

# Using Machine Learning Methods to Analyze HIV Incidence in Ukraine

Yuri Parfeniuk<sup>1,†</sup>, Dmytro Kurinniy<sup>2,†</sup>, Ksenia Bazilevych<sup>2,†</sup> and Ievgen Menialov<sup>1,†</sup>

<sup>1</sup> V.N. Karazin Kharkiv National University, Kharkiv, Ukraine

<sup>2</sup> National Aerospace University “Kharkiv Aviation Institute”, Kharkiv, Ukraine

## Abstract

HIV remains a persistent public health issue in Ukraine, with complex socio-economic and geopolitical factors influencing its incidence. This study investigates the application of machine learning techniques to analyze and predict HIV incidence trends across various regions of Ukraine. Utilizing publicly available epidemiological and demographic data, we apply different methods—including decision trees, random forests, logistic regression, and clustering algorithms—to identify key risk factors and uncover spatial and temporal patterns in HIV transmission. The results demonstrate that machine learning models can improve the accuracy of HIV incidence predictions and support data-driven decision-making for public health interventions. The study highlights the potential of machine learning tools to enhance disease surveillance and inform targeted prevention strategies in Ukraine's evolving healthcare landscape.

## Keywords

machine learning, epidemiological data, epidemic surveillance, infectious diseases

## 1. Introduction

HIV is a chronic infectious disease caused by the human immunodeficiency virus. It affects the immune system, gradually reducing its ability to resist infections and diseases. HIV is transmitted through blood, sexual contact, and from mother to child during pregnancy, childbirth, or breastfeeding. It is one of the most serious viral diseases known to humanity since the late 20th century.

Globally, HIV is one of the leading causes of death from infectious diseases [1], ranking alongside tuberculosis. According to the World Health Organization (WHO), around 650,000 people die each year from HIV-related illnesses [2]. This makes HIV/AIDS one of the most significant public health challenges. The problem is further complicated by the fact that HIV is not only a severe infection but also an issue of access to treatment. According to WHO, about 38 million people worldwide live with HIV, but only slightly more than half of them receive antiretroviral therapy (ART), which is essential for controlling the virus.

Lack of access to treatment leads to high mortality rates – without ART, the life expectancy after an AIDS diagnosis ranges from several months to three years [3]. The situation in Ukraine remains challenging. Although a national HIV/AIDS control program has been established, the prevalence of the disease remains high. In recent years, there has been a slight decrease in the number of new infections; however, the issue is still relevant. One of the main obstacles in the fight against HIV is late diagnosis, insufficient population coverage with testing, and a low level of awareness regarding preventive measures.

In 2010, the HIV prevalence rate in Ukraine was 0.9% among the adult population, which was one of the highest rates in Eastern Europe. However, the HIV/AIDS-related mortality rate remained

<sup>1</sup>ProFIT AI'25: 5th International Workshop of IT-professionals on Artificial Intelligence, October 15–17, 2025, Liverpool, UK

\*Corresponding author.

<sup>†</sup>These authors contributed equally.

 parfuriy.l@gmail.com(Y. Parfeniuk); dkurinnoj560@gmail.com (D. Kurinniy); ksenia.bazilevich@gmail.com (K. Bazilevych); evgenii.menialov@gmail.com(I. Menialov);

 0000-0001-5357-1868 (Y. Parfeniuk); 0009-0007-4186-6096 (D. Kurinniy); 0000-0001-5332-9545 (K. Bazilevych); 0000-0002-9440-8378 (I. Menialov)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

high, reaching 10.2 cases per 100,000 population [4]. The most vulnerable groups continue to be young people, individuals who use injectable drugs and sex workers, among whom the infection rate can exceed 20%.

Despite certain improvements, Ukraine continues to face serious challenges in combating HIV. The main issues include insufficient public awareness about the modes of virus transmission, late diagnosis, and inadequate coverage with antiretroviral therapy. In order to reduce the prevalence of HIV, the government has developed a national strategy to combat the epidemic, which includes improving access to testing and treatment. In addition, Ukraine has submitted an application to the Global Fund to Fight AIDS, Tuberculosis and Malaria for approximately 100 million US dollars, which will be used for HIV prevention and treatment [5]. Thus, the fight against HIV, especially among vulnerable population groups, is a key public health priority at both the national and international levels.

A modern approach to data analysis and forecasting the spread of infection can play a significant role in the fight against HIV. Machine learning (ML) has become a critically important tool in the field of healthcare [6], enabling the analysis of large volumes of data and the identification of hidden patterns that are essential for predicting the epidemiological situation. For example, methods such as decision trees, random forest, support vector machines (SVM), and neural networks demonstrate high effectiveness in predicting the spread of HIV by analyzing complex factors, including behavioral aspects, demographic data, and the level of accessibility of medical services [7].

Machine learning methods make it possible not only to forecast the spread of HIV infection but also to identify the most vulnerable regions and social groups, which allows for more effective allocation of medical resources and the implementation of targeted preventive measures [8].

The main objective of the study is to identify and implement methods that enable the analysis of HIV incidence and the identification of areas of HIV spread in Ukraine using machine learning techniques.

Object of the study: the process of HIV infection spread in Ukraine. Subject of the study: the use of machine learning methods for studying HIV incidence in Ukraine.

To achieve this objective, it was necessary to address the following tasks:

- analyze the epidemiological situation regarding HIV infection in Ukraine;
- conduct an analytical review of machine learning methods that can be used to study incidence and identify areas of HIV spread;
- develop algorithmic models to solve the research tasks;
- develop software to implement the research tasks;
- evaluate the results obtained.

The current research is part of a comprehensive information system for assessing the impact of emergencies on the spread of infectious diseases described in [9]

## **2. Development of a software application for analyzing HIV incidence in Ukraine using machine learning methods**

The study utilized five algorithmic models that formed the basis of the software: an algorithmic model for smoothing time series based on the moving average method («Moving average»); an algorithmic model based on the K-Nearest Neighbors method («K-Nearest Neighbors»); algorithmic model based on the method of «Random Forest»; algorithmic model based on ensemble methods («Ensemble»); algorithmic clustering model based on the method K-means. To evaluate the effectiveness of the models, key accuracy metrics were calculated: MAE (Mean Absolute Error); RMSE (Root Mean Squared Error);  $R^2$  (Coefficient of Determination). The study utilized an open dataset from the Public Health Center [10]. The user specifies the year for analysis, after which the following processes are performed: data preprocessing, including missing data removal and

normalization of indicators (for example, the number of new cases, testing coverage percentage, ART therapy); clustering of Ukrainian regions using the K-means algorithm, which allows grouping of regions based on similar epidemiological characteristics. To assess the quality of clustering, the purity metric was used — it determines how well a cluster corresponds to known groups (for example, by geography or infection zones) and ranges from 0 to 1. Determination of the optimal number of clusters is performed using the "elbow" method, which ensures a balance between accuracy and generalization of the models. HIV spread forecasting is carried out using the Random Forest model, which provides high accuracy due to the use of many independent decision trees. The obtained clustering results are presented as a color-coded map of Ukraine, where each region is highlighted according to its cluster. This allows for a visual assessment of:

- territorial concentration of morbidity;
- regions similar in prevalence levels;
- zones of epidemic risk.

In addition, the graphs display actual and predicted values of new HIV cases by years and regions, providing convenient data visualization and facilitating decision-making. The aim of this stage was to study the current state of the epidemic situation, the dynamics of HIV infection spread by regions, and the identification of areas with increased risks. The analysis included:

- collection and systematization of statistical data for the years 2019–2024;
- visualization of HIV prevalence on the country map;
- identification of regional differences in morbidity levels.

Forecasting morbidity included:

- preprocessing and normalization of time series data (including the moving average method using);
- splitting data into training and testing datasets;
- application of machine learning algorithms to build forecasting models (linear regression, Random Forest, k-means, gradient boosting);
- evaluation of the accuracy of the constructed models using relevant metrics.

The forecasting focused on the following specific indicators:

- the expected level of HIV incidence in future periods (based on data from previous years);
- identification of regions with potential increases or decreases in incidence;
- detection of patterns and trends that can be used for developing preventive measures.

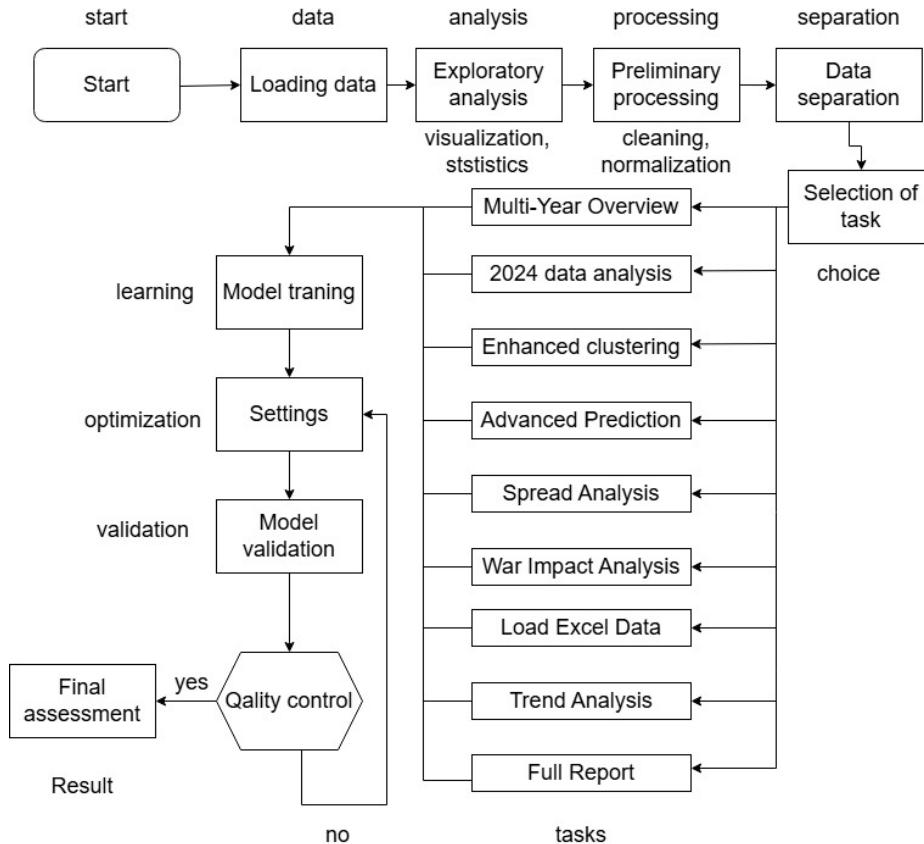
Cluster analysis was applied for grouping of Ukrainian regions based on the similarity of the epidemiological situation (HIV prevalence rate, growth dynamics, socio-economic characteristics, etc.):

- identification of typical infection spread profiles, allowing for a more targeted approach to planning countermeasures;
- improvement of visual and statistical understanding of the data prior to building forecasting models.

Figure 1 presents the overall scheme of the process for building and using the information model for analyzing HIV morbidity, which addresses the tasks described above.

The diagram covers all key stages - from data loading and processing to task selection, model training, its validation, and final evaluation. The central element is the block "Task selection",

which integrates functional modules such as Multi-Year Overview, Enhanced Clustering, Spread Analysis, War Impact Analysis. These modules represent specific directions of data analysis. Next, parameter tuning, model training, and quality verification take place with the possibility of iterative optimization. Upon achieving acceptable accuracy, the final model evaluation is performed.



**Figure 1:** Diagram of Functional Blocks.

Input data for the analysis and forecasting of HIV infection incidence are based on official statistics regarding the number of individuals newly diagnosed with HIV, as well as related medical and social indicators. The data cover quarterly or annual statistics over several years, allowing for the analysis of infection spread trends and the construction of forecasting models based on machine learning methods.

Main characteristics:

1. year and quarters – data are structured by years and quarters, allowing consideration of seasonal fluctuations and analysis of the dynamics of new HIV infection cases.
2. number of new HIV cases – for each quarter or year, the number of new infection cases is indicated. This is the primary indicator for analyzing epidemiological dynamics.
3. medical indicators – additional parameters are taken into account, such as:
  - the level of HIV testing coverage among the general population and key groups,
  - the percentage coverage of antiretroviral therapy,
  - the number of late diagnoses (CD4).

Regional data – statistics are provided for each oblast of Ukraine (except territories where official statistics are unavailable, Luhansk, parts of Donetsk, and the Autonomous Republic of Crimea).

Data on HIV prevalence in Ukraine were obtained from official materials of the State Institution "Public Health Center of the Ministry of Health of Ukraine." This allows for cluster analysis of regions to identify territories with the highest risks.

The input data are a critical element for evaluating the effectiveness of the machine learning models applied in the process of forecasting HIV incidence. After completing all stages of data processing and training, the modeling results allow not only for making predictions but also for quantitatively assessing their accuracy in relation to real data (Table 1).

**Table 1**  
Output Data After Forecasting Incidence and Prevalence.

Model	Mean Absolute Error (MAE)	R <sup>2</sup> Score	Root Mean Squared Error (RMSE)
Random Forest	34.0	0.918	45.1
K-Nearest Neighbors (KNN)	43.0	0.816	67.3
Ensemble Methods	16.0	0.977	26.5
Moving Average	92.0	0.513	90.8

As shown in Table 1, the Ensemble Methods-based model demonstrated the lowest MAE (16.0) and RMSE (26.5), indicating its highest accuracy among the tested algorithms. Since MAE and RMSE are absolute error metrics, they represent the average deviation of the forecast from the actual values in the same units as the target variable — namely, the number of HIV cases. This means that, for example, when using Random Forest, the average error is 34 cases, whereas with Ensemble Methods it is only 16 cases. RMSE, which is more sensitive to large deviations, also confirms the advantage of ensemble methods. Thus, models with lower MAE and RMSE values better reflect the actual dynamics of incidence and can be recommended for practical use in forecasting the spread of HIV.

### 3. Analysis of the obtained results

During the study, statistical indicators of HIV incidence across the regions of Ukraine for the period 2019–2024 were analyzed. Based on these data, the dynamics of key indicators were constructed: the number of new cases, testing coverage, the number of patients with CD4 counts below 350 and others.

Identified trends. Regions with high and stable incidence: Odessa, Dnipropetrovsk, and Mykolaiv regions demonstrated consistent growth in indicators, indicating a sustained high level of risk. These regions also experienced a high burden on the healthcare system. Regions with declining incidence: Khmelnytskyi, Ternopil, and Zakarpattia regions recorded a gradual decrease in the number of new cases. This may indicate the effectiveness of preventive measures or improvements in the testing system. Regions with unstable dynamics: Poltava, Cherkasy, and Kharkiv regions showed irregular data patterns, complicating the interpretation of results and requiring additional monitoring.

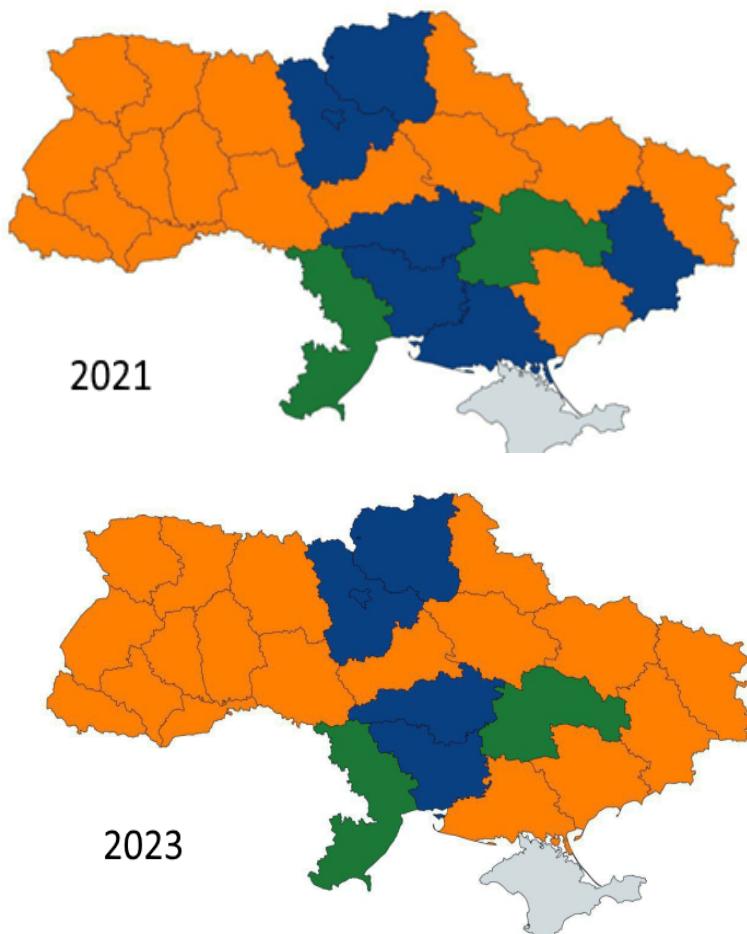
Results of cluster analysis - using the K-Means method, several clusters of regions were identified:

- Cluster 1: regions with a high level of new cases and active testing – Dnipropetrovsk, Odesa, Kyiv.
- Cluster 2: regions with a medium level of prevalence and stable dynamics – Vinnytsia, Lviv, Zaporizhzhia.
- Cluster 3: regions with a low detection rate and probable undercoverage of the population – Chernivtsi, Rivne, Volyn.

The obtained results allow us to assert that the regional approach to the prevention and treatment of HIV infection should be differentiated. Regions with a high burden require intensive support, whereas in regions with low indicators, it is important to ensure the accuracy and completeness of reporting.

Based on the clustering maps for 2019 and 2023, significant changes in the cluster structure of certain regions can be observed, in particular: Donetsk region in 2021 still belonged to cluster 2, but in 2022 and 2023 it shifted to cluster 1 (Figure 2), which may indicate a decrease in officially registered cases or problems with data access due to the armed conflict. Kherson region shows a similar dynamic, changing clusters from 2 to 1 (Figure 2). This may also be related to the temporary occupation of part of the territory, changes in the reporting system, or a decrease in case detection due to limited access to medical services.

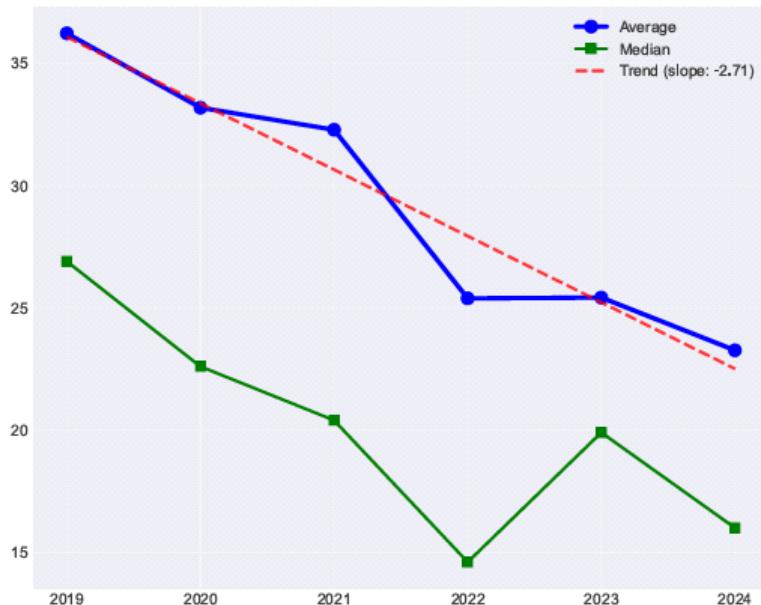
These changes indicate that the cluster structure is not static and is sensitive to socio-economic and political changes. Therefore, dynamic cluster analysis is an important element of epidemiological monitoring. Within the scope of this study, the epidemiological development of HIV infection was forecasted using machine learning methods.



**Figure 2:** Clustering maps (based on the example of 2021 and 2023).

Let's analyze an example of comparison statistical data and data obtained through clustering using k-means method in 2023. Based on the available statistics [7], it can be observed that in this year the highest incidence rates are recorded in Odesa (78.8), and Dnipropetrovsk (42.1) regions. The visualization of the clustering results provides a generalized representation of these data by grouping regions with similar levels of HIV incidence into clusters (Figure 2). Cluster One: includes the majority of Ukrainian regions with a relatively low level of HIV incidence. This cluster covers a significant part of northern, central, and western Ukraine. Cluster Two: includes regions with a higher level of HIV incidence, mainly located in the eastern part of the country. Cluster Three: corresponds to the regions with the highest level of HIV incidence (Dnipropetrovsk and Odesa), which stand out according to statistical data. These regions form a separate cluster, indicating the severity of the HIV problem in these areas.

The clustering map simplifies the understanding of the geographical distribution of HIV, as it replaces the need to analyze individual figures for each region with a visual representation of regional trends. The modeling approaches employed included linear regression (Figure 3), the Random Forest algorithm (Figure 4), and ensemble methods, notably gradient boosting.

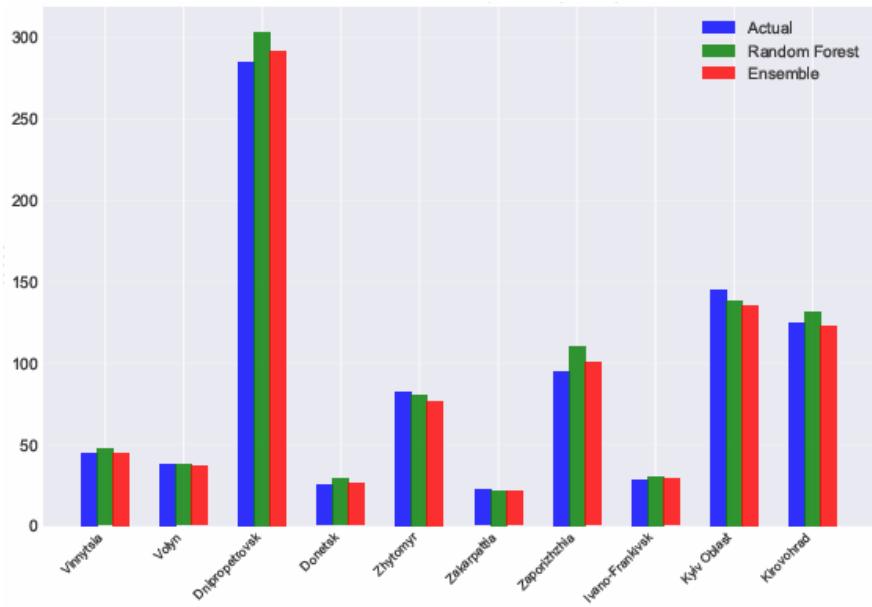


**Figure 3:** Linear Regression.

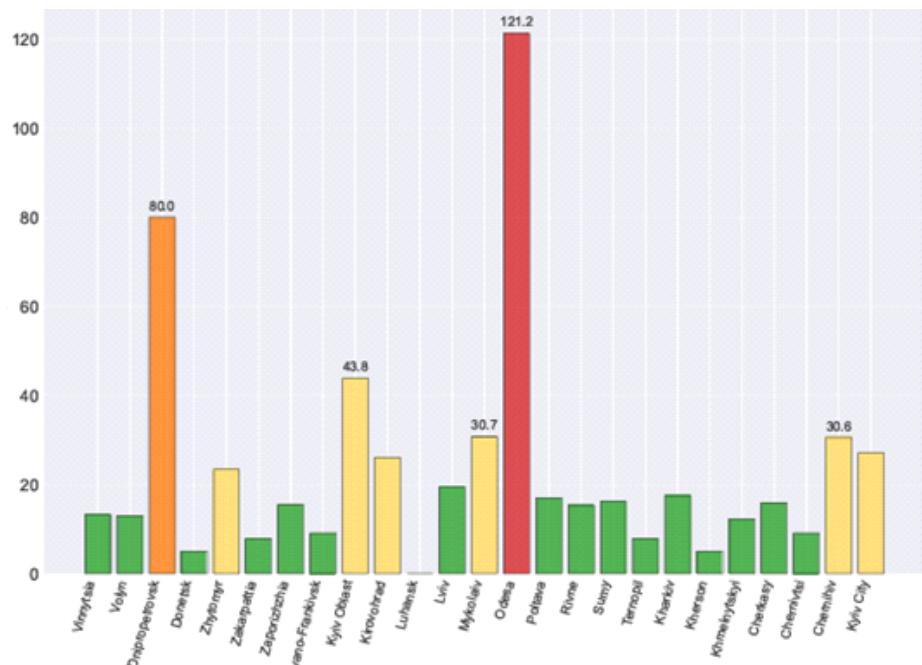
The models were trained on preprocessed time series data.

The resulting forecasts enabled the estimation of the number of new infection cases (Figure 4), the prevalence of HIV across different regions, as well as the growth rates of incidence (Figure 5).

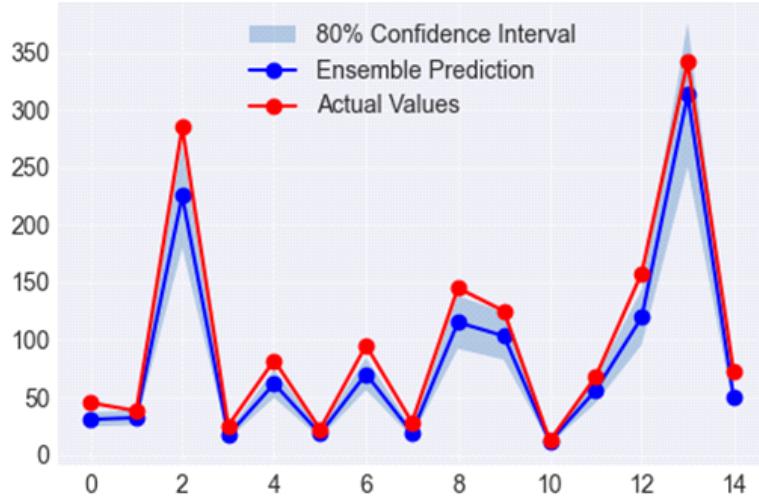
The Ensemble model demonstrated several advantages on this dataset, including the highest accuracy among all tested models, as shown in Figure 6 ( $R^2 = 0.997$ ,  $MAE = 3.33$ ). This model combines the strengths of multiple algorithms, which reduces the errors inherent to individual models and ensures greater stability of the results. Consequently, it is less sensitive to variations in the data or noise, making it a robust choice for predicting HIV incidence.



**Figure 4:** Random Forest Algorithm and Ensemble Methods.



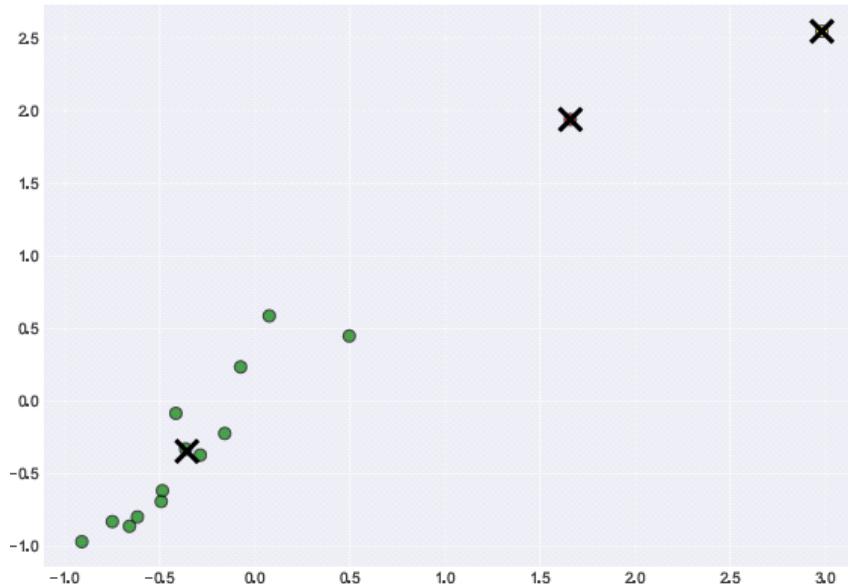
**Figure 5:** HIV prevalence across different regions and the rates of incidence growth.



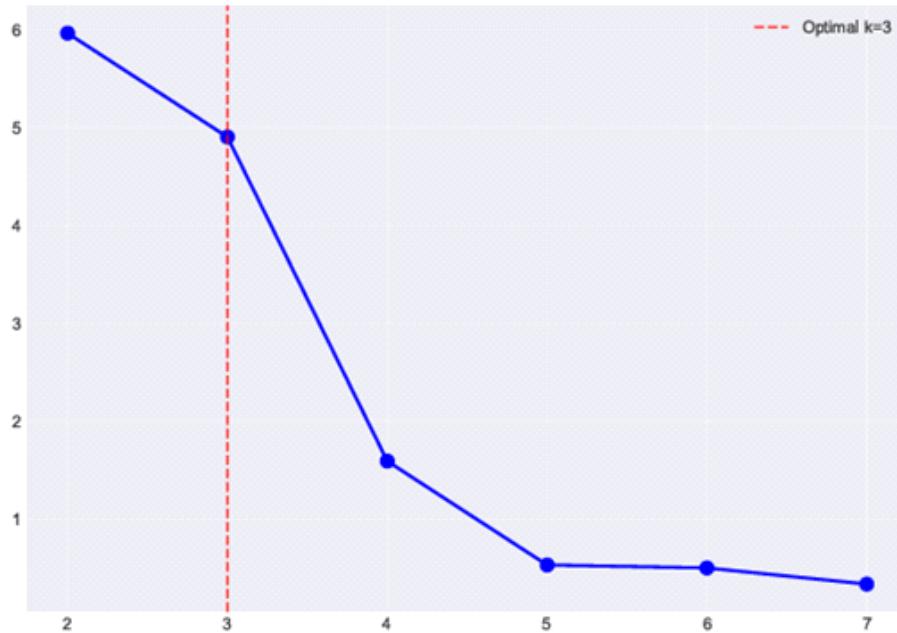
**Figure 6:** Accuracy of the ensemble model for predicting HIV incidence across regions of Ukraine.

For a deeper understanding of the changes in the epidemic situation in Ukraine, an analysis of the dynamics of HIV spread during the period 2015–2024 was conducted. The graphical representation of the data made it possible to identify both long-term trends of increasing or decreasing incidence, as well as short-term anomalies that may indicate the influence of external factors or changes in reporting practices.

To identify HIV infection spread zones in Ukraine based on epidemiological similarity, the K-means clustering algorithm was applied (Figure 7). During the analysis, the selection of three clusters was justified (Figure 8), representing regions with high, medium, and low levels of HIV prevalence. This approach enabled the structuring of data and identification of typical infection spread profiles, which is beneficial for regional planning of prevention and control measures.



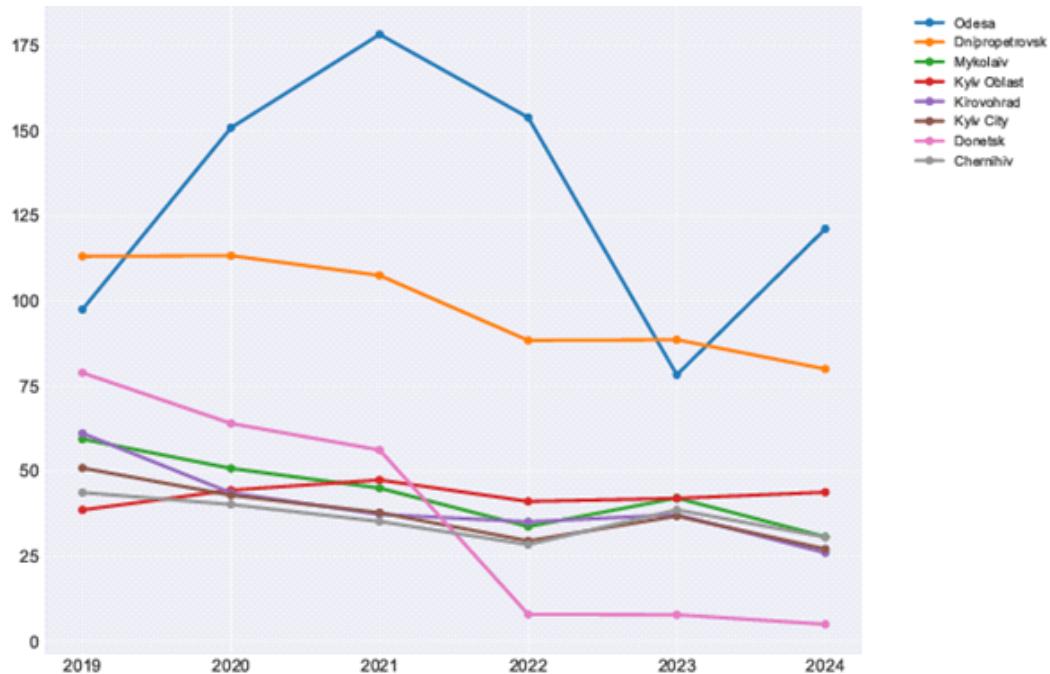
**Figure 7:** K-means Clustering Algorithm.



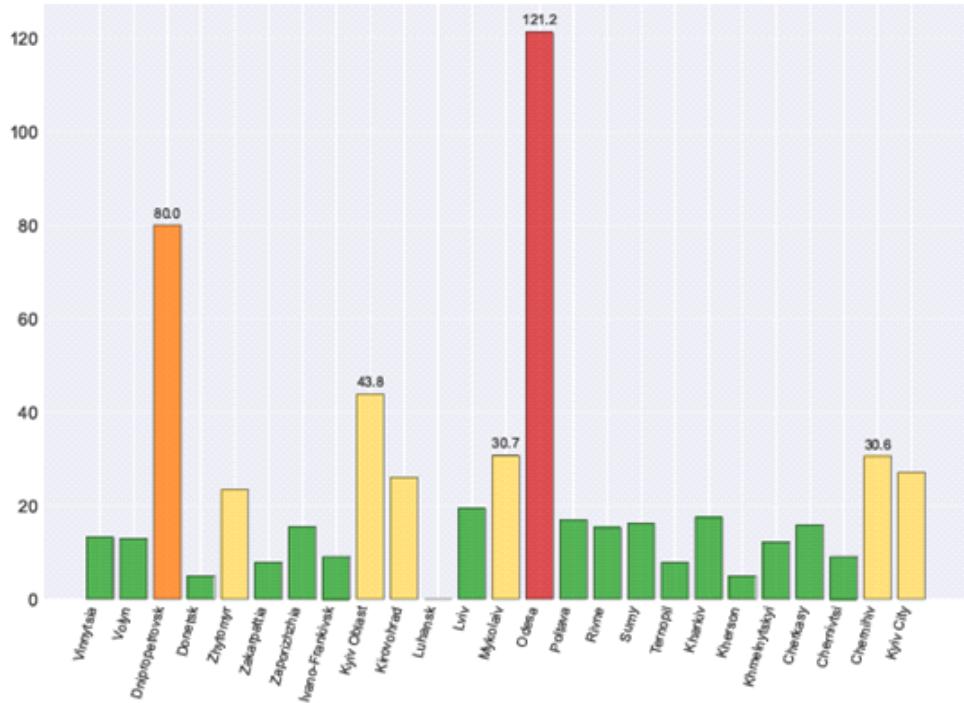
**Figure 8:** Elbow Method Plot for Selecting Three Clusters.

The results obtained demonstrate a high degree of consistency between the predicted and actual values, indicating a strong quality of the modeling process.

The collected statistical data for the period 2019–2024 were structured and analyzed, followed by the construction of a geographic visualization (Figure 9). The visual map enables prompt identification of regions with the highest incidence rates, which is critical for assessing territorial risk and strategic planning of healthcare interventions (Figure 10, example from 2024).



**Figure 9:** Geographic Visualization of HIV Infection Spread.



**Figure 10:** Visual Map for the Year 2024.

In the study also a statistical analysis was conducted in relation to emergency situations, specifically examining the impact of the COVID-19 pandemic and the war of 2022. This functionality was implemented programmatically, with the results presented in the form of a comprehensive analytical report. The analysis covered data from 2019 to 2024 and focused on the differences between the pre-war period (2019–2021) and the wartime period (2022–2024).

It was established that until 2020, the annual number of HIV antibody screening tests in Ukraine remained stable, ranging from 2.3 to 2.5 million. However, due to quarantine restrictions associated with the COVID-19 pandemic and as a consequence of the war in 2022, the number of tests dropped to a record low of 1.6 million. In 2023, the volume of examinations increased by 40%, reaching 2.25 million, primarily owing to a 70% rise in testing initiated by healthcare professionals and patients. Accordingly, in 2023, the HIV antibody testing rate per 100,000 population increased 1.6-fold compared with 2022. Between 2019 and 2023, the overall HIV prevalence decreased from 0.9% to 0.6%. This decline can be attributed to the exclusion, starting in 2022, of HIV/AIDS statistical data from the Donetsk, Luhansk, Zaporizhzhia, and Kherson regions, which were previously known for high HIV prevalence.

From 2019 to 2023, a marked decrease in the HIV incidence rate was observed against the backdrop of the COVID-19 pandemic and the war of 2022, falling from 42.5 to 28.4 per 100,000 population. Traditionally, the highest HIV incidence per 100,000 population has been recorded in the south-eastern region of Ukraine. In 2023, the highest rates were registered in Dnipropetrovsk and Odesa regions. Active migration flows in 2022–2023 contributed to a 1.5–2-fold increase in HIV incidence in the western and central regions of Ukraine. For example, in Volyn region, the rate rose from 9.7 to 16.6 per 100,000 population; in Zakarpattia, from 5.9 to 9.0; in Lviv, from 14.6 to 21.2; in Khmelnytskyi, from 11.0 to 15.7; in Chernivtsi, from 7.1 to 10.4; and in Chernihiv, from 28.4 to 38.6. In Kyiv, the indicator increased from 29.5 to 36.8.

It was also found that at the current stage of the HIV epidemic, as in previous stages, the majority of HIV-positive individuals reside in urban areas (77% of new HIV cases in 2023). Among them, 65% are men, and 78% belong to the 25–49 age group. It is noteworthy that the epidemic is aging: over the past five years, the proportion of individuals first diagnosed with HIV at age 50 or older has increased from 16% to 19%.

During the war, with the support of international organizations, it was possible to rapidly organize the provision of preventive and therapeutic HIV services in healthcare facilities of various profiles. This led to the continuation, in 2022–2023, of the pre-COVID trend in HIV transmission routes: the proportion of sexually transmitted cases increased (from 68.3% to 74.6%), while parenteral transmission through injecting drug use decreased (from 31.3% to 25.4%).

The data demonstrate a gradual decline in the share of parenteral transmission (through injecting drug use) and an increasing role of sexual transmission. This indicates a transformation of the epidemic: from concentration within key risk groups to wider dissemination in the general population, necessitating a reorientation of prevention strategies.

Conclusions based on the results of the performed calculations:

Preliminary data preprocessing revealed no missing values in the dataset. The data were scaled, checked for anomalies, and 95% of the useful information was retained after cleaning.

An initial clustering was performed using the K-means method, where the optimal number of clusters was determined by the elbow method. The clusters were successfully visualized on the map of Ukraine for different years.

For forecasting disease incidence, the Random Forest method was employed, achieving high accuracy with a mean absolute error of approximately 12.6, root mean squared error of about 15.4, and  $R^2 = 0.937$ . The model was trained in approximately 0.4 seconds. Additionally, the K-Nearest Neighbors method was tested, which demonstrated lower accuracy on the test set with an  $R^2$  of 0.82. To smooth the time series data, a moving average method was applied, which reduced noise influence and improved forecast stability. The visualization of results was carried out using an interactive map of Ukraine, where each region was automatically colored according to the risk level or cluster membership. It can be observed that HIV incidence in Ukraine exhibits a distinct geographic distribution. The highest concentration of cases is found in the eastern and southern regions, whereas most oblasts in central and western Ukraine demonstrate lower incidence rates. Given the varying levels of HIV incidence across different clusters, a differentiated approach is necessary for the development and implementation of prevention programs. For instance, in regions belonging to the third cluster (green), more intensive measures focused on prevention, testing, and treatment of HIV are required. Clustering allows for the identification of priority regions for the implementation of HIV control programs. Concentrating resources and efforts in regions with the highest incidence rates may be more effective than distributing them evenly across the entire country. The identified clusters can be used for further investigation of factors influencing the spread of HIV in each region. For example, socio-economic conditions, behavioral factors, and accessibility of medical services can be studied across different clusters.

## 4. Conclusions

This study conducted a detailed analysis of the epidemiological situation of HIV infection in Ukraine based on multi-year statistical data. Special attention was given to the dynamics of disease prevalence in the regions before and after the onset of the full-scale war, as well as to identifying trends and changes in the regional distribution of indicators.

To achieve the stated objectives, a comprehensive set of approaches was employed, including:

- clustering of Ukrainian regions using the K-means method to identify groups of regions with similar HIV prevalence characteristics;
- identification of regions with high, medium, and low levels of infection spread, enabling clear delineation of risk zones;
- analysis of cluster stability over time demonstrated that some regions change their cluster affiliation, while others remain stable (e.g., Dnipropetrovsk and Odesa).

For the forecasting tasks, machine learning models were implemented and tested, including Random Forest, ensemble methods, and the K-Nearest Neighbors algorithm.

The evaluation of model accuracy using MAE, RMSE, and  $R^2$  metrics confirmed the high effectiveness of the forecasts. The outcome of the work was the development of an information system in the form of a software application, which: allows users to select forecast parameters (year, quarter); performs clustering; visualizes results through graphs and maps; provides the ability to save results in a convenient format. The information system is designed for practical application in the field of public health, supports visual analysis of trends, and can be adapted for other infectious diseases or medical statistics indicators. Thus, the use of machine learning methods combined with visual tools significantly enhances the quality of epidemiological analysis and improves the effectiveness of decision-making in healthcare.

A promising direction of future investigation will be the assessment of complex socio-demographic factors. In conclusion, data on population migration, especially internal movements of individuals, are critically important, as parts of the movement of large groups of people can affect access to medical services, testing and treatment. Changes in the behavior of the population caused by war, such as increased use of drugs and alcohol, can also have a negative impact on the expansion of HIV. The integration of these complex factors will allow us to create more accurate models and effective approaches to counteractions to HIV/AIDS epidemic.

## Acknowledgements

This 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".

## Declaration on Generative AI

During the preparation of this work, the authors did not use Generative AI tools.

## References

- [1] N. S. Hoidyk, *Overview of the epidemiological situation of HIV/AIDS in Odesa region*, 2009.
- [2] O. A. Holubovska, O. I. Vysotska, O. V. Bezrodna, "The role of primary health care in patients with blood-borne infections (HIV infection and hepatitis B and C)," *Infectious Diseases*, no. 1 (2017): 5–8.
- [3] A. S. Dovbysh, A. V. Vasylyev, V. O. Liubchak, *Intelligent Information Technologies in E-learning*, Sumy: Sumy State University, 2013, 172 p.
- [4] Yu. P. Zaichenko, *Fundamentals of Designing Intelligent Systems: A Textbook*, Kyiv: Slovo, 2004, 352.
- [5] L. D. Kaliuzhna, L. V. Hrechanska, "Associations of sexually transmitted infections in HIV-infected individuals," *Ukrainian Journal of Dermatology, Venereology, Cosmetology*, no. 1 (2004): 78–80.
- [6] V. F. Mariievskyi, S. I. Doan, "Determining promising directions for countering HIV infection in the current epidemic situation," *Infectious Diseases*, no. 4 (2013): 17–22.
- [7] D. Chumachenko, I. Menialov, K. Bazilevych, T. Chumachenko, and S. Yakovlev, "Investigation of Statistical Machine Learning Models for COVID-19 Epidemic Process Simulation: Random Forest, K-Nearest Neighbors, Gradient Boosting," *Computation*, vol. 10, no. 6, p. 86, 2022, doi: <https://doi.org/10.3390/computation10060086>.
- [8] A. Mohammadi, et al., "Comparative study of linear regression and SIR models of COVID-19 propagation in Ukraine before vaccination," *Radioelectronic and Computer Systems*, vol. 2021, no. 3, pp. 5–18, 2021, doi: <https://doi.org/10.32620/reks.2021.3.01>.
- [9] D. Chumachenko *et al.*, "Methodology for assessing the impact of emergencies on the spread of infectious diseases," *Radioelectronic and Computer Systems*, vol. 2024, no. 3, pp. 6–26, Aug. 2024, doi: <https://doi.org/10.32620/reks.2024.3.01>.
- [10] Public Health Center of Ukraine, *HIV/AIDS Statistics*. Available at: <https://phc.org.ua/kontrol-zakhvoryuvan/vilsnid/statistika-z-vilsnidu>