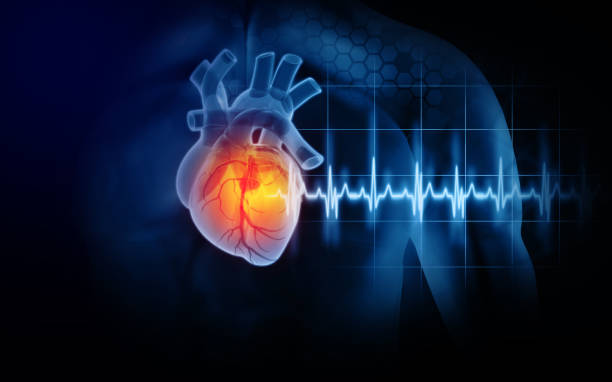
**Exploring Heart Health: A Comprehensive Data Analysis**

****

Cardiovascular diseases have increased significantly since the COVID-19 pandemic. Cardiovascular diseases are responsible for 19.7 million deaths worldwide[1]. In 2020, the annual death toll in India alone amounted to 4.77 million[2].They have spiked by 60% globally in the last 30 years, owing to increased alcoholism, blood pressure, and physical inactivity among the main reasons. Four out of every five CVD deaths occur in middle to low-income countries[3]. CVDs are detected using a number of laboratory tests and imaging studies, encompassing the patient's medical history along with his family's medical history.[4] It also includes factors like age, sex, chest pain type, resting blood pressure, cholesterol levels, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, the number of major vessels coloured by fluoroscopy, thalassemia type, and a target variable related to cardiovascular health.

The following techniques are used in case of Categorical and Continuous variables to verify relation with respect to the target variable:

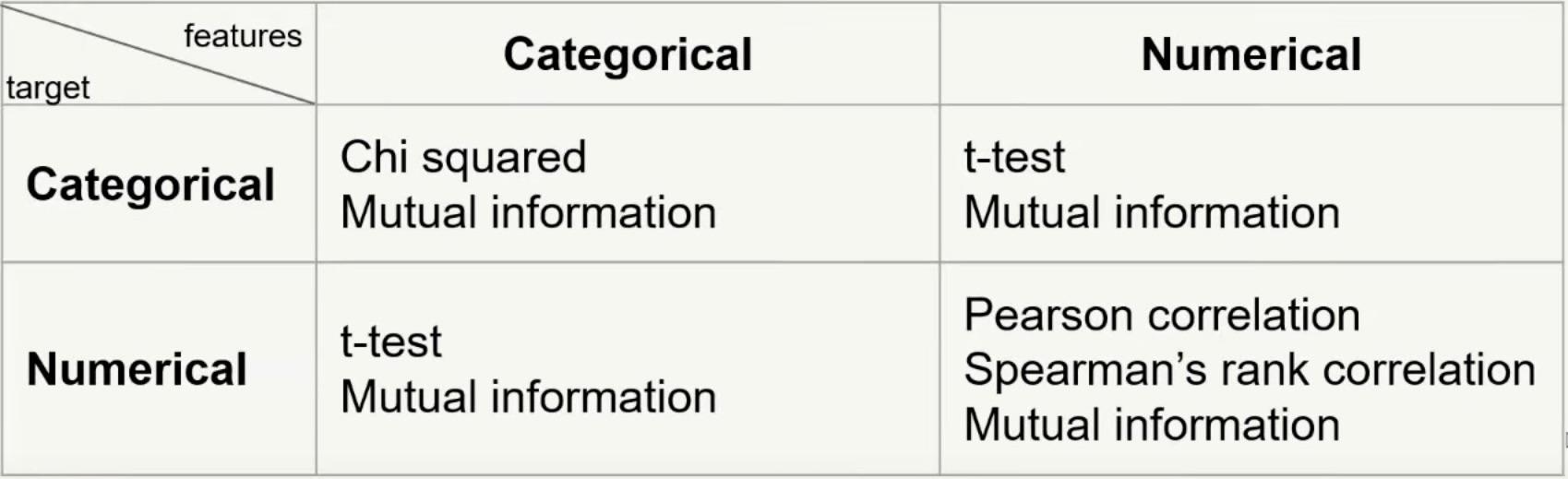


Fig 1: Tests and techniques to check relation between different types of variable

Source: “Feature Selection Techniques” [5]

1. **Relation between age and disease:**

We have divided a continuous variable age into a categorical variable age - group with an age gap of five years. We get a chi2 test value of 29.81 and p - value of 0.00047. The above values suggest a very high correlation between the two categorical variables.

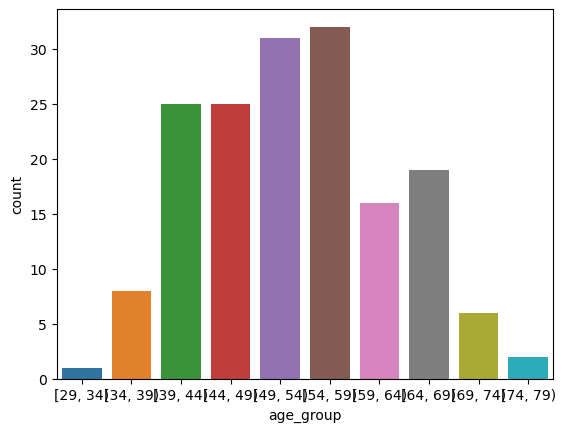


Fig 2: Count plot depicting the distribution of patients with heart disease across different age groups

**Observation: There is a clear spike for ages between 39 to 69 suggesting that ages in these age groups are most prone to heart disease.**

1. **Relation between sex and disease:**

Chi2 square is applied to investigate any relation between sex of a patient and the contraction of disease. Chi2 and p - value was found to be 22.717 and 1.876 × 10-6 respectively indicating a very high correlation. Further, ages and sex were also compared to check if the contraction of disease are at same ages for both men and women.

Additionally, a comparison was conducted between age and sex to examine whether the contraction of the disease occurs at similar ages for both men and women.

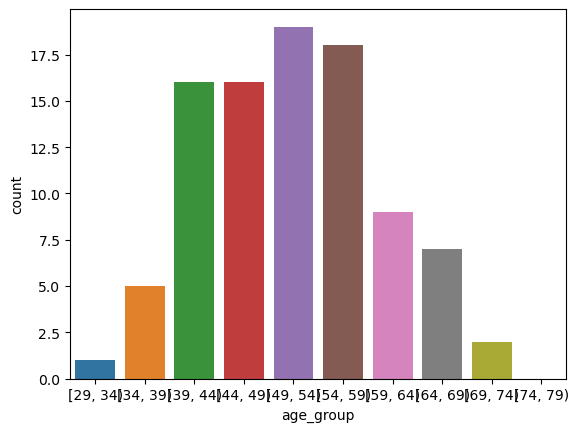


Fig 3: Count plot depicting the distribution of Male patients with heart disease across different age groups

**Observation: There is a spike between the ages 39 and 59 indicating men between these ages are highly prone to the disease.**

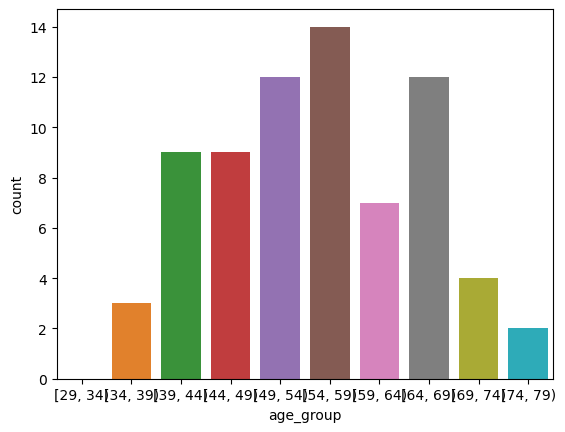


Fig 4: Count plot depicting the distribution of Female patients with heart disease across different age groups

**Observation: In case of women, the proneness to disease persists until the age of 69. The ages at risk for women range from 39 to 69.**

1. **Relation between Resting electrocardiography results and disease:**

Resting electrocardiography results are divided into three categories namely normal, having ST-T wave abnormality and showing probable or definite left ventricular hypertrophy.

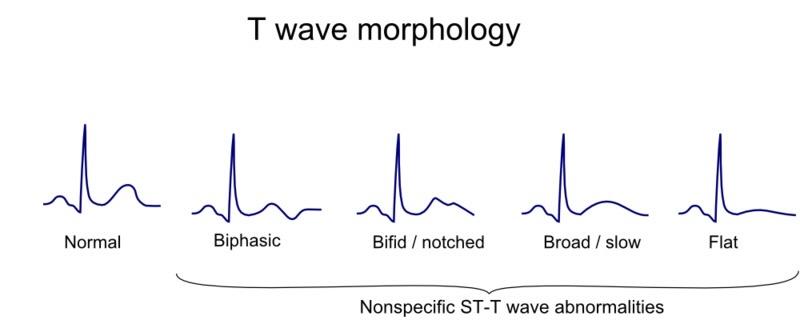


Fig 5: Normal and ST-T wave abnormalities

Source: “Wikidoc” [6]

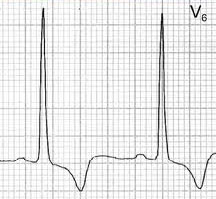


Fig 6: Left ventricular hypertrophy ECG

Source: “Medicine Hack” [7]

There is a strong relation between Rest ECG results and disease as well as Rest ECG and age. Therefore, we will compare the Rest ECG results of each category with respect to different age groups.

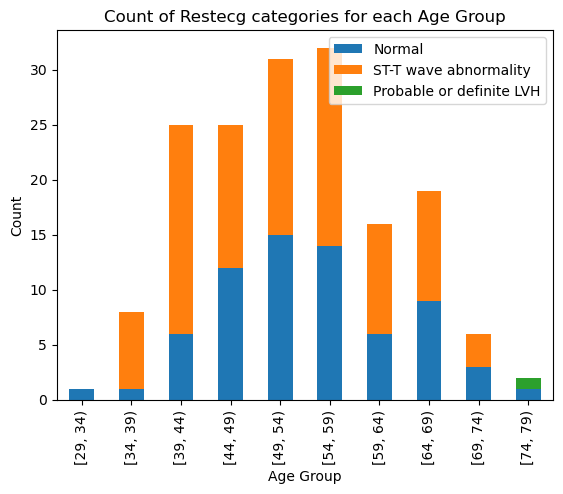


Fig 7: Count plot depicting the distribution of patients with heart disease based on different Rest ECG categories across various age groups

**Observation: It is evident that the likelihood of developing heart disease is higher for ST-T wave abnormality, showing an increasing trend with age.**

1. **Oldpeak (ST depression induced by exercise relative to rest) and disease:**

Oldpeak has a high correlation with disease and also shares a linear relation with age.

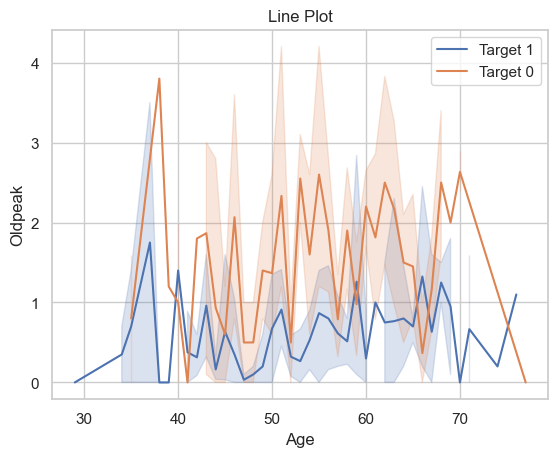


Fig 8: Oldpeak vs Age for patients as well as non-patients

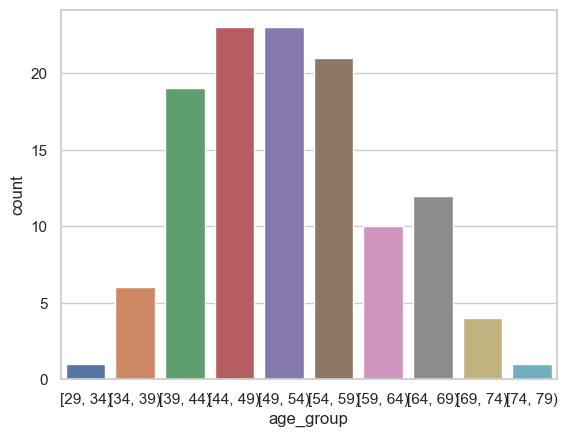


Fig 9: Number of patients when oldpeak<1

**Observation: Oldpeak below 1.0 is dangerous. 73% of patients with oldpeak below 1.0 have heart disease. Age-groups between 39-59 have a higher chance of contraction of disease due to low oldpeak.**

1. **Ca (Number of major vessels (0-3) colored by fluoroscopy) and Disease:**

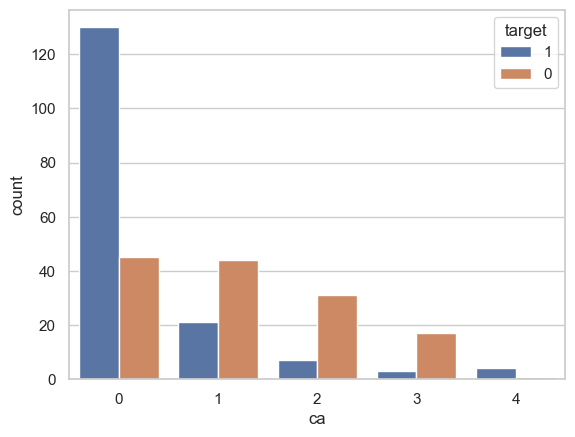


Fig 10: Patients Distribution by Number of Vessels

**Observation:**

**1) number of major vessels = 0 has next most chance of heart disease.**

**2) number of major vessels = 1 has moderate chances of heart disease.**

**3) number of major vessels = 2 has less chance of heart disease.**

**4) number of major vessels = 3 has least chances of heart disease.**

**Lower the number of major vessels, higher the chances of disease.**

REFERENCE:-

[1]: *Cardiovascular diseases*. (2019, June 11). https://www.who.int/health-topics/cardiovascular-diseases

[2]: Incidence of Cardiovascular Risk Factors in an Indian Urban Cohort: Results From the New Delhi Birth Cohort. (2011, April 26). *Journal of the American College of Cardiology*, *57*(17).

[3]: *Deaths from cardiovascular disease surged 60% globally over the last 30 years: Report - World Heart Federation*. (2023, August 9). World Heart Federation. <https://world-heart-federation.org/news/deaths-from-cardiovascular-disease-surged-60-globally-over-the-last-30-years-report/>

[4]: <https://www.news-medical.net/health/Cardiovascular-Disease-Diagnosis.aspx>

[5]: <https://www.youtube.com/watch?v=hCwTDTdYirg&t=1208s>

[6]: *T wave*. (2014, March 11). Wikidoc. <https://www.wikidoc.org/index.php/T_wave>

[7]: *Left ventricular hypertrophy - ECG*. (2011, April 27). https://www.medicinehack.com/2011/04/left-ventricular-hypertrophy-ecg.html