RESEARCH ARTICLE OPEN ACCESS

Data Mining In Medical Analysis

Mrs. Sachithra N S, Mrs. Nissana P S
1(Department of Computer Science, Arafa College of Arts and Science, Pezhakkappilly
Email: nssachithra@gmail.com)
2 (Department of Computer Science, Arafa College of Arts and Science, Pezhakkappilly
Email : nissanaps1@gmail.com)

Abstract:

Data mining has proven to be a powerful tool in medical analysis, enabling healthcare professionals to given actionable insights from vast data sets. This emerging field contributes to accurate disease diagnosis, personalized treatment plans and improved patient outcomes. Data mining technique enable healthcare providers and researchers to identify patterns, trends and associations within large data sets ultimately leading to improved patient care, efficient resource management and breakthrough discoveries.

I. INTRODUCTION

Data mining is the process of discovering patterns and extracting meaningful information from large data sets. When applied to medical analysis, data mining techniques can help healthcare professionals uncover hidden insights, identify trends and make informed decisions that can improve patient outcomes and drive healthcare innovations.

Medical data mining involves analysing various types of healthcare data such as electronic health records (EHRs), medical images, genomic data, clinical trials data and health insurance claims among others. These datasets contains a wealth of information that when effectively analysed, can provide valuable insights for medical professionals.

Data mining techniques can help identify factors that contribute to the occurrence of specific diseases and develop predictive models for early detection. Data mining enabling researchers to identify treatment response patterns and functions that influence patient outcomes, aiding in the development of personalized medicine approaches. Data mining techniques also play a crucial role in pharmacovigilance, which involves monitoring the safety and effectiveness of drugs after they have been approved for use.

II. DATA MINING TECHNIQUES

Data mining techniques play a significance role in the analysis of medical data. In medical analysis, data mining techniques are employed to discover valuable information from various resources, including electronic health records, medical literature, clinical trials and medical imaging. Different techniques are used for data mining in medical analysis.

A. Classification

Classification algorithms are widely used in data mining to predict the probability of certain diseases. These algorithms analyse historical data on patients and their diseases to learn patterns and create models that can be used for prediction. The following are the steps involved in using a classification algorithm for disease prediction.

a) Data Collection

Collect the relevant data of patients and their medical conditions. This data may include patient medical history, laboratory test results, symptoms etc.

b) Data Preprocessing

The collected data is preprocessed to remove any outliers, irrelevant features or missing values. This ensures that the data is noise free and suitable for analysis.

c) Feature Selection

The most relevant features or variables for disease prediction are selected. It involves the significance and influence of each feature on the target variable.

d) Model Training

The selected features and the corresponding disease outcomes are used to train a classification algorithm. Some common classification algorithms used in disease prediction includes logistic regression, decision trees, random forests, support vector machines (SVM) and neural networks.

e) Model Evaluation

The performance is evaluated using evaluation metrics such as accuracy, precision, recall and F1 score. Cross validation techniques may be used for robust evaluation, where the dataset is divided into multiple subsets for training and testing.

f) Model Deployment

Once the model is evaluated and found to have satisfactory performance, it can be deployed to predict the probability of diseases in new, unseen data.

g) Prediction and Interpretation

By using the deployed model new patient data can be inputted to predict the probability of certain diseases. The predictions can be interpreted and used to inform medical decisions, such as recommending preventive measures or further diagnostic tests.

Thus classification algorithms enable the analysis of large scale healthcare datasets to predict the probability of diseases. By identifying patterns and relationships in the data, these algorithms assist in early detection, risk management and personalized treatment planning.

B. Clustering

There are several clustering algorithms commonly used in data mining for medical analysis. Some of them are as follows.

h) K-Means Clustering

This algorithm partitions data into k clusters based on similarity measures. It is commonly used for clustering patients into different groups based on their medical attributes such as age, gender and medical conditions.

i) Hierarchical Clustering

This algorithm creates a hierarchical tree structure by successively merging or dividing clusters based on similarity measures. It is useful for identifying subgroups or clusters within a larger patient population.

j) DBSCAN(Density Based Spatial Clustering of Application with Noise)

This algorithm groups together data points that are closer to each other than a specified threshold, while also identifying outliers. It is often used in medical analysis to identify clusters of patients with similar characteristics or diseases.

k) Fuzzy C-means Clustering

This algorithm assigns membership weights to each data point indicating its degree of belongingness to different clusters. It is useful when the boundaries between clusters are not well defined and patients can belong to multiple clusters based on their medical attributes.

l) Self-organized maps(SOM)

This algorithm uses unsupervised learning to map high dimensional data on to a no dimensional grid, while preserving the relationships between the data points. It is commonly used for visualizing and analysing medical data, such as medical images or gene expression data.

m) Gaussian mixture model(GMM)

This algorithm represents data points as a mixture of Guassian distribution to identify clusters. it is commonly used in medical analysis to identify subtypes or groups of patients based on their medical attributes.

The clustering algorithm can help in various medical analysis tasks such as patient segmentation, disease diagnosis, and treatment planning and outcome prediction. The choice of algorithm depends on the specific requirement and characteristics of the medical data sets being analysed.

C. Association Analysis

Association analysis in data mining is a technique that explores relationship and patterns in datasets to uncover associations or connections between items. The main goal of association analysis is to find interesting or meaningful associations known as item sets, in a dataset. These item sets are sets of items that frequently occur together. The two main metrics used in association analysis are support and confidence. Support means how frequently an item set appears in the dataset, while confidence measures the strength of the relationship between two items in an item set. By combining these metrics, association analysis can determine which item sets are statistically significant and likely to be true associations.

One application of association analysis in medical analysis is in disease diagnosis. By analysing patient data, including symptoms, medical history and test results, and association analysis can identify patterns of symptoms and conditions that frequently co-occur. This can assist in the diagnosis of diseases or conditions that may have similar symptoms or manifestations.

Association analysis can also be used to identify risk factors and predictive markers for certain diseases or conditions. By analysing large medical datasets, researchers can discover associations between certain patient characteristics, such as age, gender, genetic markers, lifestyle factors or exposure to specific risk factors and the occurrence of specific diseases. This information can be valuable for preventive measures, early detection and personalized treatments.

Another application is in drug prescription analysis. By analysing prescription patterns and patient response data, association analysis can identify associations between specific medications and their effects or side effects. This can help healthcare providers make more informed decisions in prescribing medications and understanding potential interactions or adverse events. Association analysis can also be applied in clinical research and healthcare management

III. DATA SOURCES AND PREPROCESSING

In data mining for medical analysis there are various data sources that can be used including:

a) Electronic health records(EHRs)

EHRs contain patient level clinical data, such as diagnosis codes, laboratory results, medication history and demographic information.

b) Clinical trials data

Clinical trials generate data on the safety and efficacy of new treatments or interventions. This data can be used for various analysis, including drug effectiveness and side effect detection.

c) Medical Images

Medical imaging technologies such as X-rays, MRIs and CT scans, generate large volume of image data. These images can be used for diagnostic purposes or to train machine learning models for image recognition.

d) Wearable devices and health sensors

Wearable devices such as fitness trackers and smart watches collect data on various health parameters, such as heart rate, sleep patterns and activity levels. This data can be leveraged for monitoring and predicting health outcomes

e) Genomic Data

Genomic data provides information about an individual's genetic makeup. It can be used for genetic association studies, personalized medicine, and identification of genetic risk factors for diseases.

Pre-processing is an essential step in mining medical data to make it suitable for analysis. Some common pre-processing techniques in medical data mining include:

f) Cleaning and Filtering

This involves removing or correcting incomplete, inconsistent or noisy data. It may also involve removing duplicates or outliers

g) Integration

Medical data often comes from multiple sources with varying formats. Integration involves combining different datasets in to a single format for analysis.

h) Transformation

Transformation techniques are used to convert data into a more suitable format. For example date and time, data may be converted into a standardized format.

i) Discretization

Numerical variables may be discretised in to categories or bins for easier analysis. For example: age could be discretised in to age groups.

j) Feature Selection

Feature selection methods are used to identify the most relevant and informative variables for analysis, reducing dimensionality and improving efficiency.

k) Normalization and Scaling

These techniques are used to ensure that different variables are on the same scale, allowing for fair comparisons.

l) Handling missing values

Missing values in medical data can be common. Techniques like imputation or detection can be used to address missing data.

These pre processing steps help ensure the quality, consistency and suitability of the data for further analysis and mining tasks in medical research and healthcare applications.

IV. APPLICATIONS OF DATA MINING IN MEDICAL ANALYSIS

Data mining has various applications in medical analysis, including:

m) Disease diagnosis and prediction

Data mining techniques can be applied in disease diagnosis and prediction to analyse large amounts of patient data and identify patterns that can help in accurate diagnosis and predicting disease outcomes. Some specific applications include medical image analysis, electronic health records (EHRs) analysis, laboratory test analysis, patient risk stratification and early detection and disease surveillance. By leveraging data mining techniques in disease diagnosis and prediction, healthcare professionals can benefit from improved accuracy in diagnosing diseases, identifying high risk individuals, and predicting disease outcomes. This can lead to more personalized and targeted treatments, early interventions, and better overall health outcomes

n) Treatment effectiveness and personalized medicine

Data mining techniques are increasingly being utilized in the field of medical analysis to improve treatment effectiveness and enable personalized medicine.One key application of data mining in treatment effectiveness is in comparative effectiveness research(CER).CER involves analysing data from various treatment options to determine which ones yield better results for specific patient populations. Data mining can facilitate the development of personalized medicine approaches. Personalized medicine aims to optimize treatment plans based on an individual's unique genetic makeup, lifestyle, and other factors. Additionally data mining can assist in identifying adverse drug reactions and predicting treatment responses.

o) Patient Management and Disease Surveillance

Data mining can be applied in healthcare for patient management and disease surveillance. It can be used to identify high risk patients who are vulnerable to certain diseases or health conditions. Data mining can help in the early detection and diagnosis of diseases. Data mining techniques can be used for disease surveillance to identify and track outbreaks of infectious diseases. Data mining can be applied to analyse treatment outcomes and assess the effectiveness of different treatment approaches and can be used to detect adverse events related to medications, medical devices, or other healthcare interventions.

V. CHALLENGES AND ETHICAL CONSIDERATIONS

a) Ethical ,legal and social issues
b) Volume and complexity of medical data
c) Heterogeneity of medical data
d) Sensitivity of data
e) Poor mathematical characterization of medical data
f) Privacy and security of human data

VI. FUTURE DIRECTIONS

No overview has been finished in India to concentrate on the perspectives on Patients about the utilization of individual information for research. Anyway issues of privacy are probably going to acquire significance with more extensive protection inclusion. The specialist ought to expect this arrangement for what's to come. The ICMR rules permit ERBs to forgo informed assent in suitable situations where the review conveys just negligible gamble or in instances of crisis. Anyway the rules ought to likewise give stipends to assisted Commentator exception from the survey interaction. Study recommendations including clinical records would be incorporated under this class of Survey. The ICMR ought to oppose the transition to universalize the new arrangement of severe rules proposed by the European commission. India really should embrace rules of the functioning gathering .where the ERBs are liable for evaluating the expected significance of examination proposition and choosing whether or not defer the prerequisite for informed assent. Conditions under which ERBs might select to do this incorporate the accompanying circumstances. Admittance to the clinical record is fundamental for consummation of examination and assent isn't pertinent the exploration is probably going to yield data of adequate legitimacy The exploration relates to some future preparation, preventive or remedial drives which might help the patients whose records are examined.

Specialists who are non-clinicians are officially trained about their obligation of privacy and they enrol a clinical boss who officially acknowledges proficient obligation
For any break of classification, would it be a good idea for it happens exorbitant limitations on admittance to clinical information for research could hurt huge number of individuals and hamper in clinical care. An agreement strategy regarding the privileges of people also, obligations of specialists are required in India. Information mining has turned into an essential piece of medical services conveyance, arranging and the board. There have been a large number concentrates on announcing different information mining models and their viability in overseeing colossal clinical data set. These concentrates on overall utilize existing patient data. Anyway there is by all accounts a break in that, the examinations which had admittance to urgent patient data, have not gone through any moral audit process. There is a requirement for the moral thought with brief audit process before any review is attempted inside the domain of information mining research.
Presently how would we give a proof of this?
One basic methodology could be that we do a web search, (not GOOGLE) by and large on writing data sets… all things considered
Pub med or could be anything from IEEE too…
We would look through the web utilizing certain "watchwords" something like "Clinical Information mining"; whatever be the search words kindly note it down (a maximum of 2-3 words can be looked). Likewise note down the URL, date of search.
Also, we can restrict our pursuit to few additional things like Distributions from India
Time span from say between Jan first 2009 to June 30th 2009 (something like a half year.. can make it long term moreover)

Report the quantity of hits you got/the number of were screened to be incorporated, what caused you to prohibit others.
At long last utilizing the hits one acquired, data accessible as to moral survey can be checked out."A web search attempted utilizing the expressions "Clinical information mining" from

India between Jan first 2009 to June 30th 2009; uncovered 400 hits out of which 300 fitted our models in light of the data required. We observed that there was __ % who detailed having gone through moral audit. This recommends that The moral contemplations while rehearsing information mining is especially sabotaged. Steps should be taken to guarantee a compulsory necessity of Moral thought to be important for the information mining research.

VII. CONCLUSIONS

In outline, information mining in medication is unmistakable from that in different fields, on the grounds that the information are heterogeneous; extraordinary moral, lawful, and social imperatives apply to private clinical data; measurable strategies should address these heterogeneity and social issues; and on the grounds that medication itself has a unique status throughout everyday life. Information from clinical sources are voluminous, yet they come from a wide range of sources, not all of comparable construction or quality. The doctor's translations are a fundamental part of this information. They are inadequately portrayed to go with numerical models contrasted with the actual sciences. Medication is far, a long way from the scholarly highest quality level of a standard structure for its essential ideas. The moral, lawful, and social limits on clinical Information mining connect with protection and security contemplations, dread of claims, and the need to adjust the normal advantages of research against any burden or conceivable injury to the patient Strategies for clinical information mining should address the heterogeneity of information sources, information structures, and the inescapability of missing qualities for both specialized and social

ACKNOWLEDGMENT

We extend our heartfelt gratitude to the pioneers in the field of humanoids whose ground breaking research paved the way of our work.

REFERENCES

1. *Doyal Len. Informed consent in medical Research: Journals should not publish Research to which patients have not given Fully informed consent-with three excemptions.BMJ 1997; 314:1107*
2. *Medical Council of India. Indian Medical Council(Professional Conduct,Etiquuette and Ethics) Reulations,2002.Gazette of india dated 06-04-02,part III,Section 4.{cited 2006 March 25*
3. *FLEXChip Signal Processor (MC68175/D), Motorola, 1996."PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.*
4. *www.wikipedia.com*