Comparative Study on Disease Predictions Using Data Mining Techniques

[1]C. Kavitha and [2]Dr. A. Subramani
[1]Research Scholar, Mother Teresa Women's University, Kodaikanal, Dindigul, Tamilnadu, India.
[2]Assistant Professor, Dept. of Computer Science, M. V. Muthiah Govt. Arts, College for Women, Dindigul, Tamilnadu, India.

Abstract: Information mining is the method involved with removing concealed fascinating examples from enormous data set. Be that as it may, information mining with its different scientific devices and methods assumes a significant part in decreasing the utilization of lumbering tests utilized on patients to distinguish a sickness. This paper features the different information digging strategies utilized for anticipating various kinds of illness.

Keywords: *Information Mining, Sickness, Digging, Decision Tree, Support Vector Machine*

I. INRODUCTION

Information mining in medical services gives effectiveness, consistency, and quality[19]. Medical information have the attributes of illness variety, heterogeneity of therapy and result, and the intricacy of gathering, handling, and deciphering data[20]. Medical services associations produce and gather huge voluminous and heterogeneous data everyday and DM assists with revealing a few intriguing examples, which prompts the manual errands end, simple information extraction straightforwardly from records, to save lives, to diminish the expense of clinical benefits and to empower early location of diseases[21].

DATA MINING

Information mining is the method involved with figuring out enormous informational indexes to distinguish examples and connections that can assist with tackling business issues through information examination. Information mining procedures and apparatuses empower endeavors to foresee future patterns and pursue more-educated business choices.

DATA MINING ALGORITHMS AND TECHNIQUES

Different calculations and methods like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method and so on, are utilized for information disclosure from data sets.

DECISION TREE

Decision tree is utilized to construct grouping and relapse models. It is utilized to make information models that will anticipate class marks or values for the dynamic cycle. The models are worked from the preparation dataset took care of to the framework (Supervised learning).

BACKPROPAGATION ALGORITHM

The backpropagation calculation performs learning on a multi-facet feed-forward brain organization. It iteratively learns a bunch of loads for expectation of the class name of tuples. A multi-facet feed-forward brain network comprises of an info layer, at least one secret layers, and a result layer. An illustration of a multi-facet feed-forward network is displayed in underneath Figure
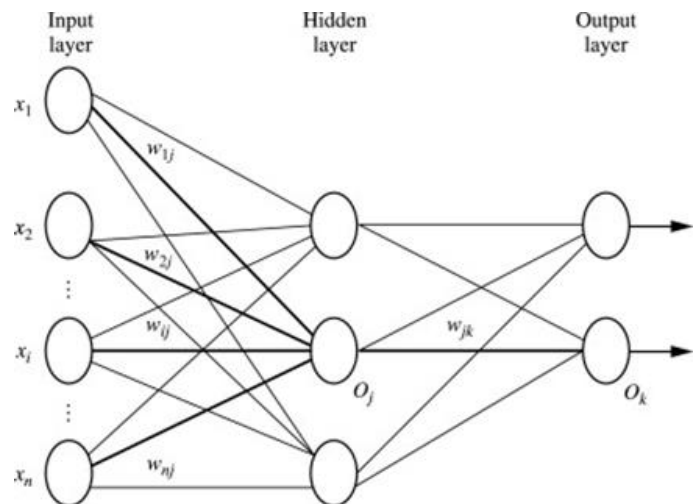

Figure 1.BackProPagation

C4.5 ALGORITHM

The C4. 5 calculation is utilized in Data Mining as a Decision Tree Classifier which can be utilized to create a choice, in light of a specific example of information (univariate or multivariate indicators). In this way, before we jump straight into C4. 5, we should examine a little about Decision Trees and how they can be utilized as classifiers.

SUPPORT VECTOR ALGORITHM

In order to improve the accuracy of breast cancer diagnosis, support vector machine (SVM), a new technology with excellent accuracy compared to existing methods in this classification, is used. The proposed method was evaluated in MATLAB, and the results' validity was evaluated using the UCI dataset, functional parameters, and various statistical criteria.
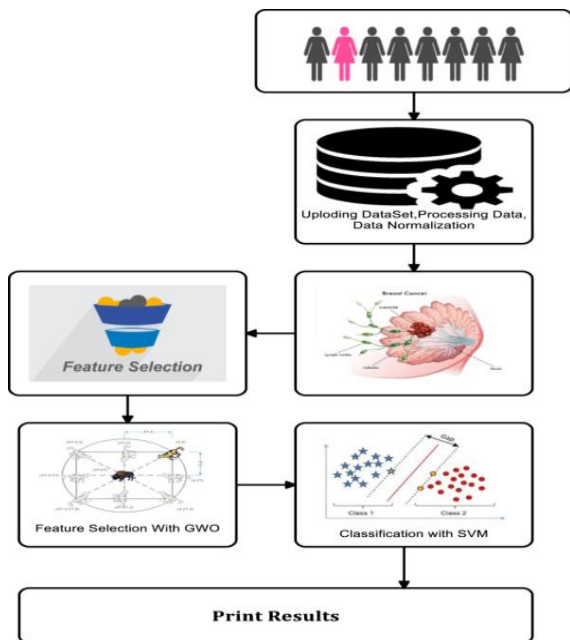
Figure 2

KAWASAKI DISEASE

Kawasaki sickness is an intense febrile vasculitic condition of youth. It is seldom found in grown-ups. Among the grown-up patients with Kawasaki sickness who have been portrayed, a lopsided number are contaminated with human immunodeficiency infection (HIV)[13].

Figure 3. Kawasaki Disease

HEART ATTACK

A coronary failure most generally results from atherosclerosis (greasy developments) in the veins that carry blood to the heart muscle. Plaque development limits the inside of the conduits, making it harder for blood to flow. The measure of harm to the heart muscle depends on the size of the area provided by the hindered vein and the time among injury and treatment. The blocked artery ought to be opened quickly to reduce heart harm.

SWINE FLU

Indications of pig influenza in pigs can incorporate fever, wretchedness, hacking (yapping), release from the nose or eyes, s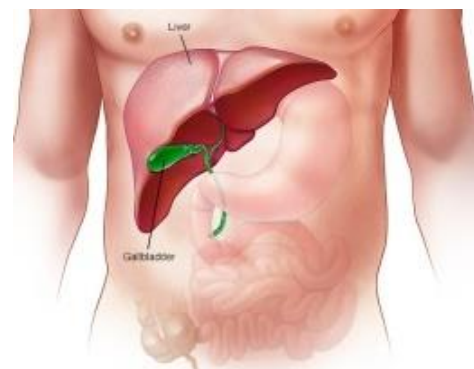niffling, breathing troubles, eye redness or irritation, and going off feed. A few pigs contaminated with flu, in any case, may give no indications of sickness at all[12].

Figure 4. Swine Flu

Pig influenza is an infectious viral disease that is spread from pigs to human. It is additionally called pig flu, pig flu, hoard influenza and pig influenza. At the point when a pig influenza patient hacks or sniffles, the large numbers of minuscule drops emerging from the nose and mouth contain the pig seasonal infection. A sound individual can get H1N1 pig influenza in the event that that individual interacts with these beads or contact surface that tainted individual has as of late contacted. The pig influenza likewise called the H1N1, is another kind of normal flu infection [9]

HEPATITIS C

Long haul disease with the hepatitis C infection is known as persistent hepatitis C. Persistent hepatitis C is normally a "quiet" disease for a long time, until the infection harms the liver enough to cause the signs and side effects of liver sickness. The hepatitis C infection (HCV) spreads through defiled blood[14].

Figure 5.Hepatitis C

BREAST CANCER

Chest sickness starts in your chest tissue. It happens when chest cells change (change) and grow out of control, making a mass of tissue (disease). Like various infections, chest illness can assault and form into the tissue enveloping your chest. It can moreover make an outing to various bits of your body and construction new developments. Right when this happens, it's called metastasis[15].

BRAIN TUMOR

A brain malignant growth is a grouping, or mass, of peculiar cells in your frontal cortex. Your skull, which encases your brain, is incredibly rigid. Any improvement inside such a restricted space can cause problems[16]

ASTHMA

Asthma is a condition wherein your aviation routes restricted and enlarge and may deliver additional bodily fluid. This can make breathing troublesome and trigger hacking, a whistling sound (wheezing) when you inhale out and windedness.
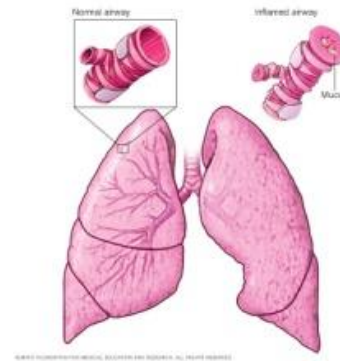

Figure 6.Asthma

SNO	PAPER TITLE	AUTHOR	TYPE OF DISEASE	DATA MINING TOOL	TECHNIQUE / ALGORITHM	ACCURACY LEVEL	DESCRIPTION
1	A Diagnostic Model for Kawasaki Disease Based on Immune Cell Characterization From Blood Samples[1]	Shangming Du	Kawasaki Disease Based on Immune Cell Characterization	LASSO regression	Machine learning algorithm, CIBERSORT	0.80 and 0.77	To estimate the immune cell composition in KD patients and febrile controls (FC), and to develop a tool for KD diagnosis.[1]
2	Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients' Recovery[2]	L. J. Muhammad 1Department of Mathematics and Computer Science	Coronavirus (COVID-19)	Logistic Regression (LR)	Decision Tree (DT) supervised learning algorithms	99.85	recovery of the infected patients from COVID-19 pandemic with the overall accuracy of 99.85% which stands to be the best model developed among the models[2]
3	Hepatitis-C Classification using Data Mining Techniques[3[	Huda Yasin Department of Computer Science, University of Karachi,	Hepatitis-C		Principal Component Analysis , Logistic regression model	90.6%	investigated factors which have higher prevalence of the risk of hepatitis C virus.[3]
4	Method Based on Data Mining Techniques for Breast Cancer Recurrence Analysis[4]	Ying Tan, Peking University, Beijing, China	Breast Cancer	WEKA	Decision Tree, Support Vector Machines, Naïve Bayes, Random Forest Algorithm	73.7 %	the possibility of using intelligent computational tools based on data mining methods for the detection of breast cancer recurrence in patients who had previously undergone surgery[4]
5	Effective heart disease prediction system using data mining techniques[5]	Poonima Singh, L. J. Institute of Engineering and Technology, Gujarat Technological University	Heart Attack	WEKA	multilayer perceptron neural network (MLPNN) with backpropagation (BP)	100%	The experimental results show that the system predicts heart disease with ~100% accuracy by using neural networks.[5]
6	Classification of Brain Tumor Based On Data Mining Techniques[6]	Aniket Sharma, Student, Information Technology, Jayawantrao Sawant College of Engineering, Maharashtra, India	Brain Tumor		Naïve Bayes, K-Means Algorithm	95 to 98%	the K-means clustering algorithm we detect the affected region over an tumor image[6]
7	Asthma Prediction Using Classification Technique[7]	J. CathrinPrincy, Bishop Heber College, Tiruchirapalli, Tamilnadu, India	Asthma		Support Vector Machine (SVM), Multilayer perception (MLP)	98.50%	classification techniques are used to find out the sensitivity, specificity, accuracy, time second of asthma prediction data. High accuracy achieved through SVM technique compare than MLP. SVM achieved 97.50% sensitivity, 98.90% specificity, accuracy 98.50% compare than other algorithms[7]
8	Prediction and Control of Stroke by Data Mining[8]	Leila Amini Department of Computer Engineering and Information Technology, Payam Noor University, Tehran, Iran	stroke	WEKA	C4.5 decision tree and K-nearest neighbor methods	C4.5 decision tree - 95.42% and K-nearest neighbor - 94.18%	C4.5 decision tree algorithm and K-nearest neighbor, can be used in order to predict stroke in high risk groups.[8]
9	Swine flu prediction using data mining techniques: A Review[9]	Dr.Gurmanikkaur,Sant Baba Bhag Singh University, Khiala, Distt: Jalandhar, Punjab, INDIA	Swine flu		Artificial Neural Network (ANN), and Random Forest(RF)	98.6%	It can effectively utilize in predicting the risk factors of swine flu, and to facilitate healthcare providers for the swine flu prediction[9]
10	HYPERTENSION PREDICTION USING DATA MINING[10]	HassanianRaheem Kareem University of Misan College of Education Republic of Iraq	Hypertension		Decision Tree, Time series, clustering and association rules	83%	The purpose was to determine which model gave the highest percentage of correct predictions for diagnosing patients with a hypertension.[10]

Figure 7 -Different Kinds of Disease Using Data Mining Various Techniques /Algorithms

In this above figure described different kinds of disease to be used various data mining techniques and algorithms for various authors to be find out solution described their research paper.

CONCLUSION

In this paper Conclude Various Disease Characterization and then diagnosis Process of all different Diseases using Data mining Techniques for using different tools such as Heart attack, Brain Tumour, Hepatitis, Kawasaki and so on

References

[1] https://www.frontiersin.org/articles/10.3389/fped.2021.769937/full

[2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7306186/

[3] https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.259.4833&rep=rep1&type=pdf.

[4] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7354816/

[5] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5863635/

[6] http://ijariie.com/AdminUploadPdf/Classification_of_Brain_Tumor_Based_On_Data_Mining_Techniques_ijariie8285.pdf

[7] https://serialsjournals.com/abstract/85429_55-176.pdf

[8] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3678226/

[9] Dr.Gurmanikkaur1 ,Er. Arwinderkaur, Swine flu prediction using data mining techniques: A Review, IJCRT | Volume 6, Issue 2 April 2018 | ISSN: 2320-2882.

[10] Hassanian Raheem Kareem, HYPERTENSION PREDICTION USING DATA MINING, 2012, Vol. – I, Issue-III www.srjis.com.

[11] https://www.techtarget.com/searchbusinessanalytics/definition/data-mining

[12] https://www.cdc.gov›flu›swineflu›keyfacts_pigs

[13] Raymond M. Johnson, J. Russell Little, Gregory A. Storch, Kawasaki-Like Syndromes Associated with Human Immunodeficiency Virus Infection, Clinical Infectious Diseases, Volume 32, Issue 11, 1 June 2001, Pages 1628–1634.

[14] https://www.mayoclinic.org/diseases-conditions/hepatitis-c/symptoms-causes/syc-20354278

[15] https://my.clevelandclinic.org/health/diseases/3986-breast-cancer

[16] https://www.healthline.com/health/brain-tumor

[17] https://www.heart.org/en/health-topics/heart-attack/warning-signs-of-a-heart-attack

[18] https://towardsdatascience.com/what-is-the-c4-5-algorithm-and-how-does-it-work-2b971a9e7db0

[19] https://www.jigsawacademy.com/blogs/data-science/data-mining-in-healthcare

[20] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7065247/

[21] https://www.computerscijournal.org/vol9no3/an-introduction-to-data-mining-applied-to-health-oriented-databases/

[22] https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0247-7