Predicting Heart Disease Risk Drivers Using Machine Learning Algorithms and Global Insights on Cardiovascular Mortalities

#603893
Louisville, Kentucky

Introduction

Cardiovascular diseases (CVD) have consistently been the leading cause of death across the globe.

- Around 18 million people die from CVD yearly, representing 32% of all global deaths.
- 85% of these CVD deaths are due to **heart attacks and stroke**.

Most CVD can be prevented by addressing behavioral risk factors such as:

- tobacco use
- unhealthy diet and obesity
- physical inactivity
- harmful use of alcohol

CVD has a heavy economic burden on society. Cost of CVD is estimated in multiple categories - screening primary prevention, acute hospital care, loss of productivity. Cost of CVD could be as high as US$20 trillion from 2010 to 2030.

Purpose

The main objective of this study is to **examine global distribution of cardiovascular death**s and its relationship with behavioral and socio-economic factors and **build machine learning models to predict heart disease risk** and understand drivers of heart disease.

The key insights from this analysis can help WHO and national and state-level governments to develop data-driven plans to identify risk factors and proactively manage this disease to reduce deaths and its economic impact on society.

Methods

Machine Learning models have been created using a curated dataset consisting of 1190 patients from the United States, United Kingdom, Switzerland, and Hungary. Global level insights have been created using multiple datasets including:

- Worldwide Socioeconomic attribute and CVD deaths
- Three decades (1990 to 2019) of country level deaths by cause

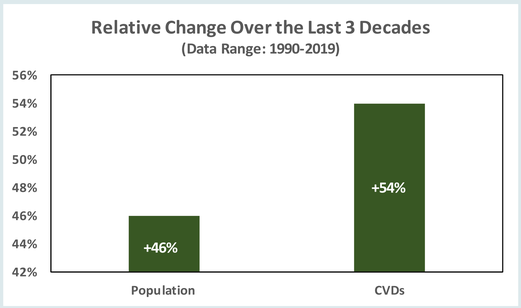
Exploratory Data Analysis is the first step performed on each dataset to understand the data in detail. **Statistical and ML techniques** are used to analyze this data to highlight underlying patterns and build models **(Decision Tree and Clustering)** to develop deeper and actionable insights.

Results (Global Distribution of Cardiovascular Deaths)

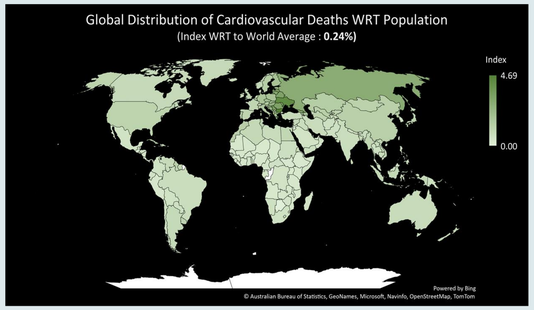
Deaths by top Diseases	1990	2019	Relative Change (2019-1990)/1990
Cardiovascular Diseases	28%	34%	**23.1%**
Neoplasms (Cancer)	13%	19%	**40.2%**
Lower respiratory infections	8%	5%	**-39.9%**
Chronic respiratory diseases	7%	7%	2.8%
Neonatal disorders	7%	3%	**-49.9%**
Diarrheal diseases	7%	3%	**-57.6%**
Digestive diseases	4%	5%	10.4%
Tuberculosis	4%	2%	**-46.9%**
Diabetes mellitus	2%	3%	**87.6%**
Alzheimer	1%	3%	**131.7%**

CVD continues to be the leading cause of deaths worldwide, increasing from 28% in 1990 to 34% in 2019.

Results (Global Distribution of Cardiovascular Deaths)



Relative change in CVD deaths from 1990 to 2019, grew by 54%, outpacing population growth in the same period.



This chart indicates higher index (higher CVD deaths WRT population) for Russia, eastern Europe and Central Asia.

Results (Predicting Heart Risk Using Decision Tree Model & Insights)

Methodology Used for Training and Testing ML Model to predict Heart Failure Risk

Classification Accuracy and Precision

Decision Tree ML Model	True Class	
Predicted	0 (Actual)	1 (Actual)
0	130	36
1	34	157

Precision for Target 1	Accuracy	Recall for Target 1
82%	80%	81%

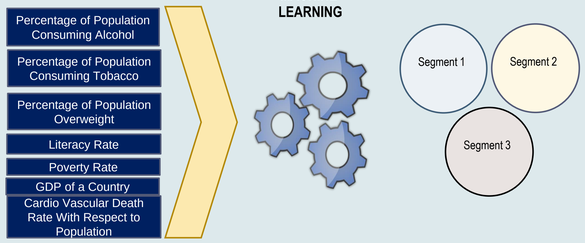
Classification accuracy on testing data set is shown above. Precision and Recall numbers for predicting Heart Risk exhibit relative stability of the model.

Attribute Importance from Strongest to Least link	
ST Slope	Slope refers to the slope of each patient's stress test
Chest Pain Type	This refers to the Angina or Ischemic chest pain
Sex	Sex refers to the assigned sex of the patient at birth
Max Heart Rate	This refers to the maximum heart rate reached during a stress test. This is a hard variable because younger people have higher heart rates while older people have lower heart rates but are more likely to have heart problems
Cholesterol	This refers to cholesterol within the blood
Exercise Angina	Exercise Angina measures whether or not a patient had Angina (severe heart pain) during exercise
Fasting Blood Sugar	Fasting blood sugar measures a person's blood sugar and how it will likely change in the long term
Oldpeak	Oldpeak refers to whether or not the patient showed ST depression induced by exercise during the stress test

ST Slope is most important attribute in predicting heart risk followed by Chest Pain Type, Sex, and Max Heart Rate.

Results (Clustering and Key Segments)

Methodology Used for Generating Segment using ML Clustering Model



Machine Learning Based Cluster Model helped Segment World's Population into 3 Key Distinct Segments

- **High Population Density Regions with Under-Indexed HDI and Moderate CVD Burden (Segment 1)**
- **Populous Regions with Over-Indexed Overweight Population and High CVD Burden (Segment 2)**
- **Developed Regions with Over-Indexed Tobacco and Alcohol Consumption and High CVD (Segment 3)**

These segments are homogeneous within and heterogeneous across and offer great opportunity to develop segment base strategies to tackle Cardiovascular health across the globe.

Conclusion

Key learnings from this study are as follows:

- **ST Slope, Chest Pain Type, and Max Heart Rate** are top drivers for predicting heart risk among patients.
- CVD globally outpaced world population growth **(54% in CVDs vs 46% in population).** Age group of 70+ years population accounts for **majority of CVD**. Russia, Eastern European, and Central Asian countries are over-indexed on CVD.
- Some of the **western countries with developed economies were able to reverse CVD trend. Major populous countries show an increase in CVDs,** posing a significant challenge.
- **Three distinct segments** identified based on aggregated attributes - % of population consuming alcohol and tobacco, % of population overweight, other socioeconomic factors, allowing for **development of segment-based strategies to tackle CVD globally.**
- **% of population using Tobacco** was found to be **positively correlated** with a **higher incidence of CVDs** in a country.
- **Higher HDI** (Human Development Index) found to be **associated** with countries with a **lower incidence and decline in CVDs.**
- Analysis suggests that **improving socio-economic factors could have a positive impact on cardiovascular health outcomes.**

Next Steps

To improve the accuracy of ML model to 95%+ accuracy, efforts should be put in to enhance richness in attributes by gathering patient level information from

- diverse geographical regions on genetic predisposal
- lifestyle and dietary data on longitudinal basis

To decrease the incidents of cardiovascular disease (CVD) deaths, it is crucial to implement more preventive measures on:

Social Factors:

- Develop AI driven wearable devices that could help predict heart risk with very high accuracy to prevent deaths. This will be very helpful for developed nations to prevent deaths due to CVD amongst aged population.
- Measures to incorporate health education into the K-12 curriculum on hazards of sedentary lifestyles, tobacco and alcohol consumption, and unhealthy diets should be undertaken. This can promote students to imbibe healthy behaviors early on and reduce the prevalence of CVD in future.
- Carry out deeper study on policies, programs and preventive measurement practices adopted by the countries that were able to reverse CVD death trends.

Economic Factors:

- WHO should encourage countries to adopt effective policies and practices focused on reducing burden of heart risk and CVD. This will help reduce premature deaths and in turn mitigate huge economic loss.

References

"About." *Our World in Data*, ourworldindata.org/about.

Bloom, D.E., Cafiero, E.T., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L.R., Fathima, S., Feigl, A.B., Gaziano, T., Mowafi, M., Pandya, A., Prettner, K., Rosenberg, L., Seligman, B., Stein, A.Z., & Weinstein, C. (2011). The Global Economic Burden of Noncommunicable Diseases. Geneva: World Economic Forum.

Bradford, Alina. "What Is Blood Sugar?" *livescience.com*, 26 May 2018, www.livescience.com/62673-what-is-blood-sugar.html.

"Cardiovascular Disease Costs Will Exceed $1 Trillion by 2035." *RTI*, 15 Feb. 2017, www.rti.org/news/cardiovascular-disease-costs-will-exceed-1-trillion-2035.

Cardiovascular Diseases | Division of Global Health Protection | Global Health | CDC. www.cdc.gov/globalhealth/healthprotection/ncd/cardiovascular-diseases.html.

Csige, Imre, et al. "The Impact of Obesity on the Cardiovascular System." *Journal of Diabetes Research*, vol. 2018, Hindawi Publishing Corporation, Nov. 2018, pp. 1–12. https://doi.org/10.1155/2018/3407306.