GROUP MEMBERS

Muhammad Talal
CSC-21F-085
Javeriya Naseem
CSC-21F-052

ARTIFICIAL INTELLIGENCE

PROJECT

Data Analytics in Heart Attack Prediction

ABSTRACT

Acute myocardial infarction (heart attack) is one of the deadliest diseases patients face. The key to cardiovascular disease management is to evaluate large scores of datasets, compare and mine for information that can be used to predict, prevent, manage and treat chronic diseases such as heart attacks. Big Data analytics, known in the corporate world for its valuable use in controlling, contrasting and managing large datasets can be applied with much success to the prediction, prevention, management and treatment of cardiovascular disease. Data mining, visualisation and Hadoop are technologies or tools of big data in mining the voluminous datasets for information.
But big data is not only about size, there is also the insight it derives from complex, noisy, heterogeneous, longitudinal, and voluminous data. Challenges, however, include capturing, storing, searching, sharing and analysing. And social communication in varied digital forms is on the increase. Already established as a novel field, data mining is now a primary method for discovering knowledge in buried patterns among the big datasets. In the health care industry, this uncharted knowledge can be utilised in different application domains, for example heart attack prediction. Data mining techniques and machine learning can be used to develop new software to assist providers and others in the health care industry to make decisions about heart attacks in the early stages [17,18].

INTRODUCTION

Acute Myocardial infarction (AMI), commonly referred to as a heart attack, is among one of the deadliest of cardiovascular diseases. AMI happens as circulation or blood flow to heart muscle is interrupted, causing the heart muscle to damage or die (become necrotic).
As the use of Big Data in healthcare increases, more useful personalised medicine will be available to individual patients.The primary reason for most heart attacks is a blockage which causes blood flow to one of the coronary arteries, vital channels through which blood travels to the heart muscle, to become reduced or obstructed. When blood flow is obstructed or reduced, the heart muscle is rapidly deprived of red blood cells which carry the necessary oxygen essential for sustaining life and consciousness in the human body. It takes as few as six to eight minutes without oxygen to cause the heart muscle to arrest, leading to the individual's death.
Big data fosters newer opportunities to predict and/or more rapidly respond to critical clinical events, generating better health outcomes and more efficient cost management. Oracle, an international technology firm, proposes that big data often has low value density if data is received in its original form. Intelligent electronic devices - used by some individuals while both at home and while traveling about their day - now capture and transmit data for analysis in the management of chronic diseases and conditions providing more frequent data about the heart, the breathing process, blood sugar or blood pressure - as the patient goes about daily life - and significantly increases a provider's ability to make appropriate clinical decisions.

DATA PREPROCESSING

The study is a presentation analysis of data on CVD in relation to individual and region-level risks, with a focus on land mining areas. The collection of data from telephone surveys of the aged 29 and above. However the data preprocessing includes several steps such as data cleaning , noise removal, and managing missing lines , data integration, data reduction or dimension reduction and data transformation.However, the large sample size, coupled with comparisons between respondents and non-respondents on key information indicate that it is not a major problem of the non-respondents data.

Data mining techniques can be utilised for pre-processing and machine learning algorithms can be utilised for implementation. Data Mining is a process of extracting valuable and significant knowledge from huge datasets. Data Pre-processing is an important process in data mining and machine learning.

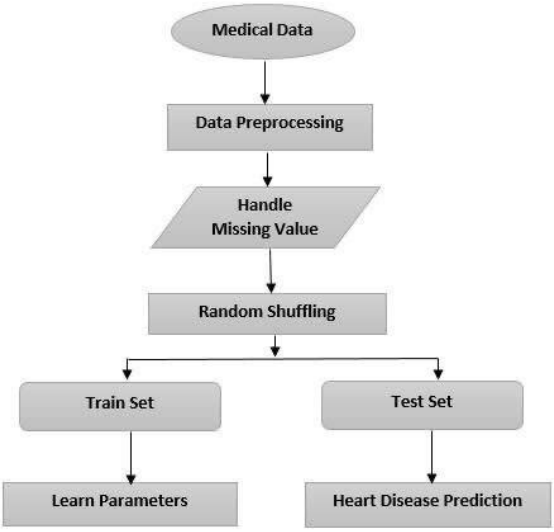
Data mining is a key tool of Big Data used to predict, prevent, and suggest the best treatment plan for heart attack.

METHODOLOGY

Our methodology includes several steps to ensure an accurate analysis of heart attack dataset.. First, we source the dataset from a reliable source that ensures its accuracy and completeness. Next, we perform descriptive statistics to summarise the columns distribution and population profiles in the dataset. We then use inferential statistical techniques such as regression analysis to examine the relationship between the independent variables (eg, age, chest pain) and the dependent variable (CA)..

Big Data varies from traditional decision support tools as it fosters collection and analysis of real-time patient data. It is extremely vital to patient care and reducing both mortality and morbidity associated with heart attacks that providers can utilise big data applications to improve or establish a heart attack prediction program. Predicting heart attacks will not only save numerous lives but will assist providers in establishing personalised medicine, one of the many applications of big data in health care currently available.

Currently, data mining can help health care insurance organizations to predict heart attack. This predictive analysis is widely used in health care. Classification is one of the data mining methods used to predict and classify the predetermined data for the specific class. There are diverse classifications procedures proposed by researchers. Different data mining techniques have been applied to predict heart disease. The accuracy of each algorithm has been verified and stated asNaive bayes , KNN algorithm , and Decision Tree .



LITERATURE REVIEW

The aim of this comprehensive literature review is to examine current standards, methods, and uses for big data to develop a heart attack prediction system to assist providers in establishing higher standards of care and a more personalised medical care plan for the patient.
This literature review was to identify usage of Big Data analytics in heart attack prediction and prevention, the use of technologies applicable to big data, privacy concerns for the patient, and challenges and future trends as well as suggestions for further use of these technologies.
The national and international databases were examined to identify studies conducted about big data analytics in healthcare, heart attack prediction and prevention, technologies used in big data, and privacy concerns. A total of 31 studies that fit these criteria were assessed.
Per the studies analysed, Big Data analytics is useful in predicting heart attack, and the technologies used in Big Data are extremely vital to the management and tailoring of treatment for cardiovascular disease. And as the use of Big Data in healthcare increases, more useful personalised medicine will be available to individual patients.
This review offers the latest information on Big Data analytics in healthcare, predicting heart attack, and tailoring medical treatment to the individual. The results will guide providers, healthcare organizations, nurses, and other treatment providers in using Big Data technologies to predict and manage heart attack as well as what privacy concerns face the use of Big Data analytics in healthcare. Effective and tailored medical treatment can be developed using these technologies.

Literature Review Questions

- What are Big Data Analytics and how is it used in health care?
- Can using Big Data Analytics be used to predict heart attacks?
- What are some of the challenges of using Big Data Analytics in health care?
- What are the future trends associated with Big Data, heart attack
- prediction and personalised medicine?
- Will society punish violations before they occur based solely on

- investigative predictions of future behaviours?

RESULTS

```python
# Initialize classifiers
classifiers = {
    'Logistic Regression': LogisticRegression(),
    'Decision Tree': DecisionTreeClassifier(),
    'Random Forest': RandomForestClassifier(),
    'Support Vector Machine': SVC(),
    'K-Nearest Neighbors': KNeighborsClassifier(),
    'Naive Bayes': GaussianNB()
}

# Training and evaluation
for clf_name, clf in classifiers.items():
    print(f"Training {clf_name}...")
    clf.fit(x_train_scaled, y_train)
    y_pred = clf.predict(x_test_scaled)
    accuracy = accuracy_score(y_test, y_pred)
    print(f"Accuracy of {clf_name}: {accuracy:.2f}")
    print(classification_report(y_test, y_pred))
    print("="*50)
```

Training Logistic Regression...
Accuracy of Logistic Regression: 0.85

	precision	recall	f1-score	support
0	0.92	0.77	0.84	103
1	0.80	0.93	0.86	102
accuracy			0.85	205
macro avg	0.86	0.85	0.85	205
weighted avg	0.86	0.85	0.85	205

==========

Training K-Nearest Neighbors...
Accuracy of K-Nearest Neighbors: 0.90

	precision	recall	f1-score	support
0	0.91	0.88	0.90	103
1	0.89	0.91	0.90	102
accuracy			0.90	205
macro avg	0.90	0.90	0.90	205
weighted avg	0.90	0.90	0.90	205

CONCLUSION

The intersection of big data analytics and healthcare presents a transformative opportunity, particularly in the prediction, prevention, and management of acute myocardial infarction (AMI), or heart attacks. The ability to analyse vast, complex datasets enables healthcare providers to gain valuable insights into cardiovascular disease, leading to more accurate and timely diagnoses, personalized treatment plans, and overall better patient outcomes.

Big data analytics involves several critical processes, including data preprocessing, data mining, and the application of machine learning algorithms. Data preprocessing steps such as cleaning, noise removal, and integration ensure that the datasets are suitable for analysis. Data mining then extracts significant patterns and knowledge, while machine learning algorithms can be employed to predict the likelihood of heart attacks based on various risk factors.