UDC 004.942: 614:4 **doi: 10.32620/reks.2024.4.01**

Mykola BUTKEVYCH, Dmytro CHUMACHENKO

National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine

TIME SERIES ANALYSIS OF LEPTOSPIROSIS INCIDENCE FOR FORECASTING IN THE BALTIC COUNTRIES USING THE ARIMA MODEL

Leptospirosis, a zoonotic disease with significant public health implications, presents considerable forecasting challenges due to its seasonal patterns and environmental sensitivity, especially in under-researched regions like the Baltic countries. This study ***aimed*** *to develop an ARIMA-based forecasting model for predicting leptospirosis incidence across Estonia, Latvia, and Lithuania, where current disease data are limited and variable. This study aims to investigate the epidemic process of leptospirosis, while its* ***subject*** *focuses on applying time series forecasting methodologies suitable for epidemiological contexts.* ***Methods:*** *The ARIMA model was applied to each country to identify temporal patterns and generate short-term morbidity forecasts using confirmed leptospirosis case data from the European Centre for Disease Prevention and Control from 2010 to 2022.* ***Results.*** *The model's performance was assessed using the Mean Absolute Percentage Error (MAPE), revealing that Lithuania had the most accurate forecast, with a MAPE of 6.841. The accuracy of Estonia and Latvia was moderate, likely reflecting case variability and differing regional epidemiological patterns. These results demonstrate that ARIMA models can effectively capture general trends and provide short-term morbidity predictions, even within diverse epidemiological settings, suggesting ARIMA's utility in low-resource and variable data environments.* ***Conclusions.*** *The scientific novelty of this study lies in its application of ARIMA modelling to leptospirosis forecasting within the Baltic region, where comprehensive time series studies on the disease are scarce. From a practical perspective, this model offers a valuable tool for public health authorities by supporting targeted interventions, more efficient resource allocation, and timely response planning for leptospirosis and similar zoonotic diseases. The ARIMA model's adaptability and straightforward application across countries demonstrate its potential for informing public health decision-making in settings with limited data on disease patterns. Future research should expand on this model by developing multivariate forecasting approaches incorporating additional factors to refine the model's predictive accuracy. This approach could further improve our understanding of leptospirosis dynamics and enhance intervention strategies.*

Keywords: *epidemic model; epidemic process; epidemic simulation; simulation; Leptospirosis; ARIMA.*

Introduction

Leptospirosis, a zoonotic disease engendered by bacteria of the Leptospira genus, has emerged as a formidable yet frequently marginalized public health challenge on a global scale [1]. The disease primarily propagates through exposure to the urine of infected animals or contamination from infected water and soil. Given its expansive range of animal hosts and vast array of human clinical presentations, this disease presents an intricate epidemiological puzzle.

The World Health Organization (WHO) approximates an annual occurrence of 1.03 million cases and 58,900 fatalities linked to leptospirosis, primarily concentrated in marginalized populations and regions with lower access to medical care [2]. Nonetheless, it is essential to consider these figures as likely underrepresentations due to substantial barriers to accurate reporting and diagnosis of the disease [3].

Additionally, the distribution and incidence of the disease are shaped by a multiplicity of environmental and socio-economic determinants, including urbanization, climate change, and occupational exposure, among others [4]. Despite escalating climatic shifts and burgeoning urbanization, it is plausible to anticipate further escalation in the prevalence of leptospirosis [5].

Recent outbreaks in regions previously untouched by the disease underscore the geographical expansion of the disease, thereby underscoring the exigency for reliable, precise, and efficient forecasting models. Such models enable effective preventive strategies, optimized allocation of health resources, and prompt response mechanisms to potential outbreaks, thereby curtailing the morbidity and mortality rates associated with the disease.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

The leptospirosis situation in the Baltic countries of Estonia, Latvia, and Lithuania presents a complex and nuanced landscape [6]. Although these countries have not traditionally experienced the brunt of the disease compared with other global regions, recent epidemiological trends indicate a potential upsurge in incidence rates. These countries are characterized by a temperate climate, rich biodiversity, and substantial rural and agricultural sectors, harboring conducive conditions for the proliferation of Leptospira bacteria, fostering the persistence and potential escalation of the disease.

Estonia, punctuated by an extensive network of water bodies and high rodent populations, particularly in rural locales, presents considerable risk factors for leptospirosis transmission [7]. Similarly, with its robust agricultural sector and a large semi-rural population, Latvia poses specific challenges in disease control and prevention [8]. Although slightly superior in terms of incidence rates, Lithuania has reported sporadic cases predominantly correlated with occupational and recreational exposure [9].

The epidemiological disposition of these countries mirrors the previously delineated global context, as factors such as urbanization, climate change, and socio-economic circumstances play significant roles. With these nations experiencing a surge in urbanization and the pervasive impacts of climate change, environments conducive to the propagation of Leptospira bacteria are becoming increasingly prevalent. In tandem with limited awareness of the disease among the general population and occasionally within the medical fraternity, this contributes to underreporting and misdiagnosis, paralleling the global context.

The simulation of epidemic processes has immense potential to influence decision-making and shape public health policies. In the context of multifaceted diseases such as leptospirosis, these models can serve as potent instruments for public health professionals and policymakers [10].

A prominent advantage of such simulation models lies in the informed decision-making they facilitate [11]. By offering robust quantitative data on disease dynamics, models like ARIMA empower policymakers to make decisions anchored in empirical evidence rather than depending on anecdotal or antiquated information [12]. This aspect is especially pivotal in controlling diseases like leptospirosis, where the epidemiological scenario is in constant flux and is molded by various determinants, including urbanization, climate change, and socio-economic circumstances.

Another dimension of the simulation model is its predictive prowess. By scrutinizing patterns and trends in historical data, these models can forecast future outbreaks or disease surges. In leptospirosis, which exhibits significant seasonal variations and can be subject to abrupt outbreaks, this predictive capability can enable health authorities to implement proactive measures, thereby preempting potential health crises [13].

Epidemic simulation models play an important role in the allocation of public health resources. In an environment where these resources are often limited, effective deployment is paramount for disease control and prevention. By identifying periods and high-risk areas, these models can inform strategies that concentrate resources where they are most needed [14].

Epidemic simulations significantly contribute to the development of health policy [15]. They provide policymakers with a platform to test various strategies in simulated environments before deploying them in real-world situations [16]. This approach minimizes risks and enhances the potential for successful disease control by enabling a proactive rather than reactive approach to public health management.

The simulation of epidemic processes can deliver a scientifically rigorous, predictive, and proactive paradigm for public health decision-making and policy development. This is of utmost importance when addressing complex diseases like leptospirosis, and helps reduce its global burden and manage its regional manifestations effectively.

Thus, this study aimed to develop an ARIMA-based forecasting model to predict the incidence of leptospirosis in Baltic countries and improve the accuracy of the prediction. This research is focused on the leptospirosis epidemic process. The research subjects were methods and models for epidemic process simulation.

To achieve the aim of the research, the following tasks were formulated:

1. Methods and models of the Leptospirosis epidemic process should be analyzed.

2. Data on Leptospirosis in Baltic countries should be analyzed.

3. A simulation model for the leptospirosis epidemic process based on the ARIMA method should be developed.

4. Tuning and verification of the proposed model were performed.

5. Estimating the Leptospirosis epidemic process dynamics model in Baltic countries is needed.

7. Results of the experimental study should be analyzed.

The contributions of this research are two-fold. First, developing models based on the ARIMA method will allow estimation of the accuracy of the method applied to the simulation of the Leptospirosis epidemic process. Second, using the Leptospirosis epidemic process dynamics data from Baltic countries will allow us to investigate the epidemic process's character.

In this paper, section 1, namely, the current research

analysis, provides the current state of leptospirosis simulation methods and models. Section 2, namely, Materials and Methods, provides an overview of the ARIMA model and metrics of the model's performance evaluation. Section 3 presents the data analysis and results of forecasting leptospirosis dynamics in Baltic countries using the developed model. The discussion section presents the obtained results. The conclusions describe the outcomes of the investigation.

The current research is part of a comprehensive information system for assessing the impact of emergencies on the spread of infectious diseases described in [17].

1. The current research analysis

This scholarly discourse concerning the simulation models for Leptospirosis reveals a growing recognition of their value in addressing this multifaceted disease. Existing research illustrates the utility of models like ARIMA in forecasting disease trajectories, thus aiding preemptive public health planning. These studies reflect an interdisciplinary approach, converging epidemiology, biostatistics, and computational biology to effectively predict Leptospirosis's incidence and spread of leptospirosis. Simulation models offer valuable insights into the disease's temporal and spatial patterns, including seasonal trends and geographical hotspots, which are integral for tailoring targeted, timely interventions. Nevertheless, while progress has been made, there is a distinct need for further research, particularly in regions like the Baltic countries where the disease's epidemiological footprint is still being ascertained. Incorporating nuanced local factors such as climate, biodiversity, and socio-economic determinants into these models could further enhance their predictive accuracy and utility in these contexts.

In study [18], an innovative Bayesian method for inference developed explicitly for the Zero-Modified Poisson (ZMP) regression model is presented, exhibiting commendable flexibility in analyzing count data, regardless of the presence of inflation or deflation of zeros in the sample. The proposed methodology incorporates a broad class of prior densities based on an information matrix for model parameters. The method employs a sensitivity study to identify compelling cases that could modify the results. The method uses Kullback-Leibler divergence as a measure, and simulation studies further bolster the findings. The model's application on real datasets of leptospirosis notifications from Bahia State in Brazil further validates its utility. In conclusion, ZMP regression models offer significant benefits in analyzing zero-inflated or zero-deflated datasets and underscore the development of a Bayesian approach based on the information matrix prior. In addition, it develops an effective measure based on the Kullback–Leibler divergence for this model. The research reveals that in real-world applications, the Human Development Index (HDI) covariate is a significant factor in explaining leptospirosis notifications, as leptospirosis notification probability increases with the HDI. However, it also warns that a low number of leptospirosis notifications in any city does not signify an absence of the disease but could reflect the health system's capacity to identify disease cases.

The present paper [19] introduces a novel methodology for numerical simulations of a newly designed fractional order Leptospirosis model (FOLM) that leverages the capabilities of stochastic numerical supervised neural networks. This research constitutes an innovative numerical examination of the Leptospirosis model, categorized into five dynamics. The problem of biological FOLM is approached by considering different fractional-order derivative values. The numerical formulations of the FOLM are produced using supervised neural networks (SNNs), and the computational performances are evaluated through the lens of Levenberg-Marquardt backpropagation (LVMBP), also known as SNNs-LVMBP. The validity of the proposed approach was assessed by comparing the obtained solutions to reference solutions. The statistics revealed that the certification and learning processes accounted for 74% and 13% of the investigation, respectively.

The authors of this paper [20] employed the Susceptible-Infected-Recovered (SIR) model to better understand leptospirosis transmission dynamics, a globally prevalent zoonosis that is typically transmitted by rodents and often leads to fatal outcomes in humans. This study is particularly prone to outbreaks following heavy rainfall and flooding and scrutinizes the factors influencing these transmission dynamics. The model identifies disease-free and endemic equilibrium points from the proposed model and conducts a local stability analysis for each point. The paper also encompasses bifurcation analysis and numerical solutions of the model, noting strong concurrence between theoretical discoveries and numerical simulations. Crucially, the research identifies the natural death rate of the rat population as a significant factor for Leptospirosis control, alongside the basic reproduction number, which holds a vital role in the epidemiology of this disease.

The paper [21] proposes an innovative framework for managing rodent-borne Leptospirosis using optimal control mathematical model theory, addressing the limitations of traditional control methods, such as rodenticide application and habitat management. Leptospirosis, primarily contracted through interaction with animals or environments contaminated with leptospires in animal urine, presents significant control challenges due to the complexity and cost of managing reservoir populations. Informed by empirical data from Salvador, Brazil, the study devised an age-structured model for leptospire

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

infection in the Norwegian rat population, extending it to include two temporary control measures, namely, rodenticide and resource reduction, and two permanent control measures, reducing rat carrying capacity and leptospire lifespan in the environment. The optimal control theory is applied to determine ideal time-dependent controls while factoring in the cost of control measures and the societal "cost" of infection. The results suggest that permanent controls can decrease leptospiral carriage prevalence in rodent populations, and temporary controls can effectively reduce the number of infected rats, thereby mitigating human infection risk. While this study focused on the Norway rat, its approach applies to other disease systems with animal and environmental reservoirs, providing a valuable tool for informed decision-making in public health.

The study [22] of the newly introduced piecewise classical-global and classical-fractional operators is applied to investigate the dynamics of the Leptospirosis disease model. This research examined the existence and uniqueness of solutions to the piecewise derivatives associated with the model by employing the piecewise iterative Newton polynomial method to derive an approximate solution. A numerical scheme was also established for the piecewise Leptospirosis model with integer and fractional orders. We observed improved dynamics and crossover behaviors based on the simulation results for both operators. It was found that the recovered human population would gradually decline over 350 days, implying that the disease would eventually dissipate. Moreover, a high value of β decreased the susceptibility of the rat and human populations, which eventually became stable. Compared to high β values, the infected human population decreased and stabilized more rapidly.

The study [23] developed and scrutinized a compartmental mathematical model to explore the influence of rodent-borne Leptospirosis on human populations, taking into account the disease's pathogenic agents in the environment and the incidence rate of human infection due to the interaction between infected rodents and the environment. The model's basic properties, equilibrium points, and stability analysis are investigated, with the basic reproduction number R0 is derived using the next-generation matrix method. Stability analysis reveals that the disease-free equilibrium is globally asymptotically stable if $R0<1$ and unstable otherwise, and the model exhibits forward bifurcation. Sensitivity analysis identifies key parameters influencing model outcomes. Numerical simulations using the fourth-order Runge-Kutta method further demonstrated the model's stability behavior and the effect of human transmission, recovery, and rodent mortality rates on the model's dynamics. Results indicate that the trajectories of the model solutions evolve toward the unique endemic equilibrium over time when $R0>1$ and that reducing transmission rates, increasing recovery rates, and controlling the rodent population significantly mitigate the spread of disease.

The authors of the paper [24] sought to explore the connection between weather parameters at different time lags and leptospirosis occurrence in Malaysia, where the incidence of leptospirosis has increased. Leveraging data mining and machine learning techniques, the study employs exploratory data analysis (EDA) to determine optimal time lags for rainfall and temperature. Further, based on backpropagation training and optimized hidden layers and nodes, an artificial neural network (ANN) model was designed to classify the selected features into disease occurrence and non-occurrence. The study revealed a strong correlation between leptospirosis occurrence and weekly average temperature at a lag of 16 weeks and weekly rainfall amount at 12...20 weeks. The ANN model, which was developed using these selected features, demonstrated high levels of accuracy, sensitivity, and specificity, increasing the accuracy of the predictive model by 13.30 ... 31.26 % from the baseline models.

The present study [25] focused on developing a cellular automata (CA)-based computational model to illustrate the spread of Leptospirosis, a disease typically transferred from bovine rats to humans. The researchers used the Susceptible-Infective-Recovered-Susceptible (SIRS) model and innovatively incorporated a voting-based rule to enhance the traditional CA rule set. By conducting simulations using actual data from Leptospirosis infections in Thailand during the years 2000 and 2001, the model demonstrated remarkable accuracy, closely aligning with real-time infection data. This implies the viability of the CA-based model and the introduced voting-based rule for realistically capturing the dynamics of Leptospirosis transmission.

In study [26], a mathematical model representing the transmission of the infectious disease Leptospirosis was examined by employing a system of nonlinear ordinary differential equations. Acknowledging the inherent difficulty of obtaining an exact solution for this system, the authors use He's homotopy perturbation method (HPM) to derive an approximate solution. The HPM results are compared with those obtained using the Runge-Kutta fourth-order (RK4) method. Illustrative plots are included to demonstrate the method's simplicity and reliability, thereby supporting the utility of the homotopy perturbation approach for analyzing complex disease transmission models like Leptospirosis.

The paper [27] scrutinized the dynamics of Leptospirosis, a public health issue predominantly spread by rodents, using a novel SI-SIR model that deviates from traditional models by assuming a logistic growth pattern in the rodent population instead of the typical exponential growth. This study offers a comprehensive exploration of the model's equilibrium stability. The equation establishes an equation for the basic reproduction number, R0,

which is determined by the rodent infection rate, birth rate, and environmental carrying capacity. This study identifies a critical threshold for the environmental carrying capacity that dictates disease extinction or persistence. The proposed method further inspects the sensitivity of R0 and proposes a method to gauge the impact of various control measures on infection dynamics. Numerical simulations are presented to provide a tangible representation of the theoretical results.

Table 1 presents the analysis of the models of Leptospirosis propagation.

The various studies analyzed illustrate the increasing importance and application of simulation and mathematical models in predicting the spread and control of Leptospirosis, a complex and significant public health issue. These papers employ various models that offer diverse perspectives for understanding disease dynamics.

Overall, these studies highlight the growing importance of utilizing mathematical and computational modeling in epidemiology and disease control and underscore the need for further research and method development, particularly in regions where the disease's epidemiological footprint is still being ascertained. The models also emphasize the importance of local factors such as climate, biodiversity, and socioeconomic determinants to enhance their predictive accuracy and utility.

2. Materials and Methods

2.1. ARIMA model

The Autoregressive Integrated Moving Average (ARIMA) model, which was introduced by Box and Jenkins, is an esteemed method for analyzing and predicting time series data [28]. This model combines the concepts of autoregression (AR), differencing (I), and moving average (MA) to identify systematic patterns in the data and use them for future forecasting. It is frequently used across various disciplines, from finance and economics to public health and environmental studies, due to its flexibility and applicability.

Conceptually, the ARIMA model is a blend of three parts:

1. Autoregressive (AR) Component (p): The AR component represents the dependency between an observed value and its initial values. An AR term of order p can be expressed as a linear function of the most recent value of p in the time series.

2. Differing (I) Component (d): Differing de-trends the time series and ensures stationarity, which is an essential assumption in ARIMA models. The idea behind differencing is to consider the changes between one observation and the next or over a defined number of lagged observations.

3. Moving Average (MA) Component (q): The MA component represents the dependency between an observed value and residual error from a moving average model applied to lagged observations.

The mathematical representation of an ARIMA model can be given as:

$$\mathrm{ARIMA}(p,d,q) = c + \sum_{i=1}^{p} \varphi_i x_{i-1} + \epsilon_i + \\ + \sum_{i=1}^{q} \theta_i \epsilon_{i-1}, \quad (1)$$

where x_i are a stationary variable;

c is constant;

φ_i are autocorrelation coefficients;

ϵ_i are white noise with zero mean;

θ_i are weights;

θ_0 is assumed to be 1.

The parameters p and q are referred to as the order of autoregressive and moving average. Using ARIMA allows you to make predictions on non-stationary data by introducing integration into the model. This is achieved by considering differences.

In the application of epidemiology, ARIMA models can be used to predict disease patterns, such as the incidence of Leptospirosis, using historical data. The time series data represent the logged historical incidence data of Leptospirosis.

The primary steps in implementing an ARIMA model involve ensuring the stationarity of the series, identifying the appropriate values of p, d, and q using autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, and then estimating the parameters for the identified model. Model diagnostics should be performed to validate the model fit before its use in forecasting future values.

The ARIMA model is a valuable tool in the field of disease forecasting and contributes significantly to public health planning and disease control. By accurately predicting the future incidence of diseases like Leptospirosis, we can better manage resources, implement preventive measures, and mitigate the impact on public health.

Advantages of ARIMA model:

1. ARIMA models are relatively easy to understand and interpret compared to more complex machine learning models. The model parameters (p, d, q) have specific interpretations related to the underlying temporal structure of the data.

2. ARIMA models are versatile and can be applied to various time series data, provided the series is stationary or can be made stationary through differencing.

3. ARIMA models are effective for short-term forecasting and can often provide robust predictions even when the data contain random noise.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Table 1

Overview of the Leptospirosis epidemic process models

Paper	Task	Method	Findings
Conceicao K.S. [18]	To analyze the frequentist properties of the Bayesian estimators of leptospirosis notifications in cities of the Bahia State (Brazil)	Bayesian approach for the ZMP regression model.	The study concluded that the ZMP regression models are highly beneficial in analyzing zero-inflated or zero-deflated datasets and identified the Human Development Index (HDI) as a significant factor in explaining leptospirosis notifications, asserting that leptospirosis notification probability increases with the HDI.
Mukdasai, K. [19]	To present numerical simulations of a newly designed fractional-order leptospirosis model using stochastic numerical supervised neural networks.	Supervised neural networks	The study established the efficacy, stability, precision, reliability, and fitness of the proposed method, with statistics showing 74% and 13% for certification and learning respectively.
Bhalraj, A. [20]	To improve our understanding of Leptospirosis transmission dynamics worldwide	Compartmental model	The natural death rate of the rat population, alongside the basic reproduction rate, plays a pivotal role in controlling the outbreak of Leptospirosis.
Minter, A. [21]	To develop an optimal control mathematical model that effectively manages the spread of leptospirosis	Compartmental model	This study concludes that both permanent and temporary controls can effectively reduce the prevalence of leptospiral carriage in the rodent population and the number of infected rats, thereby lowering the risk of human infection.
Qu, H. [22]	To examine the dynamics of leptospirosis using piecewise classical global and classical fractional operators	Compartmental model	The simulation results showed improved dynamics and crossover behaviors, with a decline in the recovered human population over 350 days and a decrease and stabilization in the susceptible rat and human populations with high β values, indicating potential disease dissipation.
Engida, H. A. [23]	To investigate the impact of rodent-borne leptospirosis on human populations, considering the incidence rate of human infection and the quantity of pathogenic agents in the environment.	Compartmental model	The study revealed that the disease-free equilibrium is globally asymptotically stable if the basic reproduction number is less than one and unstable otherwise and that decreasing transmission rates, increasing recovery rates, and controlling the rodent population have a significant role in reducing the spread of the disease in the population.
Rahmat, F. [24]	To understand the relationship between leptospirosis occurrence and weather parameters at different time lags in Malaysia.	Artificial neural network	The study identified a significant correlation between leptospirosis occurrence and weekly average temperature at a lag of 16 weeks and weekly rainfall at a lag of 12-20 weeks.
Athithan, S. [25]	To construct a computational model for Leptospirosis' spread using cellular automata and a voting-based rule.	Cellular Automata	The simulation results closely match the actual data from Thailand, demonstrating the potential applicability of this voting-rule-based approach in accurately representing Leptospirosis transmission dynamics.
Ikram, R. [26]	Applying a mathematical model to the study of leptospirosis transmission	Compartmental model	The results indicate that He's homotopy perturbation method effectively approximates a solution for the leptospirosis transmission model.
Galego, M. A. [27]	To investigate the dynamics of leptospirosis transmission between rodents and humans	Compartmental model	The researchers found a critical threshold for the environmental carrying capacity that determines whether the disease persists or becomes extinct

4. The model can accommodate various time series patterns (e.g., trend, seasonality) by adjusting the model parameters.

Disadvantages of ARIMA model:

1. The ARIMA models assume that the underlying data follows a linear pattern. This assumption may not hold true in many real-world epidemic scenarios, which often involve nonlinear processes.

2. ARIMA models require the time series data to be stationary (i.e., the properties of the series do not change over time). Many epidemic processes might exhibit non-stationary behavior due to various factors such as policy changes, intervention measures, and population behavior changes.

3. Traditional ARIMA models do not incorporate external variables or factors that might influence the epidemic process, such as socioeconomic factors, environmental conditions, or intervention strategies.

ARIMA models are most effective for short-term forecasts. Their performance tends to decline for longer-term forecasts because they rely on the assumption that future patterns will resemble past patterns.

2.2. Model Performance Metrics

The model's accuracy was assessed using the Mean Absolute Percentage Error (MAPE) to confirm its predictive ability as follows:

$$\text{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|, \qquad (2)$$

where A_t is the actual value; F_t is the forecasted value.

The MAPE is commonly used to evaluate the accuracy of forecasting models, including those used for infectious disease forecasting. The MAPE method provides a measure of the relative error of the forecast, which is particularly useful when dealing with time series data that exhibit trends or seasonality. This means that the metric considers the magnitude of actual values, providing a more accurate representation of forecasting error.

3. Results

3.1. Data analysis

The data used for that research were collected from the open Annual epidemiological reports of European Centre for Disease Prevention and Control for 2010-2013 [29], 2014-2018 [30] and 2018-2022 [31]. The data on reported confirmed leptospirosis cases and their respective rates per 100000 population in Estonia, Lithuania, and Latvia from 2010 to 2022 were used for the experimental study. The results are presented in Table 2.

Figure 1 presents the number of leptospirosis cases and their rates per 100,000 population in Estonia, Lithuania, and Latvia from 2010 to 2022. The first chart displays the number of cases, and the second chart shows the incidence rates.

The stacked bar plot shows the number of leptospirosis cases in Estonia, Lithuania, and Latvia from 2010 to 2022 (Figure 2). Each bar represents the total number of cases per year, and different colors indicate the contribution from each country. This comparative visualization underscores the differences in leptospirosis incidence among the three countries during the given period.

In this study, we employed Pearson and Spearman correlation analyses to examine the relationships within our dataset. The Pearson correlation was used to assess the strength and direction of linear associations between variables, which is valuable for understanding straightforward proportional relationships in the data. In contrast, Spearman correlation was applied to explore monotonic relationships, which do not require linearity and are thus more robust to outliers and non-normal distributions.

Table 2

Reported confirmed leptospirosis cases: numbers and rate per 100 000 population

Year	Estonia cases	Estonia rate	Lithuania cases	Lithuania rate	Latvia cases	Latvia rate
2010	1	0.1	5	0.2	2	0.1
2011	2	0.2	3	0.1	6	0.3
2012	5	0.4	20	0.7	1	0.0
2013	2	0.2	10	0.3	1	0.0
2014	2	0.15	3	0.1	7	0.35
2015	2	0.15	10	0.34	2	0.1
2016	3	0.23	18	0.62	5	0.25
2017	5	0.38	16	0.56	8	0.41
2018	6	0.45	3	0.11	4	0.21
2019	5	0.38	0	0	4	0.21
2020	10	0.75	0	0	3	0.16
2021	8	0.6	0	0	1	0.05
2022	9	0.68	1	0.04	0	0

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

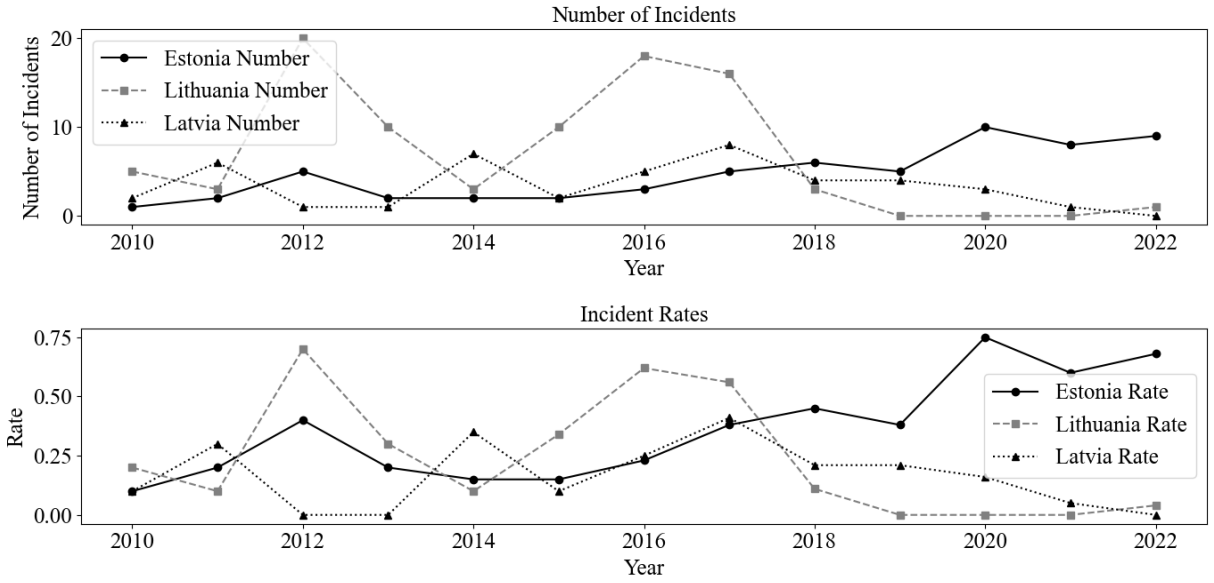

Figure 1. Trends in Reported Leptospirosis Cases and Incidence Rates in Estonia, Lithuania, and Latvia (2010-2022)

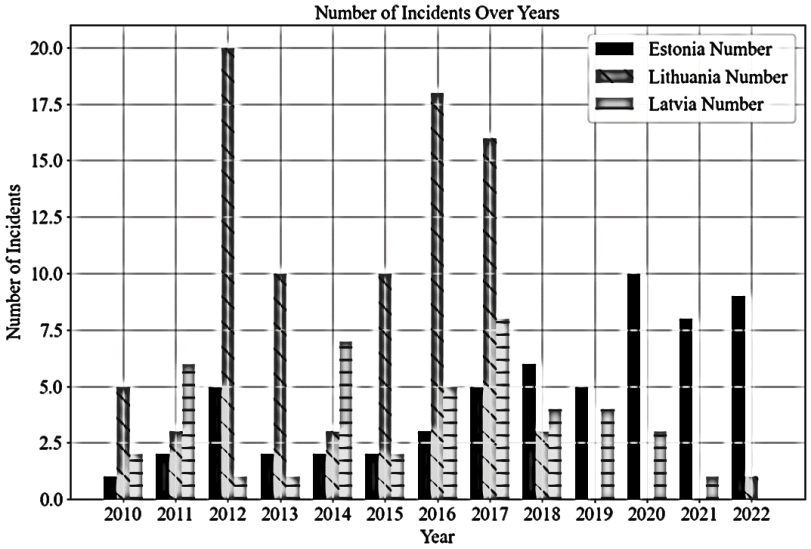

Figure 2. Stacked Bar Plot of Leptospirosis cases

This dual approach provided a comprehensive view, enabling us to confirm linear patterns while also detecting potential nonlinear associations, thereby enhancing the reliability of our exploratory data analysis.

Table 3 compares Pearson correlation coefficients.

Table 4 compares the Spearman correlation coefficients.

Figure 3 presents scatter plots for each pair of countries with the Pearson correlation coefficient displayed.

Table 3

Comparison of Pearson correlation coefficients

Countries	Pearson correlation coefficient	P-value	Interpretation
Estonia vs. Lithuania	-0.36	0.23	Moderate inverse relationship
Estonia vs. Latvia	-0.30	0.31	Weak inverse relationship. There is a slight trend towards the opposite movement of the number of incidents
Lithuania vs. Latvia	0.13	0.67	A very weak positive relationship. There is practically no linear relationship between the cases.

Table 4

Comparison of Spearman correlation coefficients

Countries	Spearman correlation coefficient	Interpretation
Estonia vs. Lithuania	-0.48	Average negative correlation. The number of incidents in one country tends to decrease as well as in the one with which it is compared.
Estonia vs. Latvia	-0.25	Weak negative correlation. There is a slight trend towards the opposite movement of the number of incidents.
Lithuania vs. Latvia	0.11	A very weak positive correlation. There is practically no linear relationship between the number of incidents.

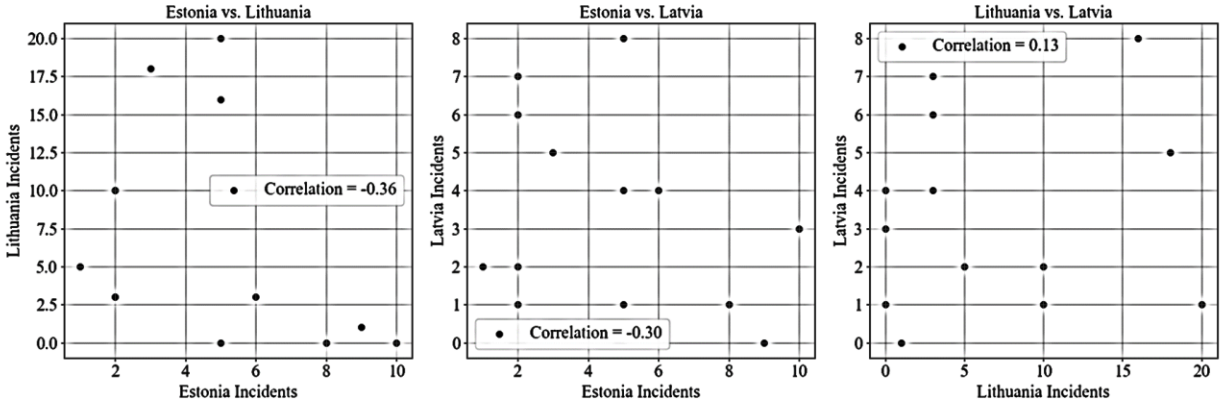

Figure 3. Comparative visualization of Pearson correlation coefficients

Pearson and Spearman correlations generally agree in identifying the type of relationship (positive or negative), but Spearman correlation sometimes reveals a stronger connection (for example, between Estonia and Lithuania).

In this analysis, both approaches show that there is the most significant negative correlation between incidents in Estonia and Lithuania, whereas there is almost no correlation between incidents in Lithuania and Latvia.

3.2. Experimental results

First, we apply data normalization by adjusting the values of each feature in the dataset to fall within a specified range, i.e., between 1 and 2. This process ensures that all features contribute equally to subsequent analyses or models, regardless of their scales. By transforming the data in this manner, we can improve the performance and interpretability of machine learning algorithms that are sensitive to feature scales.

Random splitting, which typically involves selecting a percentage of the data (e.g., 70% for training and 30% for validation) and randomly distributing the observations between the two sets. The data were split into training (the first 9 years) and validation (the last 3 years) sets.

The range is experimentally determined by defining the upper limits for the values of p, d, and q (in this example, the upper limit is set to 6).

Iterates through all possible combinations of p, d, and q values within the specified ranges. For each combination:

1. Attempt to construct an ARIMA model in the given order (p, d, q).
2. Forecast values for the validation period.
3. The MAPE was calculated to assess forecast quality.
4. If the current MAPE value is lower than the previous best, the best MAPE values, model order, and forecasts are updated.

Figure 4 presents the ARIMA model's retrospective prediction of leptospirosis morbidity in Estonia. This graph shows the model's capacity to capture observed patterns of leptospirosis incidence over time. The alignment of predicted values with actual data points suggests that the model performs reliably for Estonia, providing a clear basis for its applicability in forecasting future trends.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

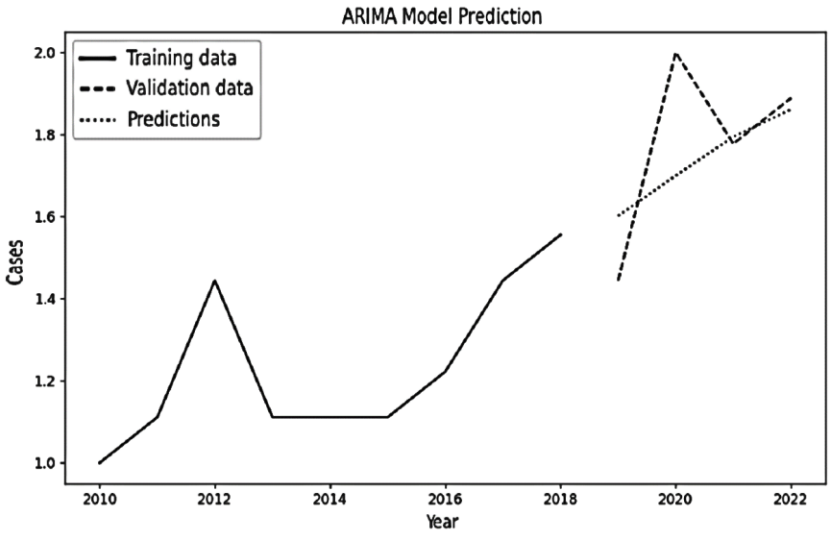

Figure 4. Prediction of leptospirosis in Estonia

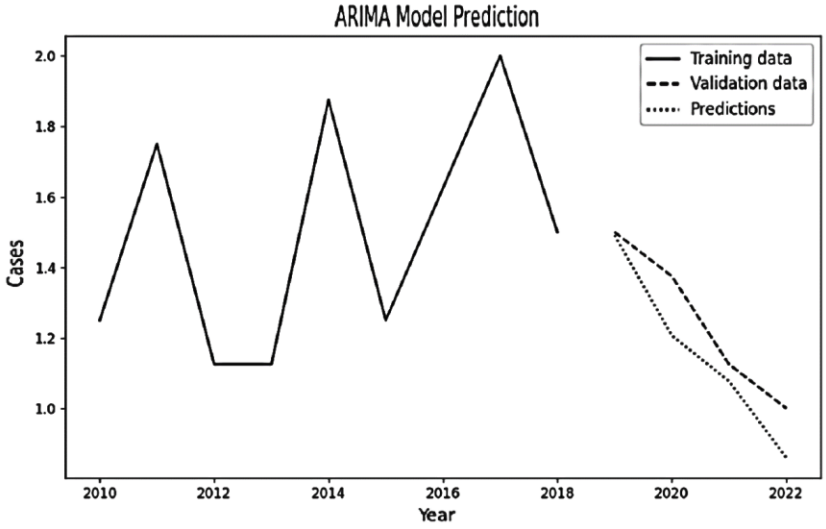

Figure 5. Prediction of leptospirosis in Lithuania

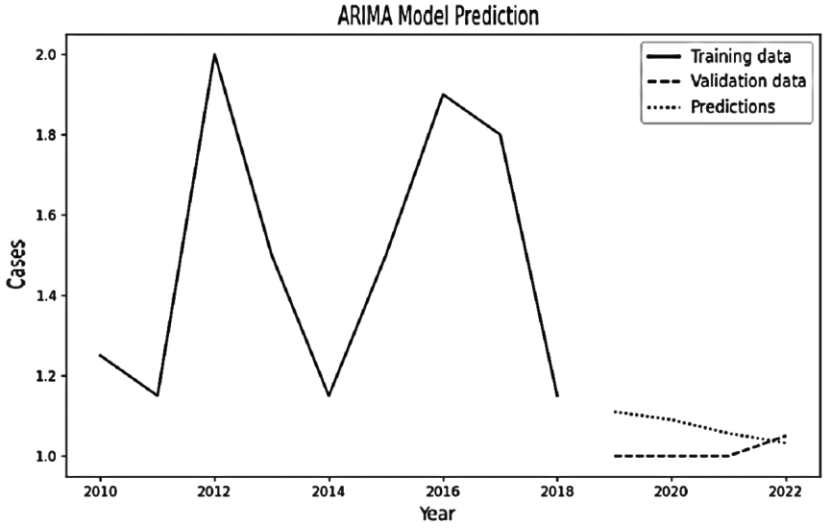

Figure 6. Prediction of leptospirosis in Latvia

Figure 5 presents a similar retrospective prediction for Lithuania, in which the ARIMA model follows the actual incidence pattern. Despite some deviations, particularly in high-incidence periods, the model adequately represents disease progression, indicating its robustness in forecasting trends within Lithuanian data.

Figure 6 shows the ARIMA prediction of leptospirosis morbidity in Latvia. The model's predictions align well with the observed values, capturing the primary trends and fluctuations in the data. This consistency underscores the model's potential utility in forecasting for Latvia, where incidences are less variable but still require reliable prediction methods.

Table 5 summarizes each country's ARIMA coefficients and Mean Absolute Percentage Error (MAPE). Estonia's model, with an order of (2, 2, 3), achieved a MAPE of 7.076, indicating an accurate fit. Lithuania's model (1, 2, 2) had the lowest MAPE of 6.841, suggesting the highest prediction accuracy among the countries. In Latvia, the (1, 2, 0) model resulted in a MAPE of 7.765, indicating slightly higher variability but still within an acceptable range. This comparative analysis showed that the ARIMA models effectively captured disease trends across the studied countries, with minor adjustments in the model order thereby optimizing the accuracy based on the regional data characteristics.

Table 5

ARIMA coefficients and predictions MAPE

Country	Best ARIMA order	MAPE
Estonia	(2, 2, 3)	7,076
Latvia	(1, 2, 0)	7,765
Lithuania	(1, 2, 2)	6,841

The ARIMA models applied to Estonia, Lithuania, and Latvia demonstrated a robust capacity to capture trends and forecast the incidence of leptospirosis in each country. The retrospective analyses indicate that although all models achieved reasonable alignment with historical data, variations in MAPE values across countries suggest that some regions exhibit slightly more challenging patterns for accurate prediction. With the lowest MAPE, Lithuania achieved the most reliable forecast, while Estonia and Latvia showed moderate accuracy, which is suitable for practical forecasting needs. These models underscore the ARIMA model's adaptability to different epidemiological profiles, highlighting its potential as a predictive tool for managing leptospirosis morbidity and aiding public health interventions across various regional contexts. The ARIMA model's performance across Estonia, Lithuania, and Latvia demonstrated significant prediction accuracy, confirming the model's effectiveness in accurately forecasting leptospirosis incidence, thereby fulfilling the research objective of improving prediction accuracy.

4. Discussion

The results of this study highlight the unique dynamics of leptospirosis morbidity across Estonia, Lithuania, and Latvia, providing insights into the factors influencing disease transmission in each country and the effectiveness of ARIMA models in predicting future cases. The initial morbidity data revealed that leptospirosis cases fluctuate annually and vary substantially between the countries. Estonia's higher incidence rate could be attributed to its rural areas and extensive water networks, which support habitats for rodents and other hosts of the Leptospira bacteria. This environmental setup, along with potentially high wildlife interaction and agricultural activities, may increase exposure risks for rural and outdoor residents. Latvia, which has a similarly large rural and agricultural sector, also experiences variable infection rates, suggesting that leptospirosis transmission may be influenced by seasonal or socio-economic factors, such as agricultural cycles, wildlife activity, or flood-related water contamination, which tend to vary annually. Lithuania's relatively low and sporadic leptospirosis rates could indicate less exposure to high-risk environments or more successful public health interventions. Differences in public awareness, healthcare access, and preventive measures could also account for the observed differences in morbidity rates.

The application of ARIMA models in this study aimed to leverage historical data to predict short-term morbidity trends. The varying MAPE values for each country's model provide insight into the model's performance and potential forecasting challenges in each setting. Lithuania's model achieved the lowest MAPE, indicating relatively stable historical patterns that may have contributed to higher predictive accuracy. This could be due to more consistent reporting practices or fewer year-to-year variations in disease incidence, which makes forecasting smoother and more reliable. Conversely, the slightly higher MAPE values for Estonia and Latvia may reflect the underlying variability in morbidity rates that complicate prediction. This variability might be tied to periodic environmental shifts, such as fluctuating water levels or annual variations in rodent populations, which the univariate ARIMA model is less equipped to account for. This suggests that although ARIMA models are beneficial for identifying overall trends, the accuracy of these forecasts may improve by including external environmental data, which could capture the seasonal and ecological drivers of leptospirosis transmission more effectively.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Correlation analysis between the three countries' morbidity rates revealed weak or inverse relationships, suggesting that leptospirosis cases in each country are likely influenced by local rather than regional factors. This lack of strong correlation implies that outbreaks in one country do not predict similar trends in neighboring areas, likely due to differences in local climate, land use, and public health strategies. The weak positive correlation between Latvia and Lithuania indicates minimal synchronization in incidence rates. This suggests that national factors, such as specific rainfall patterns and local agricultural practices, may drive infection rates. The inverse relationship observed between Estonia and Lithuania may further highlight contrasting environmental or social factors that impact transmission, such as different water management systems or preventive health policies that influence human exposure risks differently in each location.

The findings demonstrated that ARIMA models are a feasible tool for short-term forecasting of the incidence of leptospirosis in Baltic countries, although the effectiveness varies by setting. In areas where leptospirosis rates exhibit significant fluctuations, integrating additional variables like weather conditions, population movement, and socioeconomic factors could enhance predictive accuracy and provide a more comprehensive understanding of transmission dynamics. This approach would be particularly useful in countries with high case variability, where identifying the environmental and social determinants of outbreaks can aid in tailoring effective public health interventions.

This study underscores the importance of country-specific approaches to leptospirosis forecasting and control, as each region's unique epidemiological profile requires a customized model. The ARIMA models provided a foundational analysis; future work should explore multivariate forecasting approaches that incorporate climate data, rodent population trends, and human behavior patterns. Such models could offer more precise predictions and facilitate preemptive public health responses, especially in regions where the disease remains a persistent public health concern. These insights contribute to a better understanding of leptospirosis in varying geographic and environmental contexts, demonstrating the potential of predictive models to guide resource allocation and disease prevention strategies in Baltic countries and similar settings globally.

The study's limitation is the potential variability in data quality and reporting accuracy across the three countries. Differences in diagnostic capabilities, healthcare access, and reporting practices may lead to underreporting or inconsistent data, particularly in rural or resource-limited areas where leptospirosis cases are less likely to be diagnosed and documented. This variability could affect the robustness of the ARIMA model because fluctuations in reported cases may not always reflect actual changes in disease incidence, potentially impacting the reliability of the model's predictions.

To mitigate the impact of these data quality issues, future work could incorporate approaches like data imputation to handle missing or inconsistent data. Integrating external factors such as climatic conditions, agricultural practices, and socioeconomic variables into the model could help account for fluctuations in leptospirosis incidence. Multivariate models incorporating these additional predictors could enhance the robustness and generalizability of the findings by more effectively capturing the underlying drivers of leptospirosis.

Conclusions

In conclusion, this study comprehensively evaluates the ARIMA model's potential for forecasting leptospirosis morbidity across Estonia, Lithuania, and Latvia, contributing scientifically and practically to epidemiological modelling. Scientifically, this research advances our understanding of leptospirosis by demonstrating the utility of univariate time-series analysis in regions with varied ecological and socioeconomic contexts. The results underscore the adaptability of the ARIMA models in capturing distinct disease transmission patterns within the Baltic countries, where each country's unique geographic and environmental factors, such as rural landscapes and water networks, contribute to disease prevalence. This approach provides a better understanding of the regional dynamics that drive leptospirosis transmission and offers foundational insights that can broaden forecasting methods for zoonotic diseases.

From a practical perspective, the ARIMA-based forecasting model developed in this study presents a valuable tool for public health authorities in these countries. By enabling short-term prediction of leptospirosis incidence, this model can support health agencies in planning preventive measures, optimizing resource allocation, and preparing for potential outbreaks, thereby reducing morbidity and supporting proactive disease management. The model's ease of implementation, combined with the meaningful results obtained for each country, demonstrates its applicability in low-resource settings and its potential to inform timely and effective public health responses to leptospirosis and similar zoonotic diseases.

Future research should focus on enhancing predictive accuracy by integrating additional variables, such as climate data, socioeconomic factors, and animal population dynamics, which are known to influence leptospirosis transmission. The incorporation of multivariate approaches could improve the model's robustness and extend its predictive power by accounting for seasonal and environmental fluctuations affecting leptospirosis trends. In addition, exploring machine learning-based models

could provide new perspectives and increase forecasting precision, especially in areas with high variability in case data. This expanded approach will refine disease management strategies, enabling more precise targeting of high-risk periods and locations, ultimately improving public health outcomes in leptospirosis-endemic areas and beyond.

Contributions of authors: conceptualization – **Dmytro Chumachenko**; methodology – **Mykola Butkevych, Dmytro Chumachenko**; formulation of tasks – **Dmytro Chumachenko**; analysis – **Mykola Butkevych, Dmytro Chumachenko**; development of model – **Mykola Butkevych, Dmytro Chumachenko**; software – **Mykola Butkevych**; verification – **Mykola Butkevych, Dmytro Chumachenko**; analysis of results – **Mykola Butkevych, Dmytro Chumachenko**; visualization – **Mykola Butkevych**; writing – original draft preparation – **Mykola Butkevych, Dmytro Chumachenko**; writing – review and editing – **Mykola Butkevych, Dmytro Chumachenko**.

Conflict of interest

The authors declare that they have no conflict of interest concerning this research, whether financial, personal, authorship, or otherwise, that could affect the research and its results presented in this paper.

Funding

The study was funded by the National Research Foundation of Ukraine in the framework of the research project 2023.03/0197 on the topic *"Multidisciplinary study of the impact of emergency situations on the infectious diseases spreading to support management decision-making in the field of population biosafety"*.

Data availability

This study uses open datasets provided European Centre for Disease Prevention and Control available in [29 - 31].

Use of Artificial Intelligence

The authors confirm that they did not use generative artificial intelligence methods in their work.

All the authors have read and agreed to the publication of the final version of this manuscript.

References

1. Sykes, J. E., Haake, D. A., Gamage, C. D., Mills, W. Z., & Nally, J. E. A global one health perspective on leptospirosis in humans and animals. *Journal of the American Veterinary Medical Association*, 2022, vol. 260, iss. 13, pp. 1–8. DOI: 10.2460/javma.22.06.0258.

2. Costa, F., Hagan, J. E., Calcagno, J., Kane, M., Torgerson, P., Martinez-Silveira, M. S., Stein, C., Abela-Ridder, B., & Ko, A. I. Global Morbidity and Mortality of Leptospirosis: A Systematic Review. *PLoS Neglected Tropical Diseases*, 2015, vol. 9, iss. 9, article no. e0003898. DOI: 10.1371/journal.pntd.0003898.

3. Torgerson, P. R., Hagan, J. E., Costa, F., Calcagno, J., Kane, M., Martinez-Silveira, M. S., Goris, M. G. A., Stein, C., Ko, A. I., & Abela-Ridder, B. Global Burden of Leptospirosis: Estimated in Terms of Disability Adjusted Life Years. *PLOS Neglected Tropical Diseases*, 2015, vol. 9, iss. 10, article no. e0004122. DOI: 10.1371/journal.pntd.0004122.

4. Tazerji, S. S., Nardini, R., Safdar, M., Shehata, A. A., & Duarte, P. M. An Overview of Anthropogenic Actions as Drivers for Emerging and Re-Emerging Zoonotic Diseases. *Pathogens*, 2022, vol. 11, iss. 11, article no 1376. DOI: 10.3390/pathogens11111376.

5. Baharom, M., Ahmad, N., Hod, R., Fs, A., & Tangang, F. The Impact of Meteorological Factors on Communicable Disease Incidence and Its Projection: A Systematic Review. *International Journal of Environmental Research and Public Health*, 2021, vol. 18, iss. 21, pp. 11117–11117. DOI: 10.3390/ijerph182111117.

6. Sonne, C., Lakemeyer, J., Desforges, J.-P., Eulaers, I., Persson, S., Stokholm, I., Galatius, A., Gross, S., Gonnsen, K., Lehnert, K., Andersen-Ranberg, E., Olsen, M.T., Dietz, R., & Siebert, U. A review of pathogens in selected Baltic Sea indicator species. *Environment International*, 2020, vol. 137, pp. 105565–105565. DOI: 10.1016/j.envint.2020.105565.

7. Lehtla, A., Must, K., Lassen, B., Orro, T., Jokelainen, P., & Viltrop, A. Leptospira spp. in Cats in Estonia: Seroprevalence and Risk Factors for Seropositivity. *Vector-Borne and Zoonotic Diseases*, 2020, vol. 20, iss. 7, pp. 524–528. DOI: 10.1089/vbz.2019.2555.

8. *Latvia - Traveler view | Travelers' Health | CDC*. 2022, Available at: https://wwwnc.cdc.gov/travel/destinations/traveler/none/latvia (Accessed 1 Sep. 2024).

9. Jeske, K., Schulz, J., Tekemen, D., Balčiauskas, L., Balčiauskienė, L., Hiltbrunner, M., Drewes, S., Mayer-Scholl, A., Heckel, G., & Ulrich, R. G. Cocirculation of *Leptospira* spp. and multiple orthohantaviruses in rodents, Lithuania, Northern Europe. *Transboundary and emerging diseases*, 2022, vol. 69, iss. 5, pp. e3196–e3201. DOI: 10.1111/tbed.14470.

10. Bazilevych, K., Krivtsov, S., & Butkevych, M. Intelligent evaluation of the informative features of cardiac studies diagnostic data using Shannon method. *CEUR Workshop Proceedings*, 2021, vol. 3003, pp. 65-75.

11. Strilets, V., Donets, V., Ugryumov, M., Artiuch, S., Zelenskyi, R., & Goncharova, T. Agent-oriented data

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

clustering for medical monitoring. *Radioelectronic and computer systems*, 2022, vol. 2022, iss. 1, pp.103–114. DOI: 10.32620/reks.2022.1.08.

12. Mochurad, L., & Ranto, R. A parallel algorithm for the detection of eye disease. *Advances in Intelligent Systems*, 2023, vol. 158, pp. 111-125. DOI: 10.1007/978-3-031-24475-9_10.

13. Izonin, I., Tkachenko, R., Yemets, K., & Havryliuk, M. An interpretable ensemble structure with a non-iterative training algorithm to improve the predictive accuracy of healthcare data analysis. *Scientific Reports*, 2024, vol. 14, no. 1, article no. 12947. DOI: 10.1038/s41598-024-61776-y.

14. Babaiev, V. M., Kadykova, I. M., Husieva, Yu. Yu., & Chumachenko, I. V. The method of adaptation of a project-oriented organization's strategy to exogenous changes. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, 2017, iss. 2, pp. 134-140.

15. Sy, C., Bernardo, E., Miguel, A., San Juan, J. L., Mayol, A. P., Ching, P. M., Culaba, A., Ubando, A., & Mutuc, J. E. Policy Development for Pandemic Response Using System Dynamics: a Case Study on COVID-19. *Process Integration and Optimization for Sustainability*, 2020, vol. 4, iss. 4, article no. 497. DOI: 10.1007/s41660-020-00130-x.

16. Meniailov, I., & Padalko, H. Application of multidimensional scaling model for Hepatitis C data dimensionality reduction. *CEUR Workshop Proceedings*, 2022, vol. 3348, pp. 34-43.

17. Chumachenko, D., Bazilevych, K., Butkevych, M., Meniailov, I., Parfeniuk, Y., Sidenko, I., & Chumachenko, T. Methodology for assessing the impact of emergencies on the spread of infectious diseases. *Radioelectronic and Computer Science*, 2024, vol. 2024, iss. 3, pp. 6-26. DOI: 10.32620/reks.2024.3.01.

18. Conceição, K. S., Andrade, M. G., Louzada, F. Zero-modified Poisson model: Bayesian approach, influence diagnostics, and an application to a Brazilian leptospirosis notification data. *Biometrical Journal*, 2013, vol. 55, iss. 5, pp. 661–678. DOI: 10.1002/bimj.201100175.

19. Mukdasai, K., Sabir, Z., Raja, M. A. Z., Sadat, R., Ali, M. R., & Singkibud, P. A numerical simulation of the fractional order Leptospirosis model using the supervise neural network. *Alexandria Engineering Journal*, 2022, vol. 61, iss. 12, pp. 12431–12441. DOI: 10.1016/j.aej.2022.06.013.

20. Bhalraj, A., Azmi, A., & Mohd, H. Analytical and numerical solutions of Leptospirosis model. *International Journal of Mathematics and Computer Science*, 2021, vol. 16, iss. 3, pp. 949-961.

21. Minter, A., Costa, F., Khalil, A., Childs, J.E., Diggle, P. J., Ko, A. I., Begon, M. Optimal Control of Rat-Borne Leptospirosis in an Urban Environment. *Frontiers in Ecology and Evolution*, 2019, vol. 7, 209. DOI: 10.3389/fevo.2019.00209.

22. Qu, H., Saifullah, S., Khan, J., Khan, A., Rahman, M., & Zheng, G. Dynamics of leptospirosis disease in context of piecewise classical-global and classical-fractional operators. *Fractals*, 2022, vol. 30, iss. 08, article no. 2240216. DOI: 10.1142/s0218348x22402162.

23. Engida, H., Theuri, D. M., Gathungu, D. K., Gachohi, J., & Alemneh, H. T. A Mathematical Model Analysis for the Transmission Dynamics of Leptospirosis Disease in Human and Rodent Populations. *Computational and Mathematical Methods in Medicine*, 2022, vol. 2022, iss. 1, pp. 1–23. DOI: 10.1155/2022/1806585.

24. Rahmat, F., Zulkafli, Z., Juraiza Ishak, A., Mohd Noor, S. B., Yahaya, H., & Masrani, A. Exploratory Data Analysis and Artificial Neural Network for Prediction of Leptospirosis Occurrence in Seremban, Malaysia Based on Meteorological Data. *Frontiers in Earth Science*, 2020, vol. 8, article no. 377. DOI: 10.3389/feart.2020.00377.

25. Athithan, S., Shukla, V. P., & Biradar, S. R. Voting Rule Based Cellular Automata Epidemic Spread Model for Leptospirosis. *Indian Journal of Science and Technology*, 2015, vol. 8, iss. 4, article no. 337. DOI: 10.17485/ijst/2015/v8i4/60441.

26. Ikram, R., Khan, A., Khan, A., Khan, T., & Zaman, G. Analytical approximate solution of leptospirosis epidemic model with standard incidence rate. *Computational Methods for Differential Equations*, 2019, vol. 7, iss. 3, pp. 370–382.

27. Gallego, M. C., & Simoy, M. S. Mathematical modeling of leptospirosis: A dynamic regulated by environmental carrying capacity. *Chaos, Solitons and Fractals*, 2021, vol. 152, article no. 111425. DOI: 10.1016/j.chaos.2021.111425.

28. Helfenstein, U. Box-Jenkins modelling in medical research. *Statistical Methods in Medical Research*, 1996, vol. 5, iss. 1, pp. 3–22. DOI: 10.1177/096228029600500102.

29. *European Centre for Disease Prevention and Control. Annual Epidemiological Report 2016 – Leptospirosis*. Stockholm: ECDC; 2016. Available at: http://ecdc.europa.eu/en/healthtopics/leptospirosis/Pages/Annual-epidemiological-report-2016.aspx (Accessed 1 Sep. 2024).

30. *European Centre for Disease Prevention and Control. Leptospirosis. ECDC. Annual Epidemiological Report for 2018*. Stockholm: ECDC; 2022. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/Leptospirosis-AER-2018_Final.pdf (Accessed 1 Sep. 2024).

31. *European Centre for Disease Prevention and Control. Leptospirosis. ECDC. Annual Epidemiological Report for 2022*. Stockholm: ECDC; 2024. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/LEPT_AER_2022_Report.pdf (Accessed 1 Sep. 2024).

Received 19.09.2004, Accepted 18.11.2004

АНАЛІЗ ЧАСОВИХ РЯДІВ ЗАХВОРЮВАНОСТІ НА ЛЕПТОСПІРОЗ ДЛЯ ПРОГНОЗУВАННЯ В КРАЇНАХ БАЛТІЇ ЗА ДОПОМОГОЮ МОДЕЛІ ARIMA

М. В. Буткевич, Д. І. Чумаченко

Лептоспіроз, зоонозне захворювання з вагомими наслідками для громадського здоров'я, представляє значні труднощі для прогнозування через сезонні особливості та чутливість до екологічних факторів, особливо в недостатньо досліджених регіонах, таких як країни Балтії. **Метою** цього дослідження є розробка прогнозної моделі на основі ARIMA для передбачення захворюваності на лептоспіроз у Естонії, Латвії та Литві, де наявні дані про захворювання обмежені та варіативні. **Об'єктом** дослідження є епідемічний процес лептоспірозу, а **предметом** – застосування методології прогнозування часових рядів, придатних для епідеміологічних контекстів. **Методи.** Модель ARIMA була застосована до кожної країни для визначення часових закономірностей і генерації короткострокових прогнозів захворюваності, використовуючи дані про підтверджені випадки лептоспірозу з Європейського центру профілактики та контролю захворювань за 2010–2022 роки. **Результати.** Продуктивність моделі оцінювалася за середньою абсолютною відносною помилкою (MAPE), що виявила найбільш точний прогноз у Литві з MAPE 6.841. Естонія та Латвія показали помірні рівні точності, що, ймовірно, відображає варіативність випадків і відмінності в регіональних епідеміологічних особливостях. Ці результати демонструють, що моделі ARIMA можуть ефективно відображати загальні тенденції та забезпечувати короткострокові прогнози захворюваності навіть у різноманітних епідеміологічних умовах, що вказує на корисність ARIMA у середовищах з обмеженими ресурсами та змінними даними. **Висновки.** Наукова новизна цього дослідження полягає у застосуванні моделі ARIMA для прогнозування лептоспірозу в країнах Балтії, де комплексні дослідження часових рядів захворювання на лептоспіроз є рідкістю. Практична значущість полягає у розробці цінного інструменту для органів охорони здоров'я, для обґрунтування цільових втручанн, забезпечення ефективнішого розподілу ресурсів та планування своєчасного реагування на лептоспіроз та подібні зоонозні захворювання. Адаптивність моделі ARIMA та її проста реалізація в кожній країні показують її потенціал для інформування процесу прийняття рішень у сфері громадського здоров'я в умовах з обмеженими даними про закономірності захворювання.

Ключові слова: епідемічна модель; епідемічний процес; моделювання епідемії; моделювання; лептоспіроз, ARIMA.

Буткевич Микола Віталійович – здобувач каф. математичного моделювання та штучного інтелекту, Національний аерокосмічний університет ім. М. Є. Жуковського «Харківський авіаційний інститут», Харків, Україна.

Чумаченко Дмитро Ігорович – канд. техн. наук, доц., доц. каф. математичного моделювання та штучного інтелекту, Національний аерокосмічний університет ім. М. Є. Жуковського «Харківський авіаційний інститут», Харків, Україна.

Mykola Butkevych – PhD Student, the Department of Mathematical Modelling and Artificial Intelligence, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine,
e-mail: m.v.butkevych@khai.edu, ORCID: 0000-0001-8189-631X.

Dmytro Chumachenko – PhD, Associate Professor, Associate Professor at the Department of Mathematical Modelling and Artificial Intelligence, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine,
e-mail: dichumachenko@gmail.com, ORCID: 0000- 0003-2623-3294.