Use of Artificial Intelligence & Data Mining & Visualizations to Detect Numerous Illnesses

Sreenivasa Rao Veeranki

Department of Computer Science and Engineering, School of Engg. & Tech., Maharishi university of Information Technology, Lucknow, INDIA

Abstract

Physicians must physically evaluate patients to determine their ailments and diseases and as a result a vast quantity of data is collected in the medical science'domain. Data mining is crucial in illness prediction for a range of factors. For extracting or discovering highly specific data from large amounts of data, Knowledge Discovery method can be used to ease the prognosis of any kind of disease. For extracting completely undiscovered, creative, legitimate, and potentially valuable data patterns, Data mining technique can be a useful technique. Before they occur, healthcare professionals and specialists must be able to foresee a patient's ailment and internal damage. Allowing for better health outcomes. data mining forecasts and categorization aid in the discovery of relationships and trends in patient medical information.

Keywords---Artificial intelligence, Data mining, Visualizations, machine learning, Deep learning, Calcific models, random forest, k-nearest neighbour, support vector machine, Association rule mining, Regression, Naive Bayes Classifier, Artificial Neural Networks, healthcare

Introduction

In healthcare systems, substantial improvements in information developments have led to an enormous accumulation of data. Hospital information, patient information, illness information, and medication costs are all included in healthcare informatics databases. From a variety of methods and sources, this massive amount of data stems. Qualities that are not pertinent and data that is not complete are included in these data. For extracting insights from enormous amounts of illness data, using data mining algorithms can prove to be a beneficial strategy. From a vast volume of illness data, data mining uses a variety of techniques to determine insights. A number of strategies are used in data mining to insights from a large amount of disease data. Data analysis techniques such as categorization, clustering, and rule mining may be used to evaluate data and extract useful information.
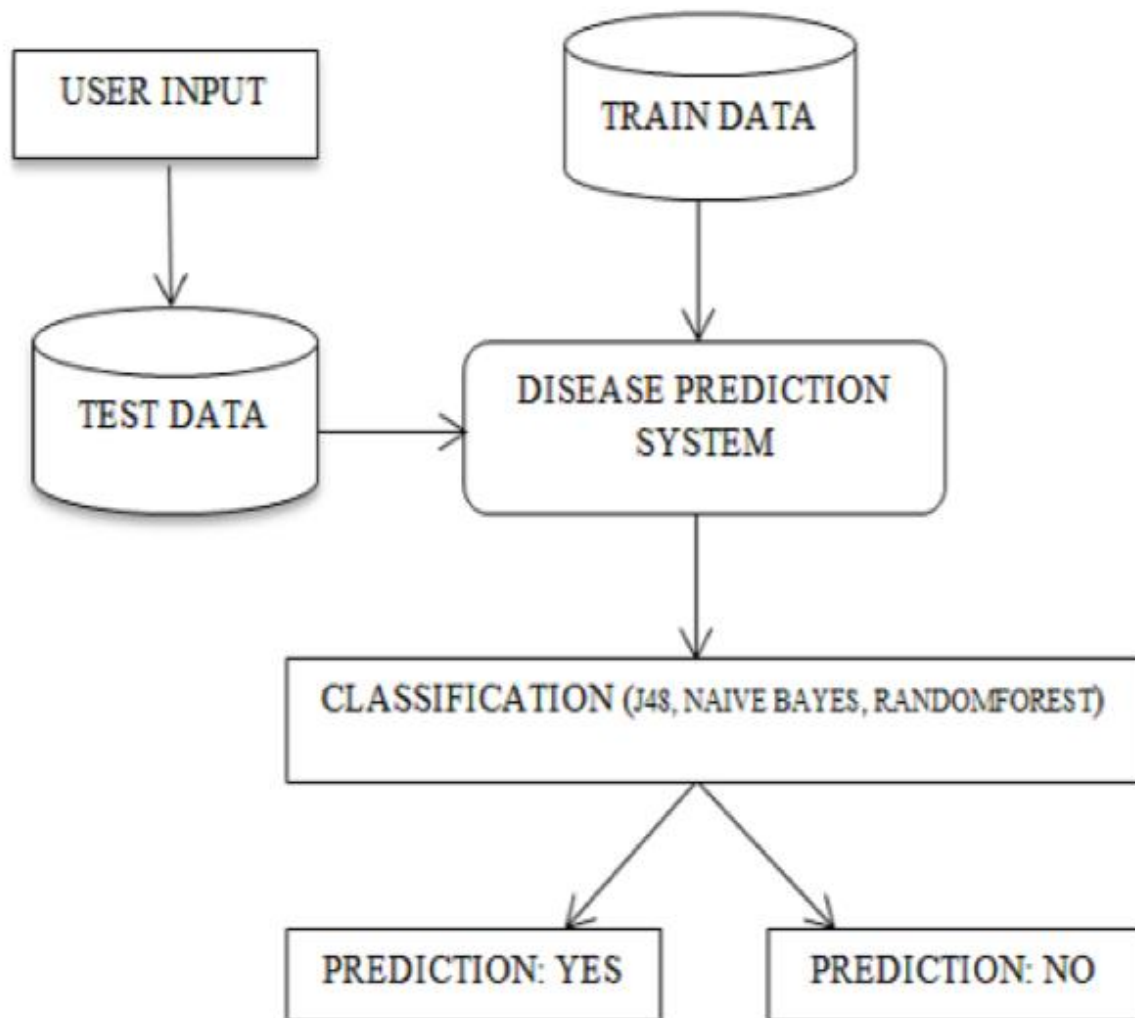
Figure 1: Data Flow Model

(Source: Journal of Information and Computational Science, 2022)

Forecasting the future prognosis of patients depending on previous data accumulated from similar diseases, diagnosing illness depending on medical information, evaluating healthcare expenditures and need for resources, preprocessing commotion, incomplete information, and reducing the time spent waiting for a disease diagnosis are just a few of the essential applications of big data mining in healthcare. Data mining tools like Rapid Miner, Orange, Weka are used to model and understand improved results for patient information

While reducing costs and shortening illness diagnostic time, prospective and updated data mining methods and technologies are applied to enhance healthcare services in the diagnosis of the disease and healthcare analytics. It is crucial to establish a system that will be integrated into the hospital's administration system to assist and advise healthcare personnel in detecting and treating different types of patients with different diseases. In illness detection with varying degrees of success, various data mining algorithms, which include Naive Bayes, Decision Trees, Neural Networks, , Automatically created groups, Bagging Algorithm, SVM and Kernel Density, are employed.

Data Mining Basics

A variety of tests must be performed on the individual to diagnose a condition. The number of trials is necessary to reduce the use of data mining software on the other hand. On productivity and time, this condensed testing data has a significant impact. Since it allows doctors to determine which characteristics, such as gender, bodyweight, and problems, are most important, healthcare data mining is critical for prognosis. This will make it easier for specialists to determine the condition. The practice of uncovering actionable information and themes in the data is known as knowledge discovery in databases. Data mining may be used to understand the meanings of information in databases.

It employs techniques to fetch the data and trends created from the database's information retrieval process. The following are the steps of information extraction in datasets. It gathers information from diverse sources during the Selection step. It eliminates the unnecessary incomplete and ambiguous data in the preprocessing step and provides a dataset that can be converted to a standard format in the transformation step. Data mining methods are used to identify the necessary results after that. During the interpretation stage, it will eventually provide the information to the end-user in a clear way.
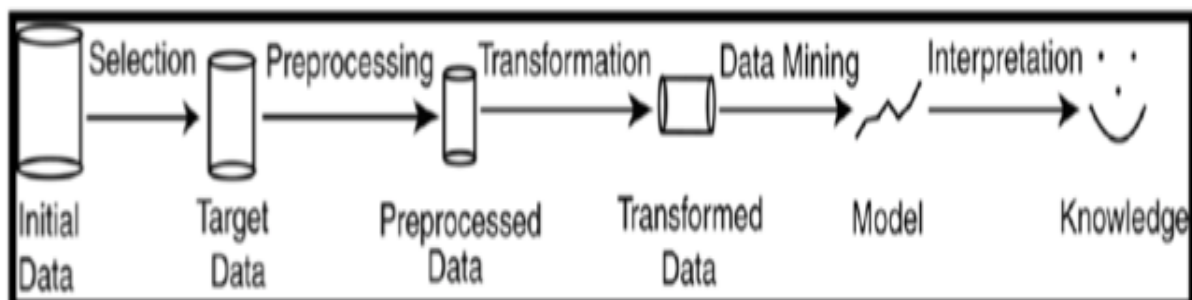

Figure 2: Healthcare knowledge Discovery Process

(Source: ResearchGate, 2017)

Data mining techniques

The commonly employed data mining methods in disease data processing are classification, clustering, regression, and association rule.

Classification- A data mining approach that relies on machine learning is known as classification. Categorising every bit of knowledge in a set of data into one of several specified categories or classes is the main purpose of this process. There are many quantitative approaches that it employs to classify data into distinct

categories. More sophisticated strategies for illness prediction are now available owing to modern categorization techniques. To classify the new unexplored and unobserved data-tuple by reference to the class label, the classifier model is built by executing operations on the training sample and afterward verifying the test statistics.

Clustering- Clustering is a data mining approach that uses an automated process to produce groups of items with similar features. The clustering algorithm generates the categories and places items in them, in which the category is not established. The ability to adapt is a trait of clustering.

Association rule mining- For uncovering intriguing relationships between various data in huge datasets, this is a powerful and well-researched approach. Depending on the input collected data, its goal is to find well-built policies in systems employing a variety of important methods. Identifying the principles, common traits, relationships, correlates, or causal structures within groups of data that may regulate relationships and indirect variables between sets of items is the key function of this data mining process. Finding links and linkages between the various services that individuals look for or ask for to gain a better understanding of their demands and requirements is possible because of this process. To diagnose illnesses, the data mining techniques mentioned can be utilized.

Regression- The outcome of the dependent variable is projected by employing the independent variable and the connection between the two factors is the regression technique of data mining. Analytics and visualization are used to depict the connection quantitatively. Simple linear regression and multiple regression are two different sorts of regression. Simple linear regression employs a single independent variable for forecasting, whereas multiple regression employs numerous variables.

Data mining tools

A large amount of data from diverse diagnostic evaluations and other individual symptoms is generated which is yet another important field of medical science. In data mining, disease prognosis plays a key role. The administration of several tests on the individual is necessary once the illness is discovered. Using data mining approaches, on the other hand, can reduce the number of tests required. In terms of implementation and time, this reduced test set performs a vital function. Since it allows professionals to discover which qualities, such as age, weight, and symptoms, are more important for finding, healthcare data extraction is a vital task. The professionals will be able to evaluate the ailment more efficiently as a result of this. The illness prediction was enhanced with the use of image mining techniques.

WEKA- Waikato University in New Zealand launched the WEKA which is an expansive software and supervised learning toolkit. WEKA can perform sample preparation, grouping, categorization, regression, presentation, and feature selection, among other things. WEKA may also be used to introduce additional methods with conventional methods based on machine learning approaches. From a variety of resources, such as files, URLs, and databases, WEKA allows doctors to retrieve data. It accepts a variety of file types, notably WEKA's custom ARFF format, CSV and many more. A contingency table, accuracy, recollection, genuine affirmative and incorrect negative, and so on are included among other performance standards in WEKA. Source code, device agnostic and mobile, graphical user interface, and a wide variety of diverse data mining techniques are just a few of the benefits of the WEKA tool.

Figure 2: WEKA

(Source: TDS, 2020)

RAPIDMINER- A conducive atmosphere for data mining is created in an open-source platform like RAPIDMINER (RM). It contains a drag and drops feature that is used to build the movement of data. It accepts a variety of file types. Various learning techniques can readily be accomplished, such as regression, classification, and grouping operations. Including a huge variety of classification and regression methods, AR, and clustering methods, Rapid Miner has multiple aspects for data pre-processing, standardization, screening, and evaluation. From a multitude of conventional and standardized datasets, it can import information.

ORANGE- The University of Ljubljana's Bioinformatics Laboratory has created an open-source data mining program called ORANGE. Scripting and development tools can be used to create applications. The Python library may be used to control information and change panels. Arranging widgets on the screen and linking their aspects of the system is how scripting is done. For extraction of information and machine learning techniques, this tool is perfect. For both data analysis doctors and unskilled individuals who wish to test and evaluate their customised methods, it is simple to use. It has the benefit of recycling as much programming as feasible.

KNIME- The Swiss corporation created and maintains a broad function open-source data mining technology known as KNIME (Konstanz Information Miner). Capabilities like information exchange, storage, investigation, and evaluation are the advantages of this tool and it is designed on the Eclipse system. Additional data mining methods, which include R and WEKA, can be combined with KNIME.

Classification Models

Decision tree Induction- A tree composition made up of vertices and connections in which the interior vertices indicate experiments on information properties, the leaf network contains class labels, and the borders or connections that reflect the test result is signified as a decision tree. A decision tree is a technique for categorization and forecasting that reflects a guideline. It is critical to acknowledge and verify the found information produced from the decision model. The vertex of a tree with a leaf node or subdivisions is known as a decision node.
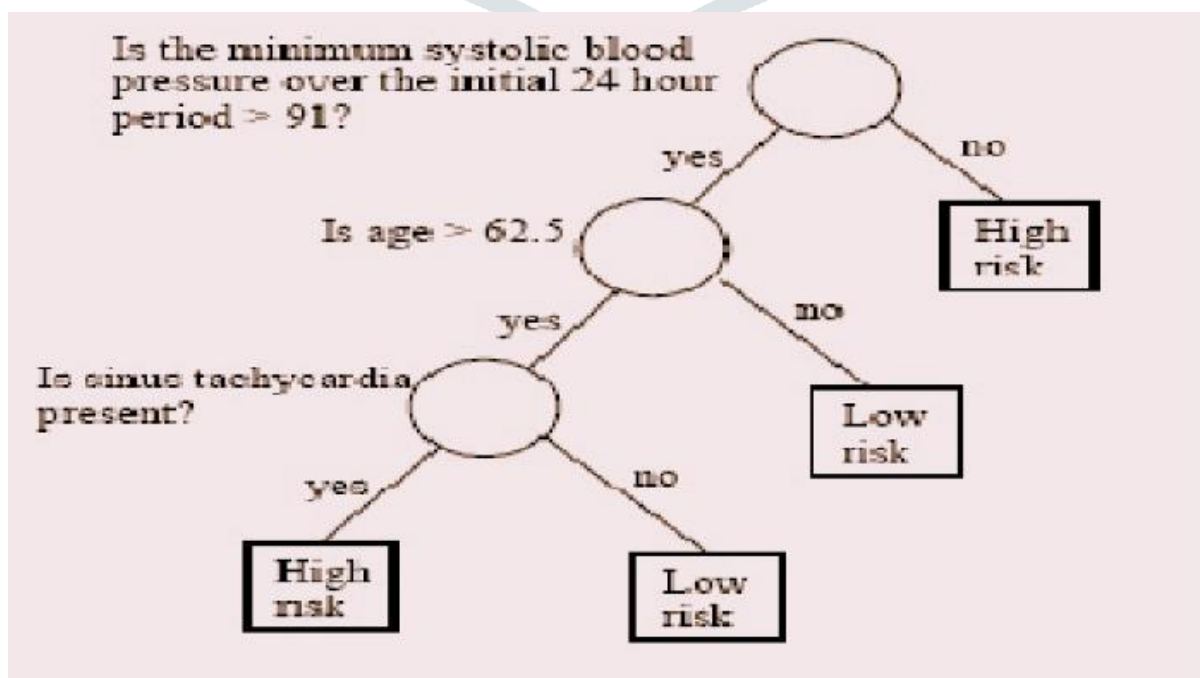
Figure 4: Decision tree

(Source: ResearchGate, 2022)

Naive Bayes Classifier – It is a sort of system in which the likelihood weights of numerous assumptions may be used to forecast their outcomes. Meta-classification is the process of combining the outcome of numerous classifiers by multiplying the average that all classifiers estimate for a particular category. With the emergence of probabilistic reasoning, often known as "Posterior Probability," Naive Bayes classification may be codified. The Naive Bayes classifier predicts the target class for an unknown collection of data samples using statistics and computations.
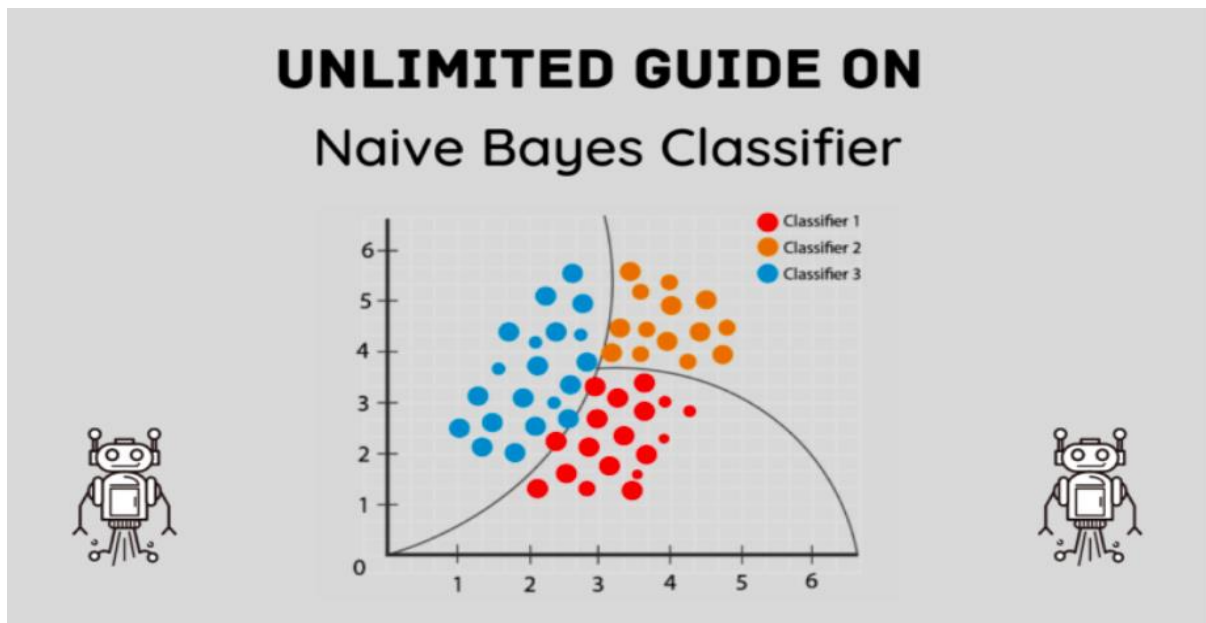
Figure 5: Naive Bayes Classifier

(Source: AnalyticLearn, 2022)

Artificial Neural Network- The most effective learning algorithms are artificial neural networks. They may express multivariate input-output maps with a wide range of advanced functionalities. Inspired by biological data processing, such as the brain, ANN is a data processing framework. By enabling communication, ANN is shown as a network of linked "neurons" that interact. The intrinsic layer, unknown layers, and outside layer are the three main levels of neural networks. A neural network is a collection of linked neurons that can communicate with one another. A strong supervised lesson plan is Artificial Neural Networks. A neural network concept that encompasses a wide range of applications and areas is known as information processing.
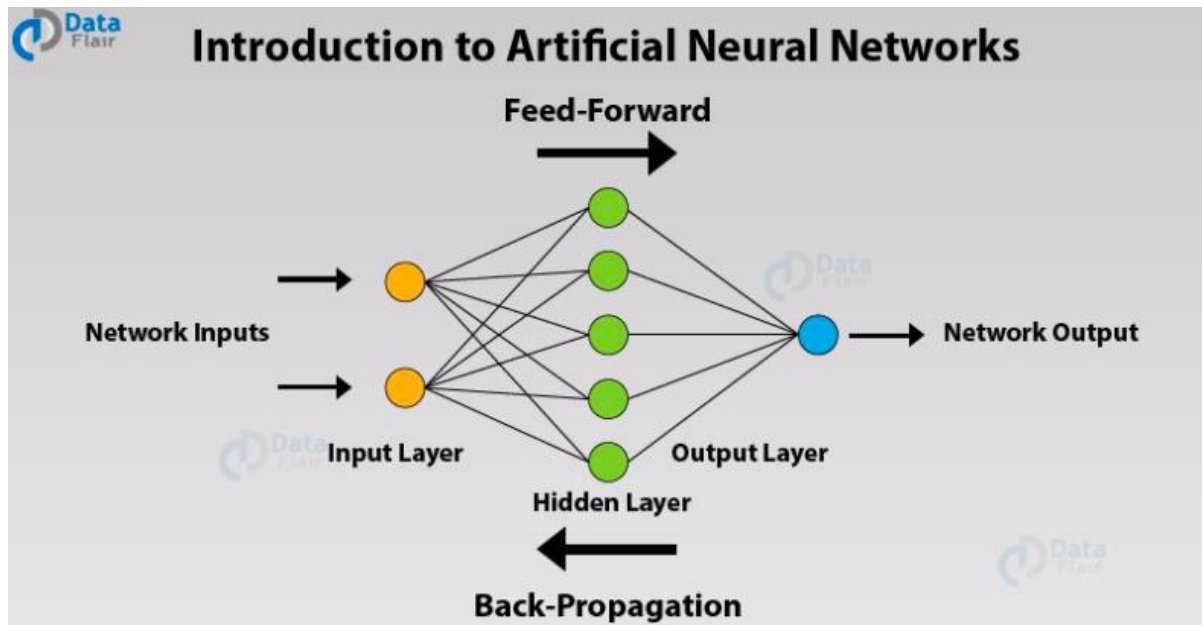
Figure 6: Artificial Neural Network

(Source: DataFlair, 2022)

Data Visualization and its Techniques

In the data set turbulent and inconsistent values might be present in the large data. Such inconsistencies gets support by visualization techniques and in graphical figures and plots the data are represented. User gets satisfaction and convenience to understand the set of data with the help of these techniques (Kirmani, 2017). Between the values of data the relationships and correlations are evaluated for processing the techniques. With the help of statistical formula, techniques of geometry, related to pixels and figures and different purposed methods are calculations are done.

Geometric technique: this data visualization technique is a process where the selected set of data are displayed in different formats of geometry. It includes shapes, and patterns too. Geometric transformation visualization and data projections. Geometric techniques types are: Scatterplot matrices, Hyper-slice and Parallel coordinates.

Graph Based Techniques: The pictorial representation of data is known as Data Visualization. Data visualization is a method which is all about graphs that consists of links and nodes. The easiest type of visualization is Graph Visualization, and it provides and improved and better way for the exploration of data (Singh, Singh & Pandi-Jain, 2018). The different types of Graph Visualization techniques are- 2D/ 3D Graphs, Bar graphs and Line graphs.

Hierarchical Bases Technique- The Visualization technique simultaneously displays various dimensions. However, this technique is not considered appropriate for datasets which are very large in size. Hierarchical visualization technique divides all the dimensions in subspaces. The subspaces are later on visualized in a hierarchical way. Hierarchical technique types are discussed below:

Tree maps: to display structured data in large amounts tree maps are used. Among various rectangles visualization space is divided which are arranged as per the different size and order of quantitative variable. A category is represented by each individual rectangle in the hierarchy level column.

Mosaic plot: in the form of successive decompositions graphical illustrations these plots are represented (Panwar et al., 2020). In this structure rectangles are represented the categorical data count and rectangle at every stage are divided into parallel form.

Data visualization advantages: data visualization applications has several advantages:

First, the exploration of data is easier and can be understood in a simpler way. Second, complex mathematical calculations are not required due to visualization. Third, for knowledge utilization and representation can be done efficiently with the help of visualization.

Result and Analysis

With forecasting and a clearer picture of medical information, data mining in healthcare is a rapidly growing subject that may help. Evaluation of healthcare institutions for improved health regulation and preventative measures of hospital mistakes, early intervention, prevention and treatment of ailments and avoidable hospital lives lost, more reasonable price and cost reductions, and identification of deceptive damage claims are all examples of data mining methods in the healthcare system. In the detection of a variety of ailments, including diabetes, hepatitis, tumor, and cardiovascular disease, data mining methods are being used.

Discussion

The practice of digitally retrieving undiscovered information from massive amounts of data is the main motive of data mining. For illness assessment and treatment, it is critical to extract meaningful facts from massive data sets and provide decision-making outcomes. By evaluating and forecasting different ailments, data mining may be utilised to obtain such information. For uncovering hidden patterns in medical information, healthcare data mining offers a lot of opportunities. With their usefulness varying depending on the healthcare data, there are a

variety of data mining methods accessible. In healthcare, data mining tools have a lot more potential and can be quite useful. In large datasets, it simplifies the task of locating predictive information. In disease prognosis, data mining is very significant.
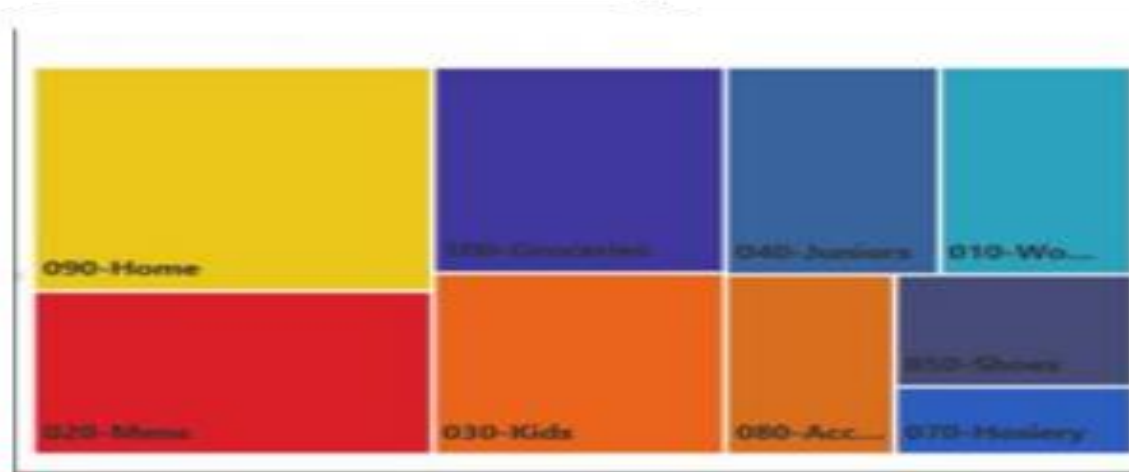

Figure 7: tree map form of data visualization category

(Source: ResearchGate, 2022)

Conclusion

It can be concluded that data mining and visualization can be used efficiently for detecting numerous illness. It has been tremendously used in healthcare sector by professionals so that diseases can be detected easily and in minimal time period treatment can be started to cure patients. These techniques help in taking healthcare sector at another level and get chance to cure various diseases.

References

Azrar, A., Ali, Y., Awais, M., & Zaheer, K. (2018). Data mining models comparison for diabetes prediction. *International Journal of Advanced Computer Science and Applications*, *9*(8).

https://pdfs.semanticscholar.org/3be5/c172b0c99e83cfd7a6e0b01ad89ed7363cfd.pdf

Lan, K., Wang, D. T., Fong, S., Liu, L. S., Wong, K. K., & Dey, N. (2018). A survey of data mining and deep learning in bioinformatics. *Journal of medical systems*, *42*(8), 1-20. https://doi.org/10.1007/s10916-018-1003-9

Learning, M. (2017). Heart disease diagnosis and prediction using machine learning and data mining techniques: a review. *Advances in Computational Sciences and Technology*, *10*(7), 2137-2159. http://www.ripublication.com/acst17/acstv10n7_13.pdf

Gu, D., Li, J., Li, X., & Liang, C. (2017). Visualizing the knowledge structure and evolution of big data research in healthcare informatics. *International journal of medical informatics*, *98*, 22-32. https://doi.org/10.1016/j.ijmedinf.2016.11.006

Babu, S., Vivek, E. M., Famina, K. P., Fida, K., Aswathi, P., Shanid, M., & Hena, M. (2017, April). Heart disease diagnosis using data mining technique. In *2017 international conference of electronics, communication and aerospace technology (ICECA)* (Vol. 1, pp. 750-753). IEEE. 10.1109/ICECA.2017.8203643

Nashif, S., Raihan, M. R., Islam, M. R., & Imam, M. H. (2018). Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system. *World Journal of Engineering and Technology*, 6(4), 854-873. https://doi.org/10.4236/wjet.2018.64057

Verma, A. K., Pal, S., & Kumar, S. (2020). Prediction of skin disease using ensemble data mining techniques and feature selection method—a comparative study. *Applied biochemistry and biotechnology*, *190*(2), 341-359. https://doi.org/10.1007/s12010-019-03093-z Alexander, C. A., & Wang, L. (2017). Big data analytics in heart attack prediction. *J Nurs Care*, *6*(393), 2167-1168. https://www.researchgate.net/profile/Lidong-Wang/publication/316851031_Big_Data_Analytics_in_Heart_Attack_Prediction/links/595fd6c4a6fdccc9b1c47e90/Big-Data-Analytics-in-Heart-Attack-Prediction.pdf

Panwar, H., Gupta, P. K., Siddiqui, M. K., Morales-Menendez, R., Bhardwaj, P., & Singh, V. (2020). A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. *Chaos, Solitons & Fractals*, *140*, 110190. https://doi.org/10.1016/j.chaos.2020.110190

Perez del Palomar, A., Cegonino, J., Montolio, A., Orduna, E., Vilades, E., Sebastian, B., ... & Garcia-Martin, E. (2019). Swept source optical coherence tomography to early detect multiple sclerosis disease. The use of machine learning techniques. *PloS one*, *14*(5), e0216410. https://doi.org/10.1371/journal.pone.0216410

Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. *IEEE access*, *9*, 39707-39716. 10.1109/ACCESS.2021.3064084

Zriqat, I. A., Altamimi, A. M., & Azzeh, M. (2017). A comparative study for predicting heart diseases using data mining classification methods. *arXiv preprint arXiv:1704.02799*. https://doi.org/10.48550/arXiv.1704.02799

Chaurasia, V., & Pal, S. (2017). A novel approach for breast cancer detection using data mining techniques. *International journal of innovative research in computer and communication engineering (An ISO 3297: 2007 Certified Organization) Vol*, 2. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2994932

Singh, P., Singh, S., & Pandi-Jain, G. S. (2018). Effective heart disease prediction system using data mining techniques. *International journal of nanomedicine*, *13*(T-NANO 2014 Abstracts), 121. 10.2147/IJN.S124998

Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics*, *36*, 82-93. https://doi.org/10.1016/j.tele.2018.11.007

Kirmani, M. M. (2017). Cardiovascular disease prediction using data mining techniques: A review. *Oriental Journal of Computer Science & Technology*, *10*(2), 520-528. http://www.computerscijournal.org/pdf/vol10no2/OJCST_Vol10_N2_p_520-528.pdf

Yahaya, L., Oye, N. D., & Garba, E. J. (2020). A comprehensive review on heart disease prediction using data mining and machine learning techniques. *American Journal of Artificial Intelligence*, *4*(1), 20-29. http://article.artifintelli.org/pdf/10.11648.j.ajai.20200401.12.pdf

Journal of Information and Computational Science, (2022). Intelligent Risk Analysis and Prediction of Heart Disease Using Data Mining Techniques. http://www.joics.org/gallery/ics-2658.pdf

ResearchGate, (2017). Analysis of Data Mining Tools for Disease Prediction. https://www.researchgate.net/publication/321520481_Analysis_of_Data_Mining_Tools_for_Disease_Prediction

TDS, (2020). A Great Way for Beginners to Get Started with Machine Learning. https://towardsdatascience.com/a-great-way-for-beginners-to-get-started-with-machine-learning-833cce028620

ResearchGate, (2022). A survey for Data Mining approaches for Healthcare. https://www.researchgate.net/figure/Classification-by-Decision-Tree-Induction_fig2_258285203

AnalyticLearn, (2022). Naive Bayes Classifier: Unlimited Guide On Naive Bayes. https://analyticslearn.com/naive-bayes-classifier-unlimited-guide-on-naive-bayes

DataFlair, (2022). Artificial Neural Networks for Machine Learning – Every aspect you need to know about. https://data-flair.training/blogs/artificial-neural-networks-for-machine-learning/

Komi, M., Li, J., Zhai, Y., & Zhang, X. (2017, June). Application of data mining methods in diabetes prediction. In *2017 2nd international conference on image, vision and computing (ICIVC)* (pp. 1006-1010). IEEE. https://doi.org/10.1109/ICIVC.2017.7984706

Verma, D., & Mishra, N. (2017, December). Analysis and prediction of breast cancer and diabetes disease datasets using data mining classification techniques. In *2017 International Conference on Intelligent Sustainable Systems (ICISS)* (pp. 533-538). IEEE. https://doi.org/10.1109/ISS1.2017.8389229

Dahiwade, D., Patle, G., & Meshram, E. (2019, March). Designing disease prediction model using machine learning approach. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1211-1215). IEEE. 10.1109/ICCMC.2019.8819782

Sreenivasa Rao Veeranki." **ARTIFICIAL INTELLIGENCE FACILITATED IN BUSINESS INTELLIGENCE",International Journal of Management Technology and Engineering,**http://ijamtes.org/VOL-10-ISSUE-3-2020/, Sreenivasa Rao Veeranki - Maharishi university of Information Technology, Lucknow, INDIA.Page No : 446-456 , DOI:16.10089.IJMTE.2020.V10I03.20.3587,ISSN NO: 2249-7455

Sreenivasa Rao Veeranki ."**ARTIFICIAL INTELLIGENCE IN HANDLING NATURAL SCIENCES". ",International Journal of Management Technology and Engineering,**http://ijamtes.org/VOL-10-ISSUE-4-2020/, Sreenivasa Rao Veeranki - Maharishi university of Information Technology, Lucknow, INDIA. Volume X, Issue IV, APRIL – 2020,Page No : 518-528,DOI:16.10089.IJMTE.2020.V10I04.20.3665, ISSN NO: 2249-7455

Sreenivasa Rao Veeranki, Manish Varshney, **"Role of Bioinformatics In Biotechnology Concerning AI",** IJRAR - International Journal of Research and Analytical Reviews (IJRAR), E-ISSN 2348-1269, P- ISSN 2349-5138, Volume.6, Issue 2, Page No pp.8-17, June 2019, Available at : http://www.ijrar.org/IJRAR19K9397.pdf

Sreenivasa Rao Veeranki, Manish Varshney,**"Bioinformatics and Data Science in Medical Research",** International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.6, Issue 1, page no.204-214, January-2019, Available :http://www.jetir.org/papers/JETIR1901G29.pdf