IMPLEMENTATION OF BIG DATA AND AI TECHNOLOGIES TO IMPROVE THE EFFICIENCY OF MEDICAL ANALYTICS

A.S. Aymurzaeva
assistant NSTU
https://doi.org/10.5281/zenodo.17510142

ARTICLE INFO

Received: 29th October 2025
Accepted: 30th October 2025
Online: 31st October 2025

KEYWORDS

To ensure the quality of the analysis and address the aforementioned issues, reliable and up-to-date sources are used

ABSTRACT

In the modern world, there is a rapid increase in the use of Big Data technologies across various areas of human activity. The study of Big Data applications in medical research is particularly relevant, as it holds great potential for significantly improving healthcare quality and preventing many diseases.

Today, significant progress has already been made in the field of medical data analysis, enabling improvements in diagnosis, treatment, and disease prevention. However, despite the existing knowledge, there are still many gaps and unanswered questions that require further investigation. In particular, the identification of patterns and the establishment of cause-and-effect relationships within large medical datasets remain insufficiently explored..

In the modern world, there is a rapid increase in the use of Big Data technologies across various areas of human activity. The study of Big Data applications in medical research is particularly relevant, as it holds great potential for significantly improving healthcare quality and preventing many diseases.

Today, significant progress has already been made in the field of medical data analysis, enabling improvements in diagnosis, treatment, and disease prevention. However, despite the existing knowledge, there are still many gaps and unanswered questions that require further investigation. In particular, the identification of patterns and the establishment of cause-and-effect relationships within large medical datasets remain insufficiently explored.

This paper focuses on reviewing modern approaches to the analysis of large-scale medical data and demonstrates how such analysis can help uncover new patterns and causal relationships in medical research. Examples of such analytical applications in medicine are presented, along with potential directions for future research in this field.

To ensure the quality of the analysis and address the aforementioned issues, reliable and up-to-date sources are used, including published scientific papers, results of international research projects, and data from statistical agencies. This approach allows for the formation of an objective and well-founded understanding of the current state of the problem and possible ways to address it.

In the last few decades, several studies have predicted the incidence of diabetes and its global prevalence for different countries around the world, using diverse data and methods of analysis. King et al. estimated diabetes prevalence by the number of diabetics aged 20 years and over for every country in the world in three time points: 1995, 2000, and 2025. Other variables were calculated, such as the gender proportion, urban–rural proportion, and age groups of the population who suffer diabetes. The data used in this study were obtained from the World Health Organisation's (WHO) global database, which was collected from 75 societies representing 32 countries. To estimate the number of diabetes cases in every country in the world, data gathered from the WHO were linked to demographic estimates and projections released by the United Nations. The study assumed that, besides ethnicity, other factors contribute to diabetes trends, such as population size, sex, age structure, and urbanisation level. All data sources were analysed using logistic regression modelling. The global prevalence of diabetes in 1995 was estimated to be 4.0%, predicted to increase to 5.4% by the year 2025. This was higher in developed than developing countries developed an updated report in 2004, adding new data and various techniques to estimate age-specific diabetes prevalence. This study estimated the prevalence of diabetes, and the number of diabetics in all age groups, for the years 2000 and 2030. For this study, diabetes prevalence data according to age and sex were collected from a restricted range of countries and extrapolated to all 191 states represented by the WHO. For people aged 20 and over, the data were obtained using population-based studies, using WHO criteria for diagnosing diabetes. In order to generate smooth, age-specific estimates, DisMod II version 1.01 software was used, which is a mathematical model for analysing estimations of disease with regard to occurrences, prevalence, and mortality rates. It was estimated that the global prevalence of diabetes for all age groups was 2.8% in 2000, projected to rise to 4.4% in 2030; a total of 171 million diabetic people in 2000 was predicted to increase to 366 million by 2030. A study by Shaw et al aimed to predict the number of diabetes cases globally for 2010 and 2030. Studies were collected from the 91 countries in which they were published between January 1989 and March 2009. A total of 133 studies that used a population-based method to evaluate the prevalence of diabetes were selected, applying the diagnostic measures of the WHO or the American Diabetes Association (ADA). Age- and sex-specific diabetes prevalence in people aged 20–79 was calculated using logistic regression modelling. These calculations were applied to the estimates of national populations to estimate the number of diabetic people for all 216 countries for 2010 and 2030. It was estimated that the global prevalence of diabetes within the 20–79 age group was 285 million adults in 2010, projected to rise to 439 million by 2030.

For this study, the Python programming language and its data analysis libraries — such as Pandas, NumPy, Matplotlib, and Seaborn — were used. The dataset utilized in the research contained information about patients, including glucose level, blood pressure, insulin level, skin thickness, body mass index (BMI), age, and number of pregnancies. All patients in the dataset were women aged 21 years or older.

Using correlation analysis, a matrix was constructed to display the relationships between different variables. The study revealed the following dependencies:

- **Insulin level** shows a strong correlation with the development of diabetes. This confirms the well-known fact that insulin secretion plays a critical role in glucose regulation and that disruptions in this process can lead to the onset of diabetes.
- **BMI > 30** (obesity) and advanced age are associated with a higher prevalence of diabetes, highlighting the need for preventive measures and monitoring of these factors.

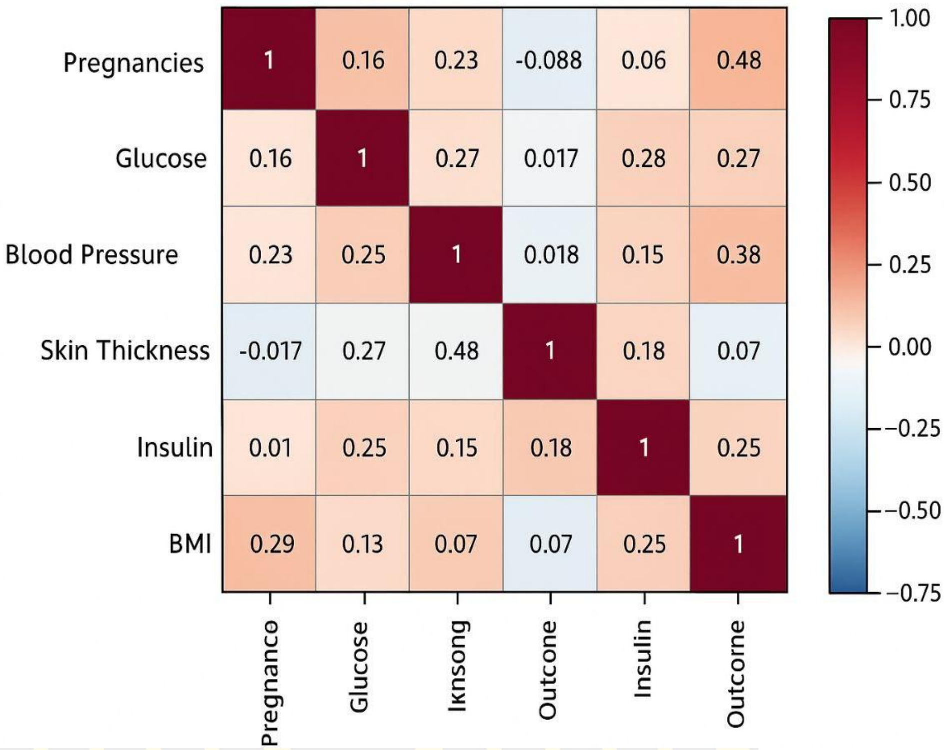

Figure 1. Correlation Matrix

- **Number of pregnancies** significantly influences the development of diabetes. This may be related to hormonal changes during pregnancy, which can temporarily disrupt insulin secretion and lead to gestational diabetes.

Despite these findings, the study has several limitations. First, the dataset may not include complete information on all factors affecting diabetes development. Larger and more diverse datasets could help refine the results and identify additional contributing factors. Second, correlation does not necessarily imply causation, and further research may be needed to determine the exact mechanisms of interaction between variables.

The study confirmed strong correlations between various medical indicators, which may suggest potential causal relationships and contribute to improving disease diagnosis and treatment.

Future research directions may include expanding the dataset and incorporating other potentially significant variables, such as genetic factors, diet, physical activity, and socio-economic status. Additionally, interaction analysis between different factors could be conducted to identify possible combined effects on diabetes development.

Further studies could also focus on determining causal links between the factors considered and diabetes onset. This could involve experimental and clinical research, as well as the application of more advanced statistical methods, such as regression analysis and machine learning.

References:

1.Kulikova, N. I. Artificial Intelligence in Healthcare: Current Approaches and Prospects. // Information Technologies in Medicine, No. 3, 2023.

2.Lavrentyev, A. A. Big Data Processing Technologies in Healthcare Systems. // Journal "Science and Innovation," No. 12, 2021.

3.Shumilov, V. P. Application of Neural Networks for Medical Data Analysis. — Moscow: Infra-M, 2022.