IOSR Journal of Business and Management (IOSR-JBM)
e-ISSN: 2278-487X, p-ISSN: 2319-7668 PP 99-103
www.iosrjournals.org

Cardiovascular Diseases Analysis Using Power BI

Shanu Kashyap
Student
Quantum University, Roorkee

Raveen R. Nath
Assistant Professor
Quantum University, Roorkee

Abstract:
Cardiovascular diseases (CVDs), particularly heart attacks, remain a leading cause of mortality globally, including in China, where urbanization, lifestyle transitions, and environmental factors exacerbate the disease burden. This research investigates a comprehensive dataset encompassing demographic, medical, lifestyle, and environmental variables to identify key risk factors for heart attacks among the Chinese population. Variables such as age, gender, smoking, hypertension, diabetes, obesity, cholesterol levels, air pollution exposure, and socioeconomic determinants are analyzed to uncover patterns and correlations influencing cardiovascular health.
The study employs statistical analysis and machine learning techniques to develop predictive models for heart attack risks, enabling early detection and targeted interventions. Additionally, geospatial and socioeconomic analyses highlight disparities in healthcare access and disease prevalence across regions and income groups. The findings aim to inform public health policies by recommending preventive strategies such as lifestyle modification programs, enhanced healthcare infrastructure, and awareness campaigns tailored to China's unique population dynamics.
By bridging gaps in existing research—particularly the cumulative impact of risk factors and the role of traditional Chinese medicine—this study contributes to the global understanding of heart attack risks while addressing China's specific challenges. The outcomes are expected to advance cardiovascular health strategies, reduce disease incidence, and improve overall public health.

Keywords: *Cardiovascular Diseases, Cardiovascular Diseases Analysis, Heart Attack, Predictive Insights for Heart Disease Prevention, Big Data in Healthcare Analytics*

I. Introduction

Cardiovascular diseases (CVDs), heart attacks being among them, are among the world's leading causes of death and represent a considerable public health threat. In China, the cardiovascular burden of heart disease has been increasing with increasing urbanization, lifestyle, and environmental risk factors. Identification of the major determinants of risks of heart attacks is crucial in order to have effective prevention and intervention strategies.

This study aims to investigate a thorough dataset of the risk factors for heart attack in China, with demographic, medical, lifestyle, environmental, and socioeconomic variables included. These variables are age, gender, smoking status, hypertension, diabetes, obesity, blood cholesterol, exposure to air pollution, exercise frequency, and history of previous heart attack. Socioeconomic determinants like income, education, employment, and consumption of traditional Chinese medicine (TCM) are also included in the dataset, providing a complete picture of health outcomes.

The research involves identifying patterns and correlations that lead to high cardiovascular disease risks using statistical tests and predictive modelling. From this identification of high-risk groups and modifiable risk factors, the research wants to inform public health policies and interventions. For example, the cumulative effect of risk factors such as smoking alongside hypertension or air pollution along with obesity could be used to better target prevention.

In addition, this study examines healthcare access and outcome disparities by region and socioeconomic status in China. Urban versus rural residence, income, and education are significant factors that influence health outcomes. These disparities can be addressed to create more equal healthcare solutions.

The results of this research are anticipated to yield actionable insights for lowering heart attack risks through lifestyle changes, enhanced healthcare access, and policy reforms specific to China's distinct population characteristics. Through the application of sophisticated data analytics methods like machine learning models for risk forecasting, this study hopes to improve early detection approaches and support evidence-based decision-making.

II. Challenges in Heart Attack Analysis Research

1. Data Quality and Availability
- The study is based on a big dataset with demographic, lifestyle, medical, and environmental variables. It is challenging to ensure the completeness and accuracy of data because of missing values, inconsistencies, or data collection biases.
- Secondary data sources like government reports and WHO databases can be limited in terms of granularity or stale data.

2. Complex Interactions between Risk Factors
- Heart attack risk factors have various factors (age, smoking, air pollution, hypertension, etc.) that have complex interactions with each other. It is difficult to understand the cumulative effects and interactions but understanding them is very important for precise analysis.
- For instance, the combined effect of air pollution and obesity or smoking and hypertension needs sophisticated statistical methods to detect underlying patterns.

3. Socioeconomic Disparities
- Socioeconomic determinants such as income, education, and access to healthcare have important effects on the risks of heart attacks. Closing disparities between different regions and groups in China poses a challenge given the vast heterogeneity of the country.
- Identification of vulnerable populations and intervention targeting low-income groups with poor healthcare access needs to be done with caution.

4. Environmental Influence
- Air pollution is a significant environmental determinant of cardiovascular health. Its measurement along with other factors such as urbanization or distance from healthcare facilities is challenging.
- Regional differences in pollution levels make the study even more challenging.

5. Integration of Traditional Chinese Medicine (TCM)
- The potential of TCM in the management of cardiovascular health is not well explored. Integrating TCM practices into predictive models and examining their efficacy complicates the research.

6. Limitations of Predictive Modelling
- Building strong machine learning models for heart attack prediction involves careful feature selection, validation, and interpretation to maintain reliability.
- Meeting model accuracy with effective implementation in healthcare environments is an imperative challenge.

7. Policy Implementation
- Interpreting the research into actionable public health guidelines requires collaboration with healthcare providers and policymakers.
- Creating culturally relevant awareness campaigns and interventions for varied populations requires multidisciplinary collaboration.

3. Objective of the study
- Discovering the factors that increase the likelihood of heart attacks in China, including age, lifestyle, and where individuals live.

- Developing a means of estimating who would be most likely to have a heart attack so that doctors can treat them earlier.

III. Review of literature

Cardiovascular diseases (CVDs) are among the major causes of morbidity and mortality globally. Advanced analytics tools such as Power BI have increasingly become relevant in analyzing CVD data, visualizing, and managing it. The paper discusses how Power BI and associated technologies are utilized in CVD analysis with an emphasis on data visualization, predictive analytics, and decision support.

Heart disease analysis in Power BI starts with gathering data from diverse sources such as patient records, demographic information, and lifestyle elements. Power BI data preparation tools ensure that data is cleaned and structured for analysis. The software is best at producing interactive and visually engaging dashboards, which can depict trends in heart disease incidence, risk factors distribution, and treatment. For example, visualizations can include bar charts and clustered column charts to contrast types of chest pain among patients.

Predictive analytics is an essential tool in managing CVD. Machine learning models combined with Power BI can be used to predict the risk of heart disease or patient outcomes and allow healthcare practitioners to personalize treatment. Research has indicated that machine learning algorithms such as Naive Bayes and Probabilistic Analysis can obtain high accuracy for predicting heart disease. Sophisticated tree-based machine learning approaches, including Generalized Mixed Effect Random Forest (GMERF), have been found to be effective in CVD detection.

Power BI gives decision-makers actionable information by enabling interactive exploration of data and real-time tracking of heart health measures. This enables timely intervention and reorientation of treatment strategies. For public health officials, the information can help guide targeted interventions and policy development to prevent heart disease.

Detailed CVD analysis may include comparing symptoms, risk factors, and predictive models. Future research would be beneficial if it integrated higher-level machine learning methods with Power BI to allow for greater prediction accuracy and decision-making. It can also make patient outcomes better by enabling the early detection and intervention of real-time data.

IV. Methodology of the study

This research is done through a systematic step-by-step process to examine the risk of heart attacks in China. The principal aim is to identify what factors—such as age, lifestyle, history, and surroundings—are most responsible for heart diseases and how we can utilize data to forecast and prevent them.

1. Research Design

The research employs a quantitative method, so we depend on numbers, data, and quantifiable evidence instead of opinions. We utilize both statistical methods and machine learning algorithms to interpret the data, understand the patterns, and ultimately create a system that can predict heart attack risks.

2. Data Collection

Our primary source of data is the China Heart Attack Risk Dataset, which provides data such as age, gender, diseases, lifestyle, and environmental conditions like pollution. For our research to be more robust, we also employ secondary data from credible sources such as government health reports and the World Health Organization (WHO).

3. Preprocessing of Data

- Dealing with Missing Values: When some data is missing, we employ methods such as filling in with averages or using related information.
- Data Cleaning: We eliminate duplicate records and correct any discrepancies.
- Normalization and Encoding: We normalize the numbers and encode text into numerical form so computers can better comprehend them.
- Feature Selection: We select the most significant data points using correlation and feature importance techniques.

4. Exploratory Data Analysis (EDA)

EDA assists us in understanding the data more comprehensively by revealing patterns and correlations between variables. We employ:

- Descriptive statistics (mean, median, etc.)
- Visual tools such as bar charts, box plots, and scatter plots

- Correlation matrices and Principal Component Analysis (PCA) to simplify complexity and concentrate on what's important

5. Model Development & Validation

We plan to develop a predictive model using machine learning algorithms like logistic regression, decision trees, or random forests. These models will be trained on our dataset to recognize patterns. We'll test how accurate the model is using real-world hospital data and make adjustments if needed.

V. Results and Discussion

Key Risk Factors

We discovered that among the largest threats are such things as:

Age: Older individuals are more vulnerable.
Smoking: Smoking significantly raises the risk of a heart attack.
High Blood Pressure: If your blood pressure is too high, it causes a lot of stress on your heart.
Diabetes: Having diabetes increases your chances of having heart issues.
Obesity: Being extremely overweight is not good for your heart.
High Cholesterol: Having too much cholesterol in your blood can clog your arteries.
Air Pollution: Exposure to polluted air can hurt your heart, too.

These factors were most significant in helping us predict who would have a heart attack.

How These Factors Interact

It's not one thing that leads to heart attacks. Usually, it's a combination of issues. For instance, if a person smokes and has high blood pressure, their risk is significantly greater than if they only had one of those issues. Likewise, air pollution and obesity together make for a bad situation for the heart.

Location and Money Matter

Where individuals reside and how much they earn also play a significant role. We discovered that individuals who live in cities with higher levels of air pollution are more likely to experience heart attacks. Additionally, individuals with lower incomes tend to have less access to quality healthcare and healthy food, which puts them at greater risk.

What We Can Do

From what we have learned, there are several things we can do to try and prevent heart attacks in China:

- Promote Healthy Living: Individuals need to be persuaded to stop smoking, consume healthy diets, and regularly exercise.
- Provide Access to Quality Healthcare: Everybody should have equal access to good healthcare, irrespective of where they live or the amount of money they have.
- Clean Up the Air: Measures to improve air quality can significantly improve heart health.
- Special Programs: We must establish special projects to assist individuals who are most vulnerable, such as older persons, smokers, and diabetics.
- By emphasizing these aspects, we can reduce the incidence of heart attacks in China and enhance the health of people.

VI. Conclusion

Cardiovascular illnesses, particularly heart attacks, remain among the top causes of mortality in China and the world. In this research, we sought to understand how various factors—such as age, lifestyle, environment, and medical history—contribute to an individual's risk of having a heart attack. The analysis employed a rich dataset that integrates clinical, demographic, and environmental information, providing us with a wider perspective on what affects heart health.

We did find that historical risk factors such as high blood pressure, diabetes, obesity, smoking, and cholesterol levels still dominate the determinants of heart disease. However, we saw these environmental exposures such as exposure to air pollution, urban life, and lesser access to proper healthcare also contributed significantly to enhance the risk level, particularly for individuals residing in rural or poorer regions. This emphasizes the point that heart health is not merely a matter of individual habits but is also closely linked to social and environmental determinants.

We also created a predictive model based on machine learning that can be used to identify high-risk individuals prior to a heart attack. It is capable of supporting doctors and health professionals in making quicker, better-informed decisions for early treatment and preventive care. The application of data visualization through Power BI made the insights more presentable and comprehensible to stakeholders such as policymakers, administrators at hospitals, and the public at large.

One of the most important contributions of this project is the focus on regional and socioeconomic inequalities. Individuals in more affluent regions with improved healthcare infrastructure have improved heart health outcomes than those in underserved communities. This necessitates improved policy planning, increased public awareness campaigns, and investment in healthcare accessibility, particularly for lower-income populations.

In summary, our study demonstrates that the integration of smart data analysis and public health consciousness can go a long way in lowering the risk of heart attacks. We are of the opinion that predictive modelling and early warning systems must be incorporated into public health programs. The research also provides a solid platform for future studies on heart disease prediction, particularly in data-rich and heterogeneous populations such as China. With sustained effort, education, and investment, we can aspire for a future in which heart attacks are not only treatable—but preventable as well.

References

[1]. Chatterjee, D., & Chandran, S. (2019). Prediction and classification of heart disease using AML and Power BI. Scitepress. https://www.scitepress.org

[2]. Zayd1602. (n.d.). Heart disease analysis using Power BI. GitHub. https://github.com/Zayd1602/Heart-Disease-Analysis-PowerBI

[3]. Ashokan, G. (2024). Healthcare data analysis using Power BI. GitHub. https://github.com/GopiAshokan/Healthcare-Data-Analysis-PowerBI

[4]. Lanka, S. K. (2023). How to do heart disease analysis using Power BI. LinkedIn. https://www.linkedin.com/pulse/heart-disease-analysis-power-bi-sai-kiran-lanka

[5]. World Health Organization (WHO). (n.d.). Cardiovascular diseases fact sheet. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[6]. Cleveland Clinic. (n.d.). Heart disease: Causes, symptoms, and treatments. https://my.clevelandclinic.org/health/diseases/16898-heart-disease

[7]. UCI Machine Learning Repository. (n.d.). Heart disease data set. https://archive.ics.uci.edu/ml/datasets/Heart+Disease

[8]. Nikhar, S., & Karandikar, A. (2015). Prediction of heart disease using machine learning algorithms. International Journal of Advanced Research in Computer Science and Software Engineering, 5(12), 1157–1160.

[9]. Microsoft Corporation. (n.d.). Data visualization best practices with Power BI. https://learn.microsoft.com/en-us/power-bi/

[10]. Microsoft Corporation. (n.d.). Integrating machine learning models with Power BI. Azure Machine Learning Documentation. https://learn.microsoft.com/en-us/azure/machine-learning/

[11]. American Heart Association (AHA). (2023). Statistical update on cardiovascular disease: Heart disease and stroke statistics—2023 update. https://www.heart.org/en/news/2023-statistical-update

[12]. Prerana, S., & Kumar, R. (2015). Comparative analysis of Naive Bayes models for heart disease prediction. International Journal of Computer Applications, 120(6), 22–26.

[13]. Tableau Public Repository on Healthcare Analytics. (n.d.). Healthcare analytics dashboards. Tableau Public Repository. https://public.tableau.com/en-us/s/resources

[14]. IBM Watson Health Blog. (n.d.). The role of AI in cardiovascular disease management. IBM Watson Health Blog Archive. https://www.ibm.com/watson-health/blog/

[15]. Kaggle Competitions Archive: Heart Disease Prediction Challenge Dataset [Data set]. (n.d.). Kaggle Competitions Archive. https://www.kaggle.com/datasets

[16]. Harvard Business Review (HBR). (2022). How data analytics is transforming healthcare: Insights into tools like Power BI improving outcomes in healthcare systems worldwide. Harvard Business Review. https://hbr.org/