UDC 004.02, 004.67, 004.891.3, 616.24-002.5-02:316.342.6:316.62:314(477)
DOI: 10.20535/SRIT.2308-8893.2025.1.02

STUDYING THE RELATIONSHIP BETWEEN TUBERCULOSIS AND SOCIOECONOMIC, MEDICAL, AND DEMOGRAPHIC FACTORS IN UKRAINE

D.V. NEVINSKYI, D.I. MARTJANOV, I.O. SEMIANIV, Y.I. VYKLYUK

Abstract. Ukraine is currently experiencing a new, ongoing tuberculosis offensive. Our study analyzes the impact of various socioeconomic and medical factors, including the number of specialized hospitals, fluoroscopic examinations of the population, the number of healthcare workers, the level of alcohol and drug abuse, and others, on the prevalence of tuberculosis among different demographic groups in Ukraine. Artificial intelligence methods made it possible to identify key factors contributing to the growth or decline in tuberculosis incidence. The results of the SHAP (SHapley Additive exPlanations) analysis, which offers a methodology for interpreting complex machine learning models, shows the most important factors that influence the incidence of tuberculosis in Ukraine. The sensitivity analysis provided more important and detailed information, which confirmed the results of the SHAP analysis.

Keywords: artificial intelligence, tuberculosis, incidence, socio-demographic factors, medical factors, demographic factors.

RELEVANCE OF THE WORK

Currently, Ukraine is experiencing a new, regular offensive of tuberculosis. In the current conditions of development of Ukrainian society, one of the important problems that needs to be addressed is the spread of tuberculosis, a disease that is closely related to socioeconomic, medical and demographic factors [1]. The fact is that tuberculosis, as a social disease, is a mirror of socioeconomic well-being in the country [2].

The analysis of the ways of spreading, negative consequences for public health and other aspects of the spread of tuberculosis has long been the focus of research [3]. At the same time, the study of socioeconomic, medical and demographic reasons that influence the spread of tuberculosis in Ukrainian society remains an unexplored area of research.

Only a medical approach to the analysis of socio-economic, medical and demographic factors that affect the incidence of tuberculosis in Ukraine is insufficient in timely forecasting the prospects for the development of the tuberculosis epidemic and developing an appropriate plan to counter its challenges, as a result of which the incidence of tuberculosis remains extremely threatening not only to the life and health of our citizens, but also gives reason to consider this situation as a threat to the WHO European region [4].

Therefore, we used mathematical analysis with the use of artificial intelligence to establish the relationship between tuberculosis and socioeconomic, medical, and demographic factors in Ukraine.

ANALYSIS OF RESEARCH

Today, scientists are conducting research and modeling of the spread of tuberculosis [5]. Another study highlights how socioeconomic conditions contribute to the spread of tuberculosis [6]. In [7], the authors analyze how access to health care affects the effectiveness of tuberculosis control. They also consider how demographic changes affect the incidence of tuberculosis [8]. An overview of progress in the use of artificial intelligence (AI) in medicine [9]

The use of artificial intelligence in tuberculosis research is becoming increasingly popular due to its ability to analyze large data sets, identify complex relationships, and predict epidemiological trends. In particular, [10] uses various machine learning algorithms to predict the incidence of tuberculosis, which allows for high accuracy predictions and identification of regions at high risk of disease spread [11]. The authors have developed a deep learning-based system to automatically detect major chest diseases, including tuberculosis, in X-rays [12]. Although this study focuses on COVID-19, the methodologies and technologies they use can be adapted to monitor and predict the spread of tuberculosis, demonstrating the potential of AI in global epidemic management [13]. In this review, the authors discuss the possibilities of machine learning in the medical field, including its ability to integrate and analyze large amounts of data on socioeconomic factors to better understand their impact on the spread of tuberculosis.

However, there are currently no studies that examine the complex impact of various factors on the spread of tuberculosis based on artificial intelligence technology.

Therefore, **the purpose of our work** is to analyze the impact of various socioeconomic, medical, and demographic factors on the incidence of tuberculosis among the urban and rural population of Ukraine, in order to identify key factors that can contribute to the development of more effective strategies for controlling and preventing the disease.

MATERIALS AND METHODS

Description of the dataset. The dataset for analyzing the impact of various socioeconomic, medical, and demographic factors on tuberculosis incidence consists of the above fields and contains 400 records. The data was collected over the last 16 years and covers all regions of Ukraine. This dataset includes information on the number of specialized hospitals, the number of fluoroscopic examinations per 100.000 population, vaccination data, the number of bacterial isolators, the incidence among urban and rural residents, and the percentage of different demographic groups (workers, employees, healthcare workers, students, pupils, pensioners, unemployed, persons returned from prison, persons without permanent residence, private workers).

The dataset also includes indicators reflecting the level of alcohol abuse and drug use, the incidence of doctors in specialized hospitals per 10 thousand health care workers, HIV/TB rates per 100 thousand people, cases of resistant TB, treatment failure, interrupted treatment, patients dropped out of follow-up, treatment outcomes for relapses and multidrug-resistant tuberculosis (MDR-TB), and the number of surgical interventions (lung and extrapulmonary TB surgeries).

Research methodology. The research consists of the following steps:

1. **Correlation analysis.** At the first stage of the study, correlation analysis is used to identify statistical relationships between various factors (e.g., number of hospitals, healthcare workers, vaccination rates) and TB incidence. This allows us to determine which variables have a potential impact on the prevalence of the disease. The use of Pearson correlation coefficient helps to assess the strength and direction of the interaction between variables.

2. **Testing different models with cross-validation.** The next step is to test different machine learning models, such as linear regression, decision trees, random forest, kNN, support vector machine (SVM), adaptive boosting (AdaBoost), stochastic gradient descent, back propagation neural networks. Cross-validation is used to check the stability of models, in our case through 5-fold cross-validation, where the data is divided into 5 subsets and the model is tested 5 times, each time using one subset as a test set and the others as training data. The consistency of the cross-validation results served as an indicator of the presence of overfitting in these machine learning models and the selection of their hyperparameters. The following hyperparameters were selected: Linear Regression — Elastic Net regularization $L1/L2 = 50/50$, Decision Trees and Random Forest — maximum depth $d = 5$, Nearest Neighbors Method — selecting the optimal value of the nearest neighbors — $k = 5$, Support Vector Machine (SVM) — selecting the parameters $C = 0.8$ and $\gamma = 0.1$, Adaptive Boosting — Limiting the number of base models: n_estimators = 50, Stochastic Gradient Descent (SGD) — Elastic Net reegularozation $L1/L2 = 50/50$, Backpropagation — $\text{regularization} = 0.001$ and Dropout $d = 0.2$.

3. **Building an ensemble of models.** Based on the obtained models, an ensemble is built that combines the forecasts of the best models to improve the accuracy and reliability of the results. The study used a stacking-based ensemble, which allowed us to consider various aspects of the data and reduce the variability of the forecast.

4. **Analysis of an ensemble of models.** This analysis evaluates the overall performance of the model ensemble. It evaluates how the combination of models performs compared to individual models, including an assessment of accuracy, specificity, and other fit metrics.

5. **Determining the importance of factors.** Factor importance analysis is conducted to identify the key variables that have the greatest impact on morbidity. This may include the use of importance metrics provided by the algorithms that are included in the ensemble model.

6. **Sensitivity analysis.** The final step of sensitivity analysis tests the robustness of the model ensemble to changes in the data or in the model parameters. This involves varying key parameters and assessing the impact of these changes on the model results.

The study was conducted in the Orange environment. The data flow diagram is shown in Fig. 1.

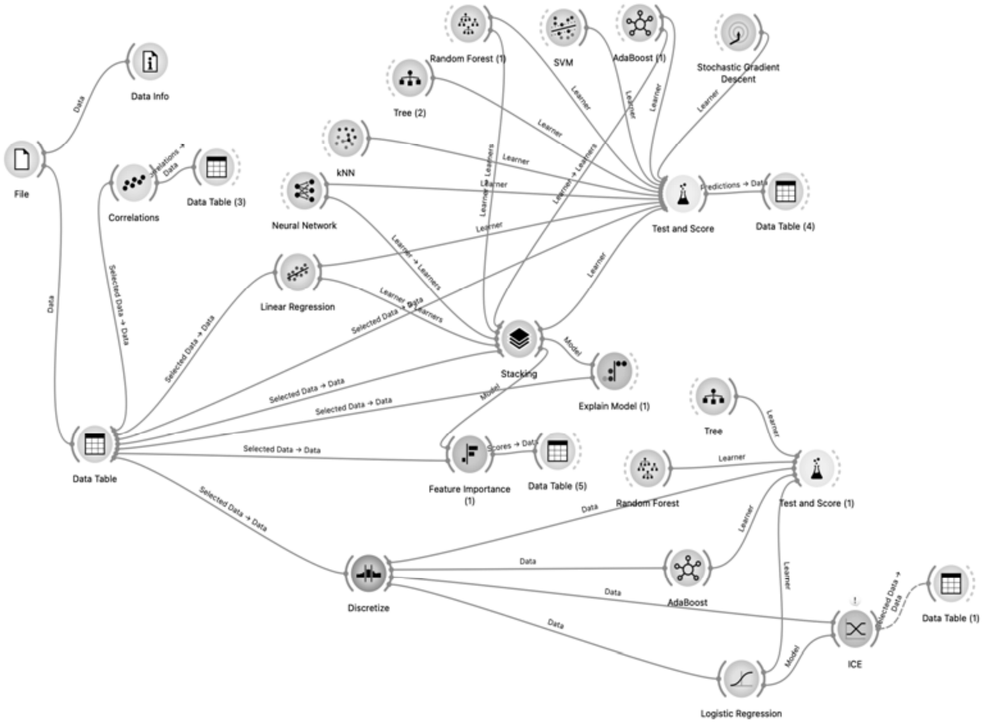

Fig. 1. The scheme of information flows of the study

RESULTS OF THE STUDY

Correlation analysis. Table 1 shows the results of the correlation analysis, which presents the values of the coefficients of determination R^2 for various factors that may affect the incidence of tuberculosis. The coefficient of determination R^2 measures the proportion of variation in a given variable that can be explained by the independent variables in the model. The key conclusions from the table include:

- **Bacterial excretion** has the highest coefficient $R^2 = 0.641$, indicating a strong relationship between the frequency of bacterial excretion in the population and the incidence of tuberculosis.
- **HIV/TB** (the ratio of HIV and TB incidence per 100.000 population) also has a significant coefficient of $R^2 = 0.542$, which emphasizes the link between these two diseases.
- **Fluoroscopic examinations** have a coefficient of $R^2 = 0.501$, which indicates the importance of regular medical examinations in detecting and controlling tuberculosis, especially in risk groups.
- **Physician morbidity** and **surgical treatment** also show relatively high R^2 values, which may reflect the impact of non-compliance with infection control conditions and the importance of surgery in some cases as an additional treatment method.
- The low R^2 coefficients for variables such as **alcohol** and **drug abuse and demographic groups** (e.g., **pensioners**, **students**, **workers**) indicate a less pronounced direct impact of these factors on morbidity compared to medical and epidemiological factors.

Table 1. Results of the correlation analysis

Factor	R^2
Bacterial excretion	0.641
HIV/TB (per 100 thousand)	0.542
Fluoroscopic examinations of the population (per 100 thousand)	0.501
Morbidity rate of doctors (per 10 thousand medical staff)	0.48
Surgical treatment (easy number of operations)	0.468
Resistant TB	0.466
Interrupted treatment	0.433
Unsuccessful treatment	0.387
Relapse rate (interrupted treatment)	0.379
Relapse rate (cured)	0.378
Expelled.	0.369
Non-operational (% of total)	0.364
MLS-TV (withdrawn)	0.335
Surgical treatment (total number of operations)	0.317
Relapse rate (unsuccessful treatment)	0.311
Recidivism rate (discharged)	0.308
Pensioners (% of total)	-0.294
Number of hospitals	0.216
Vaccinations carried out	0.2
R-treatment of MDR-TB (interrupted treatment)	0.146
Drug use (% of total)	0.118
Without a permanent place of residence (% of total)	-0.111
Alcohol abuse (% of total)	-0.107
Employees (% of total)	-0.091
MDR-TB treatment (failed treatment)	0.076
Private employees (% of total)	-0.056
Students (% of total)	0.052
Employees (% of total)	-0.047
People who returned from places of deprivation of liberty (% of the total)	-0.019
Students (% of total)	-0.01
Medical workers (% of total)	0.002

Testing different models by cross-validation. The next step was to analyze the performance of the above machine learning models in the context of tuberculosis incidence prediction using the 5-fold cross-validation method. The main parameters evaluated include the mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2). The results of the study are presented in Table 2.

Table 2. Results of testing different machine learning model

Machine Learning Model	MSE	RMSE	MAE	MAPE	R^2
Linear Regression	108.04	10.39	7.87	0.14	0.71
Neural Network	111.52	10.56	7.54	0.14	0.70
kNN	265.11	16.28	11.93	0.26	0.29
Decision Tree	191.64	13.84	9.42	0.18	0.49
Random Forest	80.92	9.00	6.64	0.13	0.78
SVM	255.39	15.98	11.69	0.25	0.32
AdaBoost	72.49	8.51	6.22	0.12	0.81
Stochastic Gradient Descent	132.32	11.50	8.52	0.16	0.65
Stacking	62.99	7.94	5.78	0.11	0.83

As can be seen from the table, the linear regression performed satisfactorily with a coefficient of determination of $R^2 = 0.71$, indicating that the model is moderately adequate for this data set. Although the RMSE and MSE are relatively high, this indicates potential deviations in predictions, especially when considering large and complex data. Neural networks are almost equal to linear regression in terms of R^2, but require more careful tuning and computational resources. This model can be particularly sensitive to overfitting due to the complexity of the model structure.

The KNN model showed the worst results with $R^2 = 0.29$, which indicates low prediction accuracy. The high MSE and RMSE values emphasize that the model does not work efficiently with the data, possibly due to insufficient data for training or mismatched model parameters. Decision trees showed average results $(R^2 = 0.49)$. This model is sensitive to changes in the data, and can create complex structures that lead to overfitting, especially in cases where tree pruning techniques are not applied.

Random Forest showed one of the best results $(R^2 = 0.78)$, demonstrating high accuracy and reliability of predictions. It efficiently manages overfitting and has a high classification and regression capability, thanks to the ensemble approach.

The support vector machine (SVM) method showed low efficiency $(R^2 = 0.32)$ with high MSE and RMSE, which may indicate the need to refine and optimize the kernel parameters to improve prediction.

Adaptive boosting (AdaBoost) showed the highest performance $(R^2 = 0.81)$ among all the models considered, with the lowest MSE and RMSE, indicating high accuracy and reliability. This model adapts well to different datasets, improving accuracy by consistently reducing the weight of errors in the training data

Stochastic Gradient Descent performed moderately well $(R^2 = 0.65)$, showing potential in situations where large datasets need to be optimized quickly. However, the method can be sensitive to noise in the data and requires careful tuning of the learning rate.

Building an ensemble of models. Based on the analysis of the performance of various machine learning models, it is proposed to create an ensemble of models using Stacking method, including the following estimators: linear regression, neural network, adaptive boosting (AdaBoost), and random forest. These models were chosen because of their high performance and complementarity in solving forecasting problems.

Stacking technology has the following advantages:

1. Complementarity of models: Random Forest and AdaBoost have demonstrated high accuracy in prediction, but they may tend to overlearn or bias in certain scenarios. Linear regression, while less accurate, offers stability and good generalization. Neural networks work effectively with non-linear relationships in data. Stacking allows you to combine their predictions, which can improve the overall accuracy and reliability of forecasting.

2. Reduce variability and errors: Stacking uses a linear model to stack predictions from the underlying models. This not only preserves the strengths of each model, but also effectively reduces the errors that can occur when using any single model.

3. Improved generalization: Using the predictions of different models as input to a "metamodel" in stacking allows the ensemble to generalize more effectively on unseen data, which is critical for real-world forecasting tasks.

Analysis of an ensemble of models. As can be seen in Table 2, the Stacking model shows the best performance among all the methods considered:

- R^2: The highest among all models, 0.83, indicating that the Stacking model explains approximately 83% of the variation in response across the dataset, outperforming its closest competitor (AdaBoost) by 0.02 points.
- MSE and RMSE: Stacking has the lowest MSE (6299) and RMSE (794), which indicates lower overall prediction errors compared to other models.
- MAE and MAPE: Also the lowest among all the models considered ($\text{MAE} = 578$ and MAPE = 0.011), which emphasizes the high accuracy of the forecasts created by the Stacking model.

Compared to individual models such as AdaBoost and Random Forest, which also showed high accuracy rates, Stacking provides an additional improvement in accuracy and stability. This demonstrates the power of a combined approach that considers different aspects of the data and the problem, while reducing the likelihood of overfitting that can occur with individual models.

Thus, stacking turned out to be the most efficient method among the analyzed ones, showing the highest performance across all evaluation criteria. This makes it an ideal candidate for use in real-world environments where high accuracy and reliability of forecasts are important.

Determining the importance of factors. The analysis of the importance of the factors, performed using a stacked model, allows us to identify the key variables that have the greatest impact on the incidence of tuberculosis. Assessment of the importance of each factor in the model allows us to better understand the dynamics of morbidity and optimize intervention strategies. Table 3 show the results of the importance of factors based on the stacked model.

As we can see from the data, the rate of bacterial shedding differs significantly from the others, which is fully supported by the literature [14]. It seems somewhat unexpected that the surgical treatment rate was among the factors with a significant impact. According to the current global TB treatment protocols, surgical treatment is indicated only in certain cases and is no longer used as often as it used to be. All other factors undoubtedly have an impact on the incidence of tuberculosis, which is confirmed by medical research data [1].

Table 3. Importance of factors in the stacking model

Feature	Importance
Bacterial excretion	0.405
Fluoroscopic examinations of the population (per 100 thousand)	0.059
Surgical treatment (easy number of operations)	0.026
MDR-TB treatment (failed treatment)	0.020
Expelled	0.016
Morbidity rate of doctors (per 10 thousand medical staff)	0.015
Resistant TB	0.015
MLS-TV (withdrawn)	0.014
HIV/TB (per 100 thousand)	0.011
People who returned from places of deprivation of liberty (% of the total)	0.009
Non-operational (% of total)	0.008
Alcohol abuse (% of total)	0.007
Pensioners (% of total)	0.007
R-treatment of MDR-TB (interrupted treatment)	0.006
Unsuccessful treatment	0.006
Vaccinations carried out	0.006
Number of hospitals	0.005
Relapse rate (unsuccessful treatment)	0.005
Without a permanent place of residence (% of total)	0.005
Relapse rate (cured)	0.004
Surgical treatment (total number of operations)	0.004
Students (% of total)	0.004
Recidivism rate (discharged)	0.004
Relapse rate (interrupted treatment)	0.004
Employees (% of total)	0.004
Medical workers (% of total)	0.003
Private employees (% of total)	0.003
Interrupted treatment	0.003
Drug use (% of total)	0.003
Employees (% of total)	0.002
Students (% of total)	0.002

As one can see from the results:

- **Bacterial shedding** is the most important factor (0.405), indicating a high level of influence on TB incidence. This emphasizes the need to focus on controlling the spread of bacterial shedding, as this indicator correlates with high incidence rates.
- **Fluoroscopic examinations** have the second most important indicator (0.059). This confirms the importance of regular medical examinations, especially

for risk groups, in detecting and preventing the disease, which allows for early identification of new cases of tuberculosis.

- **Surgical treatment** and **outcomes** for **MDR-TB** are also important variables. This reflects the importance of additional surgical interventions, in addition to chemotherapy, and the importance of successful treatment in the context of fighting resistant forms of TB and the need to improve and optimize treatment strategies.
- **The incidence of** physician-associated and **resistant TB** is also relatively high, which may indicate the risk of non-compliance with infection control measures in healthcare facilities and challenges associated with the spread of resistant forms of TB.
- Less important, but still significant, variables include **HIV/TB comorbidity**, **reentry from prison**, and **socioeconomic indicators** such as **alcohol abuse**. These variables indicate the complexity of the links between social conditions and disease, which requires a comprehensive approach to community health.

Sensitivity analysis. Sensitivity analysis and SHAP analysis are important tools for analyzing the spread of tuberculosis, which help to better understand the mechanisms of the model and its response to changes in input data.

SHAP (SHapley Additive exPlanations) analysis offers a methodology for interpreting complex machine learning models. It allows one to identify the contribution of each factor to the model's prediction, which is crucial for transparency and clarity in medical and policy decision-making. In the context of tuberculosis, SHAP analysis helps to identify which factors are most important for disease incidence, which can help to develop targeted interventions.

Sensitivity analysis is used to assess the stability and reliability of predictive models by determining how they respond to changes in input parameters. In the context of this study, this analysis allows us to test how small changes in factors, such as the number of medical examinations or demographic composition, can affect the model's conclusions. This is critical to ensure the accuracy and reproducibility of the results, especially in settings where models may be used to support public health decisions.

Fig. 2 shows the SHAP analysis of the stacking model. The graph shows the most important factors of the model. Each point on the graph corresponds to a SHAP value for each factor. The SHAP value is a measure of how much each

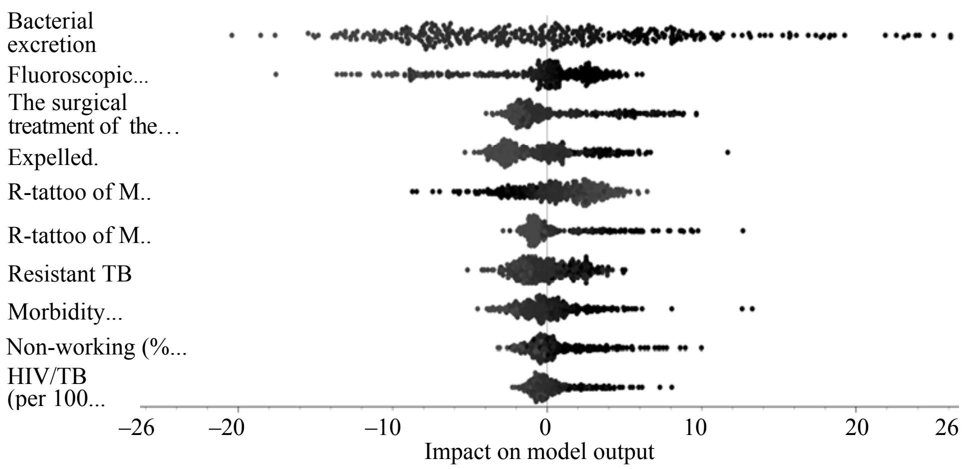

Fig. 2. SHAP analysis of the stacking model

factor influences the model outcome. A higher SHAP value (greater deviation from the center of the graph) means that the factor value has a greater impact on the prediction for the selected class. Positive SHAP values (points to the right of the center) are the values of features that influence prediction. The SHAP value shows how much the feature value affects the predicted value from the average prediction. The colors represent the value of each factor. Red represents a higher texture value and blue represents a lower value. The color range is determined based on all the values in the dataset for the object. As you can see from Fig. 2, the results of the SHAP analysis fully confirm the importance of the factors.

Sensitivity analysis provides more important information. Figs. 3–5 show the dependence of changes in tuberculosis incidence on changes in the most important factors.

All graphs are individual sensitivity plots for each individual row in the dataset. The yellow graph shows the average value of all records.

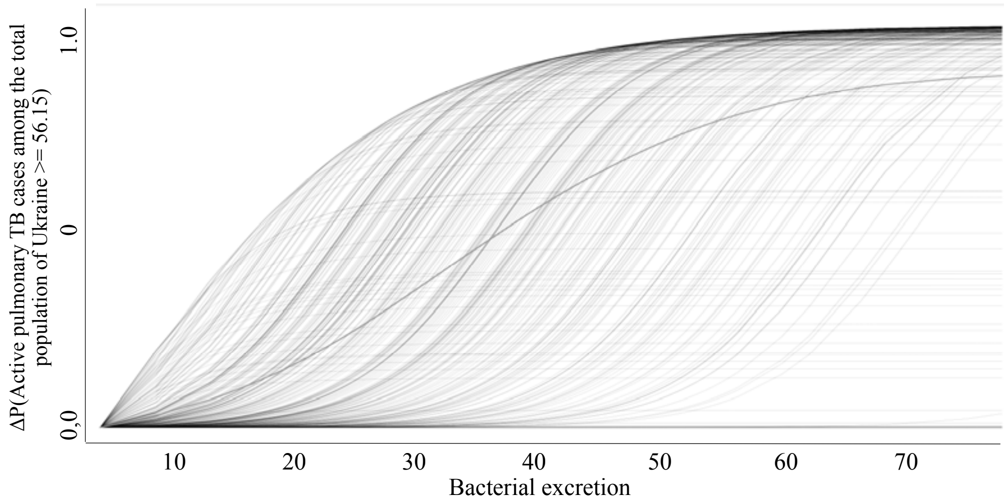

Fig. 3. The dependence of changes in tuberculosis incidence on bacterial shedding per 100 thousand people

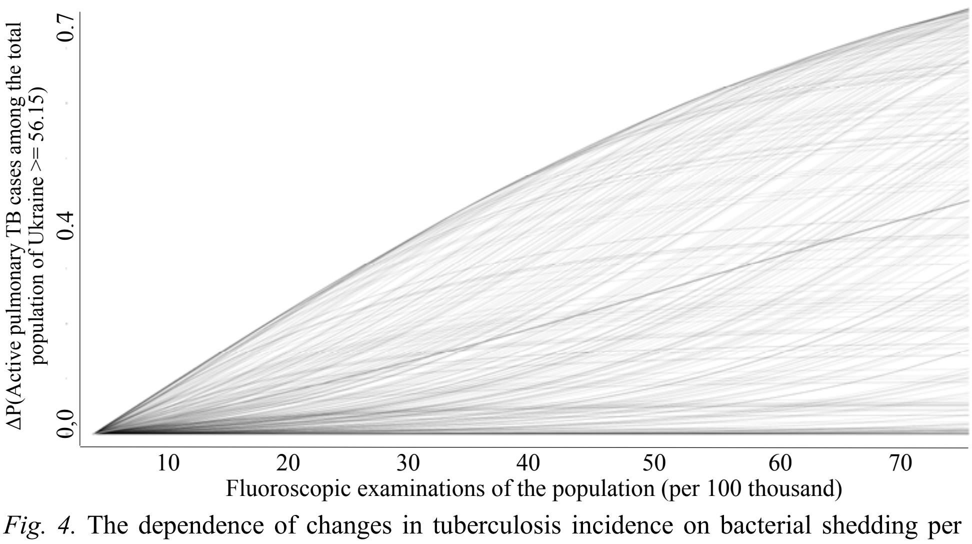

Fig. 4. The dependence of changes in tuberculosis incidence on bacterial shedding per 100 thousand people

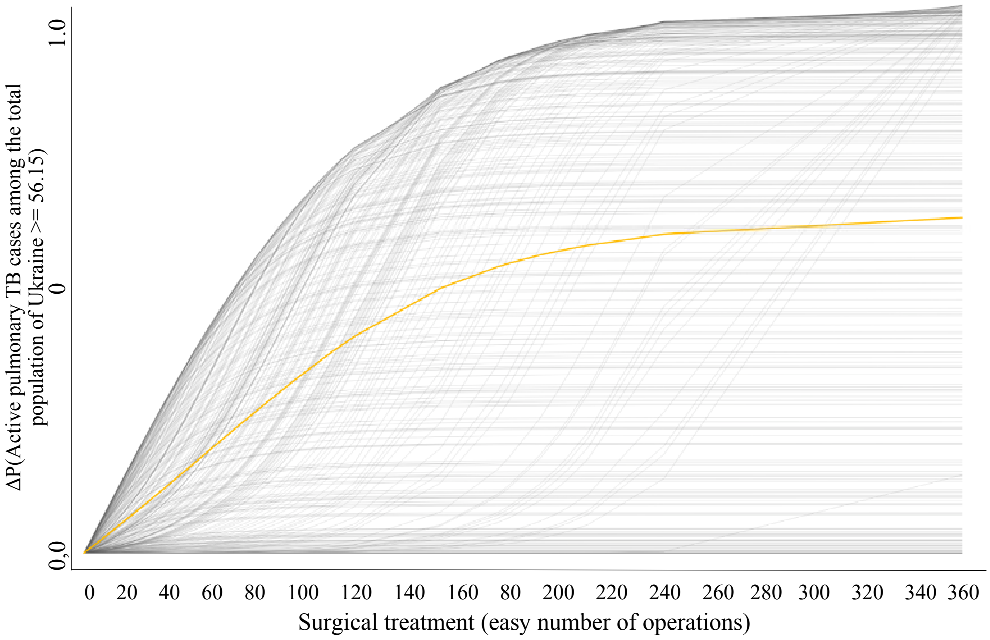

Fig. 5. Analysis of the sensitivity of tuberculosis incidence to the rate of surgical treatment of pulmonary tuberculosis (number of surgeries)

The logarithmic growth of the incidence shows a rapid increase against the background of an increase in bacterial shedding, but then a stable saturation level is determined. From a medical point of view, this is explained by the fact that active bacterial shedders quickly infect their contacts, and then the process of infection spread is suspended until new active patients start infecting others. This points to the importance of the efforts of health care systems in developed countries, which are primarily aimed at identifying and starting treatment of patients with bacterial excretion as soon as possible. Such patients pose a danger to others, often without realizing it. One undetected patient can infect 10 to 15 people who are in close daily contact with him or her. Thus, the result fully confirms the WHO epidemiological studies.

The linear increase in morbidity against the background of the fluoroscopic examination rate demonstrates a gradual, steady increase in the number of active TB patients. The importance of fluoroscopic examinations is confirmed by the latest WHO recommendations, especially the statement that fluoroscopic examinations of the population should focus on high-quality screening of risk groups rather than on random screening of everyone. Since Ukraine still has a quite high incidence of tuberculosis, and the number of internally displaced persons reached 4.9 million during the war period, all these people can be considered a risk group. The importance of regular fluoroscopic preventive examinations has been confirmed by numerous studies [15], and the fact that the sensitivity analysis ranked this indicator second in terms of its impact on morbidity is logical and understandable for the medical community.

An analysis of the sensitivity of the active TB incidence rate to the pulmonary tuberculosis surgical treatment rate shows a logarithmic increase at the beginning and a rapid transition to a stable level. This is due to the achievement of drug-free treatment of tuberculosis over a certain period. The number of surgical treatments for pulmonary tuberculosis is decreasing every year, but there are no large studies on the correlation of this indicator with the incidence of pulmonary tuberculosis [16]. The sensitivity analysis demonstrated exactly these results the surgical interventions rate, as we performed statistical processing of the data from 2007, and for most of this sixteen-year period, surgical treatment was performed along with chemotherapy for tuberculosis.

CONCLUSIONS

The use of artificial intelligence to analyze socioeconomic, medical, and demographic data has helped to identify the main factors contributing to the incidence of tuberculosis in Ukraine. In particular, the analysis confirmed the significant impact of the number of specialized hospitals, fluoroscopic examinations of the population, and the frequency of bacterial excretion on the incidence rate.

The development and validation of machine learning models, including linear regression, random forests, and adaptive boosting, allowed for accurate forecasting of tuberculosis incidence. The use of 5-fold cross-validation increased the reliability of the predictions, ensuring stability and accuracy across different demographic groups.

The results of the SHAP analysis, which offers a methodology for interpreting complex machine learning models, show the most important factors that influence the incidence of tuberculosis in Ukraine, with the greatest impact shown in bacterial excretion rates and fluoroscopic examinations of the population.

Interpretation of complex models through SHAP analysis and sensitivity analysis provided a deep understanding of the impact of individual factors, allowing for the formulation of targeted strategies for TB control and prevention. This creates the basis for informed decision-making in the field of public health and optimization of health care resources.

REFERENCES

1. S.S. Chiang et al., “Clinical manifestations and epidemiology of adolescent tuberculosis in Ukraine,” *ERJ Open Res*, 6(3):00308-2020, 2020. doi: https://doi.org/10.1183/23120541.00308-2020
2. I. Margineanu et al., “TB therapeutic drug monitoring - analysis of opportunities in Romania and Ukraine,” *Int. J. Tuberc. Lung Dis.*, 27(11), pp. 816–821, 2023. doi: 10.5588/ijtld.22.0667
3. O.S. Shevchenko, L.D. Todoriko, I.A. Ovcharenko, O.O. Pogorelova, and I.O. Semianiv, “A mathematical model for predicting the outcome of treatment of multidrug-resistant tuberculosis,” *Wiad. Lek.*, 74(7), pp. 1649–1654, 2021. doi: 10.36740 WLek202107117
4. D. Butov et al., “National survey on the impact of the war in Ukraine on TB diagnostics and treatment services in 2022,” *Int. J. Tuberc. Lung. Dis.*, 27(1), pp. 86–88, 2023. doi: 10.5588/ijtld.22.0563
5. K. Lönnroth, E. Jaramillo, B.G. Williams, C. Dye, and M. Raviglione, “Drivers of tuberculosis epidemics: the role of risk factors and social determinants,” *Soc. Sci. Med.*, 68(12), pp. 2240–2246, 2009. doi: 10.1016/j.socscimed.2009.03.041
6. Rifat Atun, Diana E.C. Weil, Mao Tan Eang, and David Mwakyusa, “Health-system strengthening and tuberculosis control,” *The Lancet*, 375(9732), pp. 2169–2178, 2010. doi: 10.1016/S0140-6736(10)60493-X
7. M.A. Mujtaba et al., “Demographic and Clinical Determinants of Tuberculosis and TB Recurrence: A Double-Edged Retrospective Study from Pakistan,” *J. Trop. Med.*, vol. 2022, article ID 4408306, 2022. doi: 10.1155/2022/4408306
8. E.J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” *Nat. Med.*, vol. 25, pp. 44–56, 2019. doi: https://doi.org/10.1038/s41591-018-0300-7

9. P. Farmer, "The major infectious diseases in the world--to treat or not to treat?" *N. Engl. J. Med.*, 345(3), pp. 208–210, 2001. doi: 10.1056/NEJM200107193450310
10. N. Tang et al., "Machine Learning Prediction Model of Tuberculosis Incidence Based on Meteorological Factors and Air Pollutants," *Int. J. Environ. Res. Public Health*, 20(5), 3910, 2023. doi: 10.3390/ijerph20053910
11. E.J. Hwang et al., "Development and Validation of a Deep Learning-Based Automated Detection Algorithm for Major Thoracic Diseases on Chest Radiographs [published correction appears in JAMA Netw Open. 2019 Apr 5;2(4):e193260]," *JAMA Netw Open*, 2(3):e191095, 2019. doi: 10.1001/jamanetworkopen.2019.1095
12. S. Tuli, S. Tuli, R. Tuli, and S.S. Gill, "Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing," *Internet of Things*, 11:100222, 2020. doi: 10.1016/j.iot.2020.100222
13. A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *N. Engl. J. Med.*, 380(14), pp. 1347–1358, 2019. doi: 10.1056/NEJMra1814259
14. K.E. Wiens et al., "Global variation in bacterial strains that cause tuberculosis disease: a systematic review and meta-analysis," *BMC Med.*, 16(1), article no. 196, 2018. doi: 10.1186/s12916-018-1180-x
15. V. Smelov et al., "Rationale and Purpose: The FLUTE Study to Evaluate Fluorography Mass Screening for Tuberculosis and Other Diseases, as Conducted in Eastern Europe and Central Asia Countries," *Int. J. Environ. Res. Public. Health*, 19(14), 8706, 2022. doi: 10.3390/ijerph19148706
16. R. Zaleskis, A.W. Mariani, F. Inzirillo, and I. Vasilyeva, "The Role of Surgery in Tuberculosis Management: Indications and Contraindications," in *G.B. Migliori, M.C. Raviglione (eds) Essential Tuberculosis*. Springer, Cham, 2021. doi: https://doi.org/10.1007/978-3-030-66703-0_15

Received 10.05.2024

INFORMATION ON THE ARTICLE

Denys V. Nevinskyi, ORCID: 0000-0002-0962-072X, Lviv Polytechnic National University, Ukraine, e-mail: nevinskiy90@gmail.com

Dmytro I. Martjanov, ORCID: 0009-0003-3919-4412, Lviv Polytechnic National University, Ukraine, e-mail: d.martjnoff@gmail.com

Ihor O. Semianiv, ORCID: 0000-0003-0340-0766, Bukovinian State Medical University, Ukraine, e-mail: igor_semianiv@bsmu.edu.ua

Yaroslav I. Vyklyuk, ORCID: 0000-0003-4766-4659, Lviv Polytechnic National University, Ukraine, e-mail: vyklyuk@ukr.net

ВИВЧЕННЯ ЗВ'ЯЗКУ МІЖ ТУБЕРКУЛЬОЗОМ ТА СОЦІАЛЬНО-ЕКОНОМІЧНИМИ, МЕДИЧНИМИ, ДЕМОГРАФІЧНИМИ ЧИННИКАМИ В УКРАЇНІ / Д.В. Невінський, Д.І. Мартьянов, І.О. Сем'янів, Я.І. Виклюк

Анотація. Натепер Україна переживає новий, черговий наступ туберкульозу. Це дослідження аналізує вплив різних соціально-економічних та медичних факторів, включаючи: кількість спеціалізованих лікарень, флюорографічні огляди населення, кількість медичних працівників, рівень зловживання алкоголем та наркотиками та інші на поширеність туберкульозу серед різних демографічних груп населення в Україні. Використання методів штучного інтелекту дало змогу визначити ключові чинники, що сприяють зростанню або зниженню захворюваності на туберкульоз. Результати SHAP (SHapley Additive exPlanations) аналізу, який пропонує методологію для інтерпретації складних моделей машинного навчання, показує найважливіші фактори, які впливають на захворюваність туберкульозом в Україні. Більш важливу інформацію несе аналіз чутливості, який підтвердив отримані показники в SHAP аналізі.

Ключові слова: штучний інтелект, туберкульоз, захворюваність, соціально-демографічні чинники, медичні чинники, демографічні чинники.