EasyChair Preprint

№ 15184

Healthcare Databases Analysing Using Predictive Data Analysis Approach

Lucky Adhikari, Sudipta Hazra, Senjuti Banerjee,
Siddhartha Chatterjee and Aaliya Nashat

October 1, 2024

Healthcare Databases Analysing Using Predictive Data Analysis Approach

Lucky Adhikari[1], Sudipta Hazra[2], Senjuti Banerjee[3], Siddhartha Chatterjee[4] and Aaliya Nashat[5]

[1,3,5]Hospital Management, Brainware University, India

[1]luckyadhikary97@gmail.com

[2]Computer Science and Engineering, Asansol Engineering College, India

[2]sudiptahazra.nitdgp@gmail.com

[4]Computer Science and Engineering, College of Engineering and Management Kolaghat, India

[4]siddhartha.chatterjee31@gmail.com

Abstract: ***The healthcare databases are coagulating at a poriferous rate, making it difficult to evaluate the many parameters and find hidden patterns for further knowledge development. Over the past ten years, data analytics has demonstrated a wide range of success in collecting or extracting information from central databases. It has raised awareness among scientists and researchers all around the world to develop technologies for knowledge discovery in massive datasets. In fact, with the advent of cutting-edge IT-based technologies for medical diagnostics, healthcare databases have experienced exponential growth in volume and dimensionality. According to medical databases, congenital heart disease ranks among the leading causes of new born fatalities in both industrialised and developing nations. Patients with genetic abnormalities are living longer thanks to potential cures for congenital disabilities that have surfaced as a result of recent advancements in the healthcare industry. As a result, the affected individuals can now live longer. The current study used a clustering technique to locate hidden ways to find hidden patterns from congenital cardiac databases for future medical diagnosis. The designs were examined between 2006 and 2016 in India based on the death rate from congenital cardiac abnormalities in different age groups.***

Keywords: Healthcare Databases, Predictive Data Analysis, Machine Learning, Data Mining, Healthcare Analytics.

1. INTRODUCTION

Healthcare databases are coagulating at a poriferous rate, making it difficult to evaluate the many metrics and find hidden patterns for further knowledge discovery. Over the past ten years, data analytics has demonstrated a wide range of success in collecting or extracting information from central databases. This has raised awareness among scientists and academics all across the world on the need to develop technology for large-scale database knowledge discovery. In fact, with the advent of cutting-edge IT-based technologies for medical diagnostics, healthcare databases have experienced exponential growth in volume and dimensionality. The exponential growth of healthcare technology has produced detection patterns for upcoming medical diagnoses. As a result, scientists and researchers are devoting their greatest efforts to uncovering secret data and a range of variables that may be identified as the illness's underlying cause. Healthcare databases are generally quite complex and constantly expanding, making it difficult to handle data anomalies.

The Centres for Disease Control and Prevention (CDC) estimate that approximately 40,000 new-borns in the US are born with COHD each year. Between 1999 and 2006, there were approximately 41,494 recorded deaths in the US as a result of congenital cardiac abnormalities. In the US, congenital heart disease affects about 1 in 150 persons, according to the AHA. Ventricular septal defects, atrial septal defects, pulmonary stenosis, and patent ductus arteriosus are the most common types of congenital heart disease (CHD) seen in children, although tetralogy of fallot and atrial septal defects are more common in adults [1]. One of the main causes of new born fatalities worldwide is thought to be congenital heart disease. According to this study, one of the most prevalent congenital anomalies for which it can be difficult to anticipate a disease's course before birth is congenital heart defects (CHD). The severity of congenital heart defects (CHD) can range from a little hole between the chambers or the heart's valves, which can be immediately fixed, to major abnormalities that may need multiple surgeries in infancy, youth, or adulthood [2]. However, managing the complexity of real-time healthcare information presents a worldwide problem for academics. Numerous diverse optimisation methods and instruments are created to assess the precision and expertise derived from large datasets. However, because of the exponential velocity of the data stream and its complicated character, which can inevitably limit knowledge discovery, making decisions from large databases is a simple challenge. Forfeited, data mining and knowledge discovery are used as a difficult task to find relevant information for making decisions.

Today's health sector has access to private data that may be crucial to decision-making [3]. However, because of its complexity, patterns become hard to find using conventional computing technologies. Therefore, medical data mining is a suitable technique that has a great ability to uncover hidden patterns in the data sets related to the medical domain in order to overcome the complexity of the data. Clinical diagnosis can be made with the help of these patterns. By diagnosing and drawing connections between seemingly unrelated data, data mining can help healthcare organisations anticipate trends in the health and behaviour of their patients [4]. Nevertheless, there are currently fewer tools available for assessing and analysing large amounts of clinical data after they have been saved [5]. One sort of knowledge discovery technique is data mining, which extracts unknown data relationships and arrangements. Innovative research methodologies have been made possible by the most recent advancements in the medical profession. Researchers are now able to handle and analyse complex analyst data because to the increased expansion of data mining. As a result, the specialists may now review a large collection of data [6–9]. Gathering naïve data from the relevant bodies is more expensive and more difficult to handle than using the data that is already in the databases [10]. Numerous databases are created expressly to give individuals worldwide a better understanding of congenital cardiac defects.

The goal of the current study is to forecast a disease's trend using the data sets that patients have submitted [10]. In the healthcare sector, the two most important objectives of data mining are patterns and predictions. Data mining is a novel approach to discovery that involves several processes, and it requires input from the specialist field at every stage of the process [11]. Classification, clustering, and rule mining are some of the data mining techniques that are helpful for deciphering data and producing insightful information. Applications of data mining in the healthcare industry include predicting the effects of diseases in the future based on historical data, diagnosing illnesses, and analysing treatment costs [4, 12]. The goal is to employ data analytics to identify trends in the massive COHD databases. The information was gathered from the global health data exchange repository, which oversees the management of health data worldwide. The several methods are retrieved according to the mortality among congenital heart abnormalities for different age groups between 2006 and 2016. The data is divided into mixed age groups, ranging from early infancy to 64 years of age. Furthermore, a broad spectrum of medical knowledge can be shared between beneficiaries and professionals in the health industries thanks to computer applications in the healthcare sector. The standard collection and integration of data from multiple internal databases is what makes up data mining tools. Databases, dataset extraction from those databases, and data sample prior to algorithm application are all part of the knowledge discovery process [13].

2. DATA MINING FOR HEART DEFECTS AT BIRTH

These days, hospitals handle patient data with computational applications. Patient information, treatment costs, and other pertinent information are typically included in hospital statistics. Generally, these statistics are periodically updated whether they are presented in numerical, diagrammatical, or graphical form. Nonetheless, the majority of hospital data are intricate and challenging to handle [14]. One sort of knowledge discovery technique is data mining, which extracts unknown data relationships and arrangements. The most recent advancements in medicine have led to innovative research methodologies. Researchers are now able to handle and analyse complex data because to data mining's enhanced development. As a result, the now examines a large amount of data [15]. Gathering Naïve data from the relevant bodies is more expensive and more difficult to maintain than the data offered on the databases. Numerous databases are specifically created to give individuals worldwide a better understanding of congenital cardiac diseases. The goal of the current study is to forecast a disease's pattern using the patient data set that has been provided. In the healthcare sector, the two most crucial objectives of data mining are patterns and predictions. Data mining is a multi-step, unique strategy that requires input from the specialist area at every stage of the discovery process [16]. Classification, clustering, and rule mining are a few examples of data mining techniques that are helpful for deciphering data and producing actionable insights. Applications of data mining in the healthcare industry include predicting the effects of diseases in the future based on historical data, diagnosing illnesses, analysing treatment costs, and other aspects [17]. Furthermore, a vast range of medical knowledge can be shared between beneficiaries and professionals in the health industries thanks to computer applications in the healthcare sector. The standard collection and integration of data from several internal databases is known as data mining tools. Databases are used in the knowledge discovery process, datasets are extracted from such databases, and data is sampled before any algorithms are applied. Combining machine learning and statistical congenital heart defects with data mining allows extracting hidden correlations from massive databases. The purpose of each data mining technique varies based on the modelling goal. The following are the two essential modelling goals: prediction and classification. Prediction models provide information about the continuously valued function, while classification methods provide categorical labels [18].

An investigation of the differences in mortality rates among adults with COHD based on gender was conducted. The Quebec database of congenital cardiac abnormalities for people aged 18 to 65 who were hospitalised to hospitals between 1996 and 2005 was used to gather data. Regression analysis indicates that men have higher testosterone than women [19]. An additional investigation into the risk variables influencing the mortality of COHD newborns, children, and adults following heart transplantation. 488 cases of heart transplants for congenital cardiac defects were included in the study. Prior to transplantation, 93% of cases underwent at least one procedure. After three months, there was a notable difference in the survival rates of children with cardiomyopathy and COHD patients. If COHD patients can make it through the post-operative phase, they will likely have good late survival [20]. If we investigate the cause of the fluctuations in COHD incidence. The incidence of COHD varies by state, ranging from 4 in 1000 to 50 in 1000 live newborns. Variable frequency drive was connected to the incident (VSD). Out of the 62 reports that were analysed, the incidence was looked at after 1955. For moderate and critical COHD cases, the incidence rate was 6 per 1000 live births [11]. To determine the prevalence rates of COHD in infants, kids, and adults, Bayesian models are also utilised. Information on patients with congenital heart defects (CHD) was available in the Quebec database from 1983 to 2010. From 1998 to 2005, the prevalence rate among newborns was 8.21 per 1000 live births. Starting in 2010, the total frequency rose to 13.11 per 1000 in children and 6.12 per 1000 in adults during the course of the following five years. Thus, between 2000 and 2010, the prevalence rate increased by 11% and 57% in children and adults, respectively. Critical COHD cases, on the other hand, have a significantly higher rate, rising by 19% in children and 55% in adults. Adults accounted for 66% of all COHD cases by 2010 [21]. Between 1985 and 2000, the prevalence rate and age distribution of adults and children with major and mild congenital cardiac defects are also examined. Tetralogy of Fallot, arteriosus, TGA, and abnormalities of the univentricular heart are the identified kinds of severe congenital heart disease. In 1995, 2000, 1990, and 1995, the prevalence rate of critical and minor COHD cases was examined. The information that was needed was gathered from Quebec. Children under the age of eighteen were covered in the study. In 2000, there were 4.09 cases of COHD for every 1000 adults and 0.38 cases of catastrophic COHD for every 1000 individuals. Of the two sexes combined, women made for 57% of all adult COHD patients. In 1985, the median age of all individuals with critical COHD was 11 years old; by 2000, that age had increased to 17 years old. Between 1985 and 2000, the prevalence rate of significant COHD increased. The increase in the COHD rate in adults was found to be substantially greater than in children. Therefore, the total significant COHD population in 2000 consisted of 49% of adults [22].

3. METHODOLOGIES

The healthcare industry generates vast amounts of data from various sources, including electronic health records (EHRs), laboratory results, imaging studies, and administrative databases. Predictive data analysis involves using statistical techniques and machine learning algorithms to analyze historical data and make predictions about future events.

3.1 DATA COLLECTION AND INTEGRATION

Effective predictive data analysis requires comprehensive and high-quality data from diverse sources. Key steps include:

Data Integration: Combining data from various sources such as EHRs, laboratory systems, and imaging databases to create a unified dataset.

Data Cleaning: Addressing missing values, correcting errors, and standardizing data formats to ensure data quality.

Feature Engineering: Extracting relevant features and creating new variables that capture important aspects of the data.

3.2 ML TECHNIQUES

Several machine learning techniques are employed in predictive data analysis:

Supervised Learning: Algorithms like logistic regression, decision trees, and neural networks are trained on labeled data to predict outcomes such as disease diagnosis or patient readmission [23].

Unsupervised Learning: Techniques like clustering and dimensionality reduction are used to identify patterns and group similar patients without predefined labels [24].

Ensemble Methods: Combining multiple models to improve prediction accuracy and robustness.

Supervised and unsupervised learning are present in figure. 1 and 2 respectively.

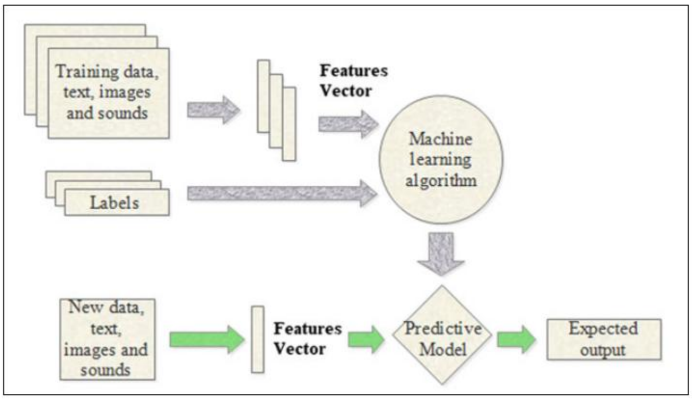

Fig.1. Supervised learning

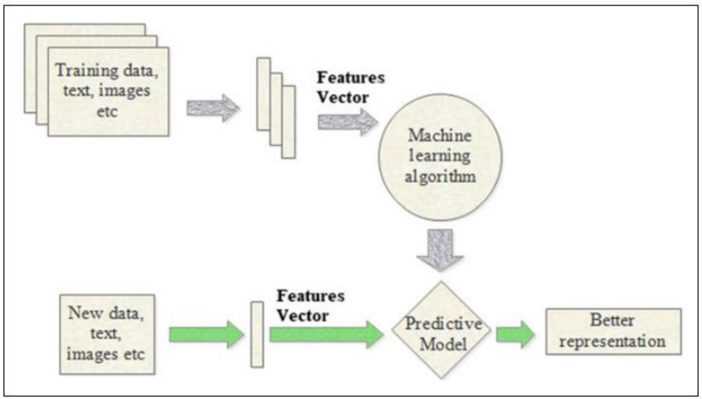

Fig.2. Unsupervised learning

3.3 PREDICTIVE MODELING

Building predictive models involves:

Model Selection: Choosing appropriate algorithms based on the nature of the problem and the data.

Training and Validation: Splitting the data into training and validation sets to develop and test the model.

Evaluation Metrics: Using metrics such as accuracy, precision, recall, F1 score, and area under the ROC curve (AUC-ROC) to evaluate model performance.

3.4 AN APPROACH TO DATA ANALYTICS WITH HIERARCHICAL CLUSTERING

Finding information and hidden patterns in massive databases is the goal of the data analytics technique, which will be used in future prediction modelling. The main goal of the research is to use unsupervised techniques to extract knowledge from large-scale datasets. Unsupervised learning approaches often rely on no class labels and estimate the degree of similarity between the data using a variety of statistical techniques [26-30]. While clustering approaches are utilised to determine the degree of similarity between the data, each algorithm is unique due to the choice of algorithms, parameters, and other factors. The partitioning, hierarchical, density, and grid-based clustering approaches are the most commonly discussed clustering algorithms. The contribution of our work centres on hierarchical clustering, which homogenises data by classifying it according to related qualities. In each phase of the clustering process, a new cluster is formed or an existing cluster group is joined. The fundamental clustering procedures are crucial because the previously created collection is compared to the other groups. Nonetheless, there are two methods for measuring Hierarchical Clustering techniques: the divisive method or the agglomerative congenital heart defects approach.

In order to ensure that the primary number of collections equals the total number of groups, the agglomerative approach divides each group into a distinct cluster only during the initial steps. Subsequent stages involve merging the discovered clusters until each group is categorised into a single collection. Conversely, the dividing approach has the opposite effect. With this approach, each group first forms a single cluster before splitting off into distinct groups. The initial stage in clustering is to measure the separation between the groupings. The groups are increasingly similar the closer they are together. There are different distance matrices available, and the selection is based on whether the data values are discrete or continuous. The results are displayed using a hierarchical cluster approach in the form of a dendrogram. On the dendrogram, the groups that are most similar to one another are first grouped together. The classes of clusters according to the agglomerative method are depicted by the vertical lines. The distance between linked groups is also specified in these lines. When the length increases and the similarity starts to decline, these lines further move. The disparity in the clusters' respective distances is depicted by the horizontal lines. These lines serve as a link between every set in a group, which is crucial for examining the final collections. As a result, the largest horizontal line displays the significant variations.

4. RESULTS AND DISCUSSIONS

Many people view data mining techniques as a special tool for finding information and hidden patterns in huge collections. In order to find hidden knowledge for future medical diagnosis, the study's methodology, which focused on interfering with COHD databases, identifies patterns from large-scale databases. Age, year, gender, and death number were among the variable factors that were used to discretize the data [25]. From new born days to 89 years old, data for a variety of characteristics, including age groupings, were taken into consideration. The decade from 2006 to 2016 was chosen as the time period.

Figure 3 shows the variation in mortality rates between 2006 and 2016 for different age groups. Based on statistical study, post neonatal fatalities accounted for the greatest number of deaths in 2008 and 2009. Second, from 2007 to 2016, there was a decline in the rates of early, late, and postnatal neonatal mortality. Thirdly, starting in 2013, there has been a minor rise in the death rate for patients in the following age groups: 20–24, 25–29, 30-34, 35–39, and 40–44. A gradual growth is being noted, rather than one that is highly pronounced. Hierarchical clustering provides more evidence in favour of this finding.

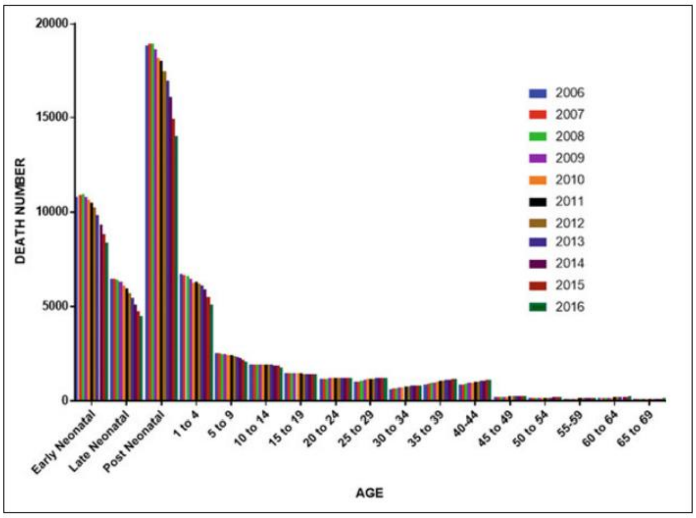

Fig.3. Mortality rate utilising hierarchical clustering

The dendrogram's vertical lines show how far apart the clusters are from one another. According to the analysis in Fig. 4, cluster 3 had the greatest number of deaths, followed by set 2 and cluster 1. We attempted to integrate the scatter plot with the clusters in order to visualise the shifting pattern of death across various age groups. We plotted the number of deaths each year on the Y-axis and the age on the X-axis. Based on age clusters, the data is categorised. This scatter plot showed that, between 2006 and 2016, there was a steady rise in the number of deaths in collection 1, particularly in the age groups of 25–29, 30–34, 35–39, and 40–44. Even while the trend is still rising gradually, this change in the data points to improvements in the care of congenital patients, which extends their lives.

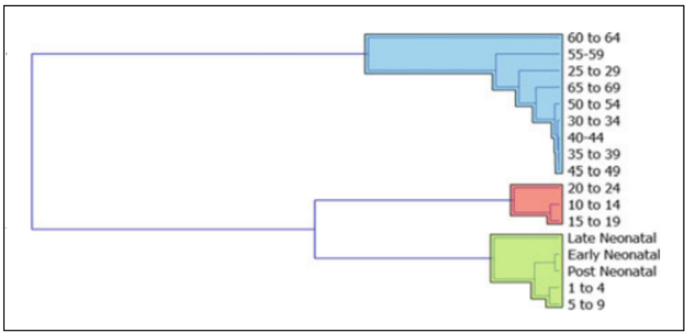

Fig.4. Age group cluster

5. CONCLUSION

Predictive data analysis holds immense potential for transforming healthcare by enabling proactive and personalized care. Despite challenges related to data quality, privacy, and model interpretability, advancements in machine learning and data integration are paving the way for more effective and efficient healthcare delivery. Continued research and collaboration are essential for overcoming these challenges and harnessing the full potential of predictive analytics in healthcare. Over the past ten years, people with congenital cardiac abnormalities have lived longer because to recent medical improvements. Significant mortality in 2006–07 largely happened in the post-neonatal age range, but since 2013, the longevity of congenital heart patients' disability has gradually increased, reaching 30–40 years.

REFERENCES

[1] Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, et al. Heart disease and stroke statistics—2013 update: a report from the American Heart Association. Circulation. 2013;127(1):e6–e245.

[2] van der Linde, D., Konings, E.E., Slager, M.A., Witsenburg, M., Helbing, W.A., Takkenberg, J.J., Roos-Hesselink, J.W.: Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. J. Am. Coll. Cardiol 58(21), 2241–2247 (2011)

[3] Hosseinkhah, F., Ashktorab, H., Veen, R.: Challenges in data mining on medical databases. In: Database Technologies: Concepts, Methodologies, Tools, and Applications, IGI Global, USA, pp. 1393–1404 (2009)

[4] Milley, A.: Healthcare and data mining. Health Manag. Technol. 21(8), 44–45 (2000)

[5] Masethe, H.D., Masethe, M.A.: Prediction of heart disease using classification algorithms. Proc. World Congr. Eng. Comput. Sci. 2, 22–24 (2014)

[6] Khairy, P., Ionescu-Ittu, R., Mackie, A.S., Abrahamowicz, M., Pilote, L., Marelli, A.J.: Changing mortality in congenital heart disease. J. Am. Coll. Cardiol. 56(14), 1149–1157 (2010)

[7] Kaur, H., Chauhan, R., Wasan, S.K.: A Bayesian network model for probability estimation. In: Encyclopedia of Information Science and Technology, Third Edition, pp. 1551–1558, IGI Global, USA (2015)

[8] Hazra, S., Ghosal, S., Mondal, A., Dey, P. (2024). Forecasting of Rainfall in Subdivisions of India Using Machine Learning. In: Bhattacharyya, S., Das, G., De, S., Mrsic, L. (eds) Recent Trends in Intelligence Enabled Research. DoSIER 2023. Advances in Intelligent Systems and Computing, vol 1457. Springer, Singapore. https://doi.org/10.1007/978-981-97-2321-8_18

[9] Hazra, S., Chatterjee, S., Mandal, A., Sarkar, M., Mandal, B.K. (2023). An Analysis of Duckworth-Lewis-Stern Method in the Context of Interrupted Limited over Cricket Matches. In: Chaki, N., Roy, N.D., Debnath, P., Saeed, K. (eds) Proceedings of International Conference on Data Analytics and Insights, ICDAI 2023. ICDAI 2023. Lecture Notes in Networks and Systems, vol 727. Springer, Singapore. https://doi.org/10.1007/978-981-99-3878-0_46

[10] T˘aranu, I.: Data mining in healthcare: decision making and precision. Database Syst. J. 6(4), 33–40 (2016)

[11] Hoffman, J.I., Kaplan, S.: The incidence of congenital heart disease. J. Am. Coll. Cardiol. 39(12), 1890–1900 (2002)

[12] Kaur, H., Chauhan, R., Ahmed, Z.: Role of data mining in establishing strategic policies for the efficient management of healthcare system—A case study from Washington DC area using retrospective discharge data. BMC J. Health. Serv. Res. 12(1), 12 (2012)

[13] Lemke, F., Mueller, J.A.: Medical data analysis using self-organizing data mining technologies. Syst. Anal. Modell. Simul. 43(10), 1399–1408 (2003)

[14] Sowmiya, C., Sumitra, P.: Comparative study of predicting heart disease by means of data mining. Int. J. Eng. Comput. Sci. 5(12), 19580–19582 (2016)

[15] Esfandiari, N., Babavalian, M.R., Moghadam, A.M.E., Tabar, V.K.: Knowledge discovery in medicine: current issue and future trend. Expert Syst. Appl. 41(9), 4434–4463 (2014)

[16] Xing, Y., Wang, J., Zhao, Z.: Combination data mining methods with new medical data to predicting outcome of coronary heart disease. In: Proceedings of IEEE International Conference on Convergence Information Technology, pp. 868–872 (2007)

[17] Jothi, N., Husain, W.: Data mining in healthcare–a review. Proc. Comput. Sci. 72, 306–313 (2015)

[18] Srinivas, K., Rani, B.K., Govrdhan, A.: Applications of data mining techniques in healthcare and prediction of heart attacks. Int. J. Comput. Sci. Eng. 2(02), 250–255 (2010)

[19] Zomer, A.C., Ionescu-Ittu, R., Vaartjes, I., Pilote, L., Mackie, A.S., Therrien, J., Langemeijer, M.M., Grobbee, D.E., Mulder, B.J.M., Marelli, A.J.: Sex differences in hospital mortality in adults with congenital heart disease: the impact of reproductive health. J. Am. Coll. Cardiol. 62(1), 58–67 (2013)

[20] Lamour, J.M., Kanter, K.R., Naftel, D.C., Chrisant, M.R., Morrow, W.R., Clemson, B.S., Kirklin, J.K.: The effect of age, diagnosis, and previous surgery in children and adults undergoing heart transplantation for congenital heart disease. J. Am. Coll. Cardiol. 54(2), 160–165 (2009)

[21] Marelli, A.J., Ionescu-Ittu, R., Mackie, A.S., Guo, L., Dendukuri, N., Kaouache, M.: Lifetime prevalence of congenital heart disease in the general population from 2000 to 2010. Circulation 130(9), 749–756 (2014)

[22] Marelli, A.J., Mackie, A.S., Ionescu-Ittu, R., Rahme, E., Pilote, L.: Congenital heart disease in the general population: changing prevalence and age distribution. Circulation 115(2), 163–172 (2007)

[23] Hazra, Sudipta, Swagata Mahapatra, Siddhartha Chatterjee, and Dipanwita Pal. "Automated Risk Prediction of Liver Disorders Using Machine Learning." In *the proceedings of 1st International conference on Latest Trends on Applied Science, Management, Humanities and Information Technology (SAICON-IC-LTASMHIT-2023) on 19th June*, pp. 301-306. 2023.

[24] Banerjee, S., Hazra, S., Kumar, B. (2023). Application of Big Data in Banking—A Predictive Analysis on Bank Loans. In: Peng, SL., Jhanjhi, N.Z., Pal, S., Amsaad, F. (eds) Proceedings of 3rd International Conference on Mathematical Modeling and Computational Science. ICMMCS 2023. Advances in Intelligent Systems and Computing, vol 1450. Springer, Singapore. https://doi.org/10.1007/978-981-99-3611-3_40

[25] Gon, Anudeepa, Sudipta Hazra, Siddhartha Chatterjee, and Anup Kumar Ghosh. "Application of Machine Learning Algorithms for Automatic Detection of Risk in Heart Disease." In *Cognitive Cardiac Rehabilitation Using IoT and AI Tools*, pp. 166-188. IGI Global, 2023

[26] Sima Das, Siddhartha Chatterjee, Sutapa Bhattacharya, Solanki Mitra, Arpan Adhikary and Nimai Chandra Giri “Movie’s-Emotracker: Movie Induced Emotion Detection by using EEG and AI Tools”, In the proceedings of the 4th International conference on Communication, Devices and Computing (ICCDC 2023), Springer-LNEE SCOPUS Indexed, DOI: 10.1007/978-981-99-2710-4_46, pp.583-595, vol. 1046 on 28th July, 2023.

[27] Payel Ghosh, Sudipta Hazra and Siddhartha Chatterjee, “Future Prospects Analysis in Healthcare Management Using Machine Learning Algorithms” In the International Journal of Engineering and Science Invention (IJESI), ISSN (online): 2319-6734, ISSN (print): 2319-6726, Vol.12, Issue 6, pp. 52-56, Impact Factor – 5.962, UGC Sl. No.- 2573, Journal No.- 43302, DOI: 10.35629/6734-12065256, June 17, 2023.

[28] Soumen Das, Siddhartha Chatterjee, Debasree Sarkar and Soumi Dutta, “Comparison Based Analysis and Prediction for Earlier Detection of Breast Cancer Using Different Supervised ML Approach” In Emerging Technologies in Data Mining and Information security (IEMIS 2022), Advances in Intelligent Systems and Computing (AISC, volume 1348, pp 255 – 267 on 30 September 2022,http://doi.org/10.1007/978-981-19-4193-9 (Springer).

[29] Ghosh, P., Dutta, R., Agarwal, N., Chatterjee, S., & Mitra, S. (2023). Social media sentiment analysis on third booster dosage for COVID-19 vaccination: a holistic machine learning approach. *Intelligent Systems and Human Machine Collaboration: Select Proceedings of ICISHMC 2022*, 179-190.

[30] Sudipta Hazra, Surjyasikha Das, Rituparna Mondal, Prerona Sanyal, Anwesa Naskar, Pratiksha Hazra, Kuntal Bose, Shirsha Mullick, Swarnakshi Ghosh and Siddhartha Chatterjee “Pervasive Nature of AI in the Health Care Industry: High-Performance Medicine”, In International Journal of Research and Analysis in Science and Engineering (IJRASE), Peer Reviewed UGC Sponsored, ISSN: 2582-8118, Vol. 4, Issue. 1, pp. 1-16 on 10th January, 2024.