini P, Hosseini M, Almasi-Hashiani A, Mohammadi M, Navid B, Omani-Samani R: **Determinants of Cesarean Section among Primiparas: A Comparison of Classification Methods**. *Iranian journal of public health* 2018, **47**(12):1913.
20. Xiao T, Zhu J, Liu T: **Bagging and boosting statistical machine translation systems**. *Artificial Intelligence* 2013, **195**:496-527.
21. Rish I: **An empirical study of the naive Bayes classifier**. In: *IJCAI 2001 workshop on empirical methods in artificial intelligence: 2001*; 2001: 41-46.
22. Tapak L, Shirmohammadi-Khorram N, Amini P, Alafchi B, Hamidi O, Poorolajal J: **Prediction of survival and metastasis in breast cancer patients using machine learning classifiers**. *Clinical Epidemiology and Global Health* 2019, **7**(3):293-299.
23. Amini P, Ahmadiania H, Poorolajal J, Amiri MM: **Evaluating the high risk groups for suicide: A comparison of logistic regression, support vector machine, decision tree and artificial neural network**. *Iranian journal of public health* 2016, **45**(9):1179.
24. Yang M, Zamba GK, Cavanaugh JE: **Markov regression models for count time series with excess zeros: A partial likelihood approach**. *Statistical Methodology* 2013, **14**:26-38.
25. Yang M, Cavanaugh JE, Zamba GK: **State-space models for count time series with excess zeros**. *Statistical Modelling* 2015, **15**(1):70-90.
26. Yang M, Zamba G, Cavanaugh J: **ZIM: Zero-inflated models for count time series with excess zeros**. *R package version* 2014, **1**(2).
27. Xiong Y, Wang D, Lin W, Tang H, Chen S, Ni J: **Age-related changes in serological susceptibility patterns to measles: results from a seroepidemiological study in Dongguan, China**. *Human vaccines & immunotherapeutics* 2014, **10**(4):1097-1103.
28. Zahraei SM, Eshрати B, Gouya MM, Mohammadbeigi A, Kamran A: **Is there still an immunity gap in high-level national immunization coverage, Iran?** *Archives of Iranian medicine* 2014, **17**(10):0-0.
29. World Health Organization: **New measles surveillance data for 2019**. Retrieved August 2019, **24**:2019.
30. Prevention ECfD, Control: **Monthly measles and rubella monitoring report, September 2018**. In.: ECDC Stockholm; 2018.
31. Hughes SL, Bolotin S, Khan S, Li Y, Johnson C, Friedman L, Tricco AC, Hahné SJ, Heffernan JM, Dabbagh A: **The effect of time since measles vaccination and age at first dose on measles vaccine effectiveness—A systematic review**. *Vaccine* 2020, **38**(3):460-469.

32. Soleimanpour S, Hamed Asl D, Tadayon K, Farazi AA, Keshavarz R, Soleymani K, Seddighinia FS, Mosavari N: **Extensive genetic diversity among clinical isolates of Mycobacterium tuberculosis in central province of Iran.** *Tuberculosis research and treatment* 2014, **2014**.
33. Velayati AA, Farnia P, Mirsaeidi M, Reza Masjedi M: **The most prevalent Mycobacterium tuberculosis superfamilies among Iranian and Afghan TB cases.** *Scandinavian journal of infectious diseases* 2006, **38**(6-7):463-468.
34. Ahmad WMTW, Ab Ghani NL, Drus SM: **Minimizing False Negatives of Measles Prediction Model: An Experimentation of Feature Selection Based On Domain Knowledge and Random Forest Classifier.** *International Journal of Engineering and Advanced Technology* 2019, **9**(1):3411-3414.
35. Ford B, Yolken R, Dickerson F, Teague T, Irwin M, Paulus M, Savitz J: **Increased measles seronegativity in adults with major depressive disorder.** *Brain, Behavior, and Immunity* 2017, **66**:e18.
36. Nasir JA, Imran M, Zaidi SAA: **FORECASTING MEASLES COVERAGE USING ARTIFICIAL NEURAL NETWORK.** *Journal of University Medical & Dental College* 2018, **9**(4):25-31.
37. Mizumoto K, Kobayashi T, Chowell G: **Transmission potential of modified measles during an outbreak, Japan, March–May 2018.** *Eurosurveillance* 2018, **23**(24).
38. Yang K, Wang D, Li H: **Threshold autoregression analysis for finite-range time series of counts with an application on measles data.** *Journal of Statistical Computation and Simulation* 2018, **88**(3):597-614.

Figures

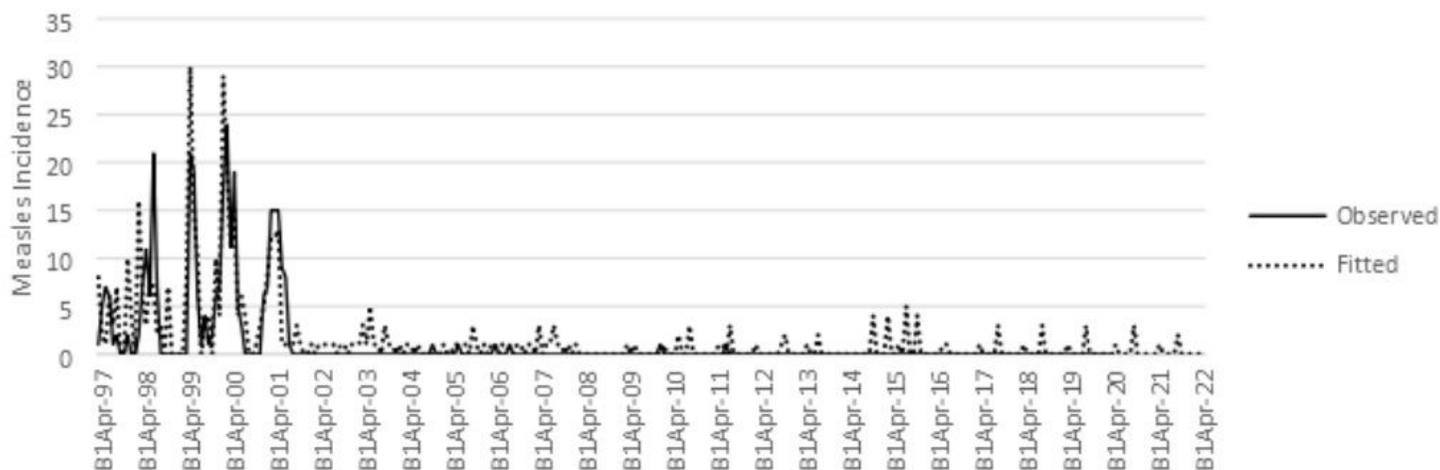


Figure 1

Measles incidence