ISSN 2455 – 9091

IJCSET JUNE Volume 10 Issue 6

International Journal of Computer Science Engineering and Technology (IJCSET)

REVIEW ON DATA MINING STRATEGIES FOR EARLY ALZHEIMER'S DETECTION

[1] R. Malarvizhi, [2] Dr. R. Rangaraj
[1, 2] Assistant Professor,
[1, 2] Department of Computer Science,
[1] N.M.S.S.Vellaichamy Nadar College, Madurai.
[2] Hindusthan College of Arts & Science, Coimbatore.

ABSTRACT: - This extensive review critically examines a range of methodologies for data mining approaches for early Alzheimer's detection. It investigate the effectiveness of machine learning, pattern recognition, and statistical techniques on a variety of datasets to provide insight into how well they might detect early indicators of Alzheimer's disease. The poll underscores the importance of utilizing cutting-edge analytics and innovative algorithms to improve early detection accuracy and efficiency. The knowledge gained from this investigation can greatly aid in the current attempts to identify Alzheimer's early on, opening the door to prompt therapies and better patient outcomes.

Keywords: [Alzheimer's disease, early detection, data mining, machine learning, pattern recognition, statistical methods.]

1. INTRODUCTION

Alzheimer's disease (AD) is expected to become more common in the upcoming decades, posing a serious threat to world health. Timely intervention is hampered by the deceptive nature of AD and the absence of reliable early diagnostic instruments. The incorporation of data mining techniques presents a viable approach for discovering complex patterns in various datasets, thereby aiding in the prompt identification of Alzheimer's disease in answer to this urgent need. Given the complexity and diversity of the disease, conventional diagnostic techniques frequently prove inadequate for reliable early assessments. Thus, a comprehensive strategy that utilizes data mining from multiple sources, including genetics, brain imaging, and clinical records, provides a comprehensive viewpoint.

This method looks for minute patterns that point to Alzheimer's disease's pre-symptomatic stages. This review of the literature aims to carefully investigate the state of the art, with a particular emphasis on integrated data mining techniques for the early detection of Alzheimer's disease. This review attempts to investigate the developments, difficulties, and new trends in the field of machine learning, statistical analysis, and artificial intelligence by combining results from a multitude of research. The main goal is to present a comprehensive overview of approaches and to critically assess how well they differentiate between cognitive decline and normal aging. The use of sophisticated algorithms to explore brain imaging data, genetic markers, and clinical factors is one of the main research topics in this research. With the use of brain imaging methods like MRIs and PET scans, scientists may examine the anatomical and functional features of the brain. Conversely, genetic markers provide information about the genetic variants and predispositions linked to Alzheimer's disease.

A wide range of factors are included in clinical variables, such as medical history, cognitive tests, and demographic data. It may be possible to identify new biomarkers that are essential for the early identification of Alzheimer's disease and to unearth complex linkages by incorporating these many data sources into data mining models. The discipline has evolved substantially with the introduction of machine learning techniques, which enable the extraction of intricate patterns and connections from massive datasets. When paired with machine learning algorithms, statistical analysis improves the capacity to identify minute changes that might be suggestive of Alzheimer's disease in its early stages. Furthermore, deep learning models in particular show that artificial intelligence is capable of digesting complex data patterns, which helps to create predictive models that are more accurate.

The research evaluation does, however, critically recognize the difficulties in integrating data mining techniques for Alzheimer's detection despite these developments. There are many obstacles because data sources are heterogeneous, data quality varies, and consistent processes are required. In addition, it is imperative to tackle matters concerning data security and privacy in order to promote confidence and ease the exchange of datasets for cooperative research endeavors. The assessment highlights how data integration methods are always changing and stresses the value of cooperative efforts. In order to create more reliable and broadly applicable models, these cooperative efforts seek to combine various datasets that include a wide range of variables and data types. Collaborative techniques play a critical role in guaranteeing the generalizability and efficacy of data mining models across heterogeneous populations as datasets increase in size and complexity.

2. LITERATURE REVIEW

1. B. V. Baiju (2019) et.al proposed Disease Influence Measure Based Diabetic Prediction with Medical Data Set Using Data Mining. This study extensively tackles the challenge of predicting diabetes through a thorough exploration of diverse data mining methods. Employing a medical dataset for predicting diabetes mellitus, the research conducts a comprehensive review on approaches to disease prediction. While disease identification relies on symptoms, accuracy varies with distinct features and methods. Introducing a Disease Influence Measure (DIM)-based diabetic prediction, the method involves preprocessing, noise removal, DIM estimation, and subsequent prediction. The research systematically compares different prediction approaches, highlighting their performance variations. It underscores the significant impact of changing lifestyles on health, particularly the onset of diabetes mellitus, emphasizing the critical role of insulin in blood sugar regulation. By evaluating various prediction methods and introducing a DIM-based algorithm, the study contributes to the refinement of diabetic prediction techniques. The evaluation, based on a dataset featuring 19 features and 605 medical records, lays a foundation for future advancements in predictive methodologies for diabetes.

2. D. B. Mehta (2019) et.al proposed Newfangled Approach for Early Detection and Prevention of Ischemic Heart Disease using Data Mining. In the ever-expanding realms of biotechnology and health sciences, copious datasets, especially concerning heart diseases, pose a formidable challenge for identification based solely on symptoms. This study employs data mining classification techniques on ischemic stroke data to establish a recommendation system for assessing the risk of Ischemic Heart Disease (IHD). Early detection of IHD is pivotal for timely intervention and heightened survival rates. Applying Logistic Regression, Decision Tree, K Nearest Neighbor, Naïve Bayes, and SVM algorithms to an Ischemic Stroke Dataset, the research attains a commendable 97.91% accuracy, particularly with Support Vector Machine. The novel recommendation system not only discerns risk levels but also offers tailored interventions, encompassing diet and exercise recommendations, along with insights into the nearest heart hospitals and specialists. This pioneering research introduces a user-friendly approach, empowering individuals to comprehend their IHD risk factors, fostering early awareness and promoting preventive measures for a healthier lifespan.

3. G. Choudhary (2019) et.al proposed Prediction of Cardiovascular Disease using Data Mining Technique. This research delves into the analysis of cardiovascular diseases across diverse age groups, employing a preprocessed dataset derived from 299 patient records containing 16 attributes. Noteworthy medicinal features such as 'dig,' 'prop,' 'nitr,' and 'fast grume sugar' are thoroughly examined for their predictive capabilities regarding cardiovascular illness. The study evaluates three data mining classification techniques—naïve Bayes, random tree, and REPTree and finds that the random tree algorithm excels in providing accurate prognostic insights. Given the global impact of cardiovascular diseases, constituting approximately 31% of all fatalities, this investigation underscores the critical importance of early detection. The paper introduces a prognostic system utilizing Naïve Bayesian, Random Tree, and REPTree classifiers, emphasizing the superior precision and accuracy of the random tree algorithm in forecasting early cardiovascular illnesses. Nevertheless, the study advocates for potential improvements through the inclusion of additional attributes and diverse data mining approaches. In addressing the imperative for efficient classifiers in healthcare data mining, this work contributes to advancing early detection and intervention strategies for cardiovascular diseases.

4. S. Cheng (2020) et.al proposed Pathogenic Factors Mining Based on Medical Big Data Using Association Analysis Algorithm: A Case of Anorectic Diseases. This study tackles the intricate and varied pathogenic factors associated with anorectic diseases in Guangdong Province. It introduces an association analysis algorithm designed to extract valuable insights from extensive medical datasets. An enhanced Apriori algorithm, more efficient than its classical counterpart, improves runtime by eliminating irrelevant itemsets. Through simulations, the algorithm exhibits superior performance in identifying disease association rules and potential pathogenic factors. The incorporation of big data technology in medicine proves pivotal, fostering collaboration among medical institutions, researchers, and data analysts to enhance comprehension of disease causation. The improved Apriori algorithm excels in expediting runtime without compromising efficacy, as evidenced by practical simulations.

5. R. A. Canessane (2019) et.al proposed HUSP Mining Techniques to Detect Most Weighted Disease and Most Affected Diseases for the Healthcare Industry. Data mining, a vital element of Knowledge Discovery in Databases (KDD), plays a crucial role in extracting valuable information, particularly in the healthcare domain where vast patient records often remain underutilized. This research introduces the HUSP (High Utility Sequence Pattern) algorithm, emphasizing its significance in efficiently detecting and prioritizing impactful diseases. The proposed approach encompasses comprehensive stages, including data collection, pre-processing, feature selection, clustering, and classification, with a focus on identifying the most significant diseases. Sequential pattern mining, a core component, utilizes the HUSP algorithm within the broader context of machine learning. The study underscores the effectiveness of the classification procedure, aiming to derive meaningful patterns from extensive healthcare datasets. With the goal of enhancing decision-making for healthcare officials and improving service quality, this research highlights the untapped potential within healthcare data and emphasizes the pivotal role of data mining in revealing valuable patterns and relationships. The findings suggest promising prospects for disease classification, addressing challenges related to accuracy and performance observed in previous experiments.

6. M. F. Kabir (2018) et.al proposed Rule Discovery from Breast Cancer Risk Factors using Association Rule Mining.

Breast cancer, a pervasive global concern among women, underscores the imperative of risk factor reduction for effective prevention. Identifying pivotal insights and patterns within breast cancer risk factor datasets is vital for developing preventive strategies. The primary objective of this study is to uncover and analyze pertinent information in the form of rules extracted from a breast cancer risk factor dataset, aiming to provide actionable insights for prevention initiatives. Leveraging association rule mining, a key data mining technique, allows the extraction of meaningful rules that succinctly express valuable information. These rules serve as comprehensible guidelines for end-users, facilitating a deeper understanding of potential risk factors and aiding in the initiation of targeted prevention strategies. The utilization of association rule mining in breast cancer risk factor analysis contributes to the ongoing efforts in harnessing data-driven approaches for the advancement of preventive measures and the reduction of the disease's impact on the global population.

7. Tafish (2018) et.al proposed Breast Cancer Severity Degree Predication Using Data Mining Techniques in the Gaza Strip. Data mining has become an integral strategy in medical applications, particularly in the field of medicine. It is defined as a method that extracts valuable information from raw datasets by analyzing and compressing them from various perspectives. In the realm of healthcare, medical data mining involves techniques that extract valuable and novel insights from healthcare databases to aid physicians in achieving optimal diagnoses. In the Gaza Strip, cancer and diabetes have emerged as prevalent health concerns in recent years. Given the costly and prolonged nature of diagnostic tests for these diseases, data mining proves to be a widely applicable solution. Building on previous breast cancer research, this study proposes a model to address the challenge of determining the risk level for cancer, aiming to enhance efficiency, reduce time and costs based on data collected from hospitals in the Gaza Strip. The analysis includes association rule mining to unveil relationships between attributes and the development of predictive models (classifiers) utilizing data mining techniques such as SVM, ANN, and KNN, followed by an evaluation and comparison of the intelligent models and their outcomes.

8. N. Ramkumar (2017) et.al proposed Prediction of liver cancer using Conditional probability Bayes theorem. Cancer stands as one of the most perilous diseases globally, manifesting in various forms such as lung, breast, bone, and notably, liver cancer, which is exceptionally hazardous and often has lifelong implications. Symptoms of liver cancer encompass jaundice, weight loss, yellow-shaded urine, vomiting, pain in the upper right abdomen, sweats, fever, and an enlarged liver. Liver cancer is categorized into primary, originating within the liver, and secondary, spreading from other body parts. The liver, a vital organ, is significantly impacted, with around 30 out of 100,000 individuals experiencing liver cancer, predominantly affecting African and Asian nations. Hepatocellular carcinoma is the most prevalent form, notably affecting males more than females. Mainly attributed to excessive alcohol consumption, the prevalence of liver cancer has surged, making it a prominent global health concern. Data mining algorithms and artificial intelligence concepts, notably the application of Bayes' theorem with the WEKA tool, are increasingly utilized for predicting the likelihood of liver cancer, contributing to early detection and intervention.

9. Ghasemi Z (2020) et.al proposed automated chagas disease vectors identification using data mining techniques. Chagas disease (CD), a widespread vector-borne zoonotic illness, poses a significant global health threat, transmitted primarily by triatomine insects or kissing bugs. Detectable at any stage, CD diagnosis relies on analyzing clinical, epidemiological, and laboratory data. With distinct acute and chronic phases, early detection in the acute stage is crucial for effective control and treatment. Despite numerous clinical trials, progress in automatic identification research has been slow. This study introduces two automatic CD vector identification systems, employing pre-processing, feature extraction, and classification phases. Utilizing Principal Component Analysis (PCA) for feature extraction and Random Forest (RF) and Support Vector Machine (SVM) for classification, the proposed methods exhibit promising results. The PCA–SVM approach achieves 87.62% accuracy for 12 Mexican and 75.26% for 39 Brazilian species, while the PCA–RF approach attains 100% accuracy for both Brazilian and Mexican species. These outcomes surpass existing automatic identification systems, showcasing the efficacy and potential of the proposed methods in CD vector classification.

10. C. N. Aher (2020) et.al proposed Soft Computing based Approaches for Classifying Diseases using Medical Diagnosis Dataset. The medical field grapples with vast data volumes regularly, and conventional methods may be inadequate for deriving meaningful insights. In contemporary biology research, the simultaneous measurement of thousands of gene expressions has become routine, necessitating advanced data analysis approaches. Machine learning algorithms emerge as invaluable tools in medical research, particularly for disease prediction based on molecular-level information encoded in gene expression profiles. Swift and accurate disease identification is paramount for prescribing patient medications and guiding physicians. Utilizing algorithms such as Decision Trees, Support Vector Machines, Artificial Neural Networks, Bayes classification, and k-Nearest Neighbors, researchers can discern patterns indicative of various diseases. Machine learning enhances the efficiency of disease prediction, especially with the DNA micro-array method capable of tracking extensive gene expressions. Researchers leverage this wealth of gene data to probabilistically diagnose diseases, resulting in the development of several techniques that yield promising and reliable results in predicting and understanding various medical conditions.

11. M. I. Qrenawi (2018) et.al proposed Identification of Cardiovascular Diseases Risk Factors among Diabetes Patients Using Ontological Data Mining Techniques. Data mining, an integral facet of Knowledge Discovery in Databases (KDD), holds paramount importance in extracting substantial information. In healthcare, where extensive

patient records often remain underutilized, data mining proves indispensable. This study introduces the HUSP (High Utility Sequence Pattern) algorithm, highlighting its crucial role in detecting and prioritizing the most impactful diseases. The proposed approach encompasses comprehensive stages, including data collection, pre-processing, feature selection, clustering, and classification, aimed at identifying diseases with the highest impact. Sequential pattern mining, a central component, incorporates the HUSP algorithm within the broader context of machine learning. The research accentuates the efficacy of the classification procedure, seeking to unveil meaningful patterns within the vast healthcare dataset. With a focus on enhancing decision-making for healthcare officials and improving service quality, this study underscores the untapped potential within healthcare data and the pivotal role of data mining in revealing valuable patterns and relationships. The findings suggest promising prospects for classifying diverse diseases, addressing accuracy and performance challenges observed in previous experiments.

12. R. Patra (2019) et.al proposed Predictive Analysis of Rapid Spread of Heart Disease with Data Mining. Cardiovascular disease, a leading global cause of mortality across all age groups and genders, is on the rise among both older and younger populations. Effectual health monitoring strategies are imperative to curb this escalating epidemic. This study addresses the critical need for early detection and treatment by proposing a heart disease prediction method using the UCI machine learning repository's Cleveland heart disease database. Employing the Information Gain concept for attribute selection, features are processed using Weka and Python, with a focus on addressing the research gap in heart disease prediction. Utilizing Python Anaconda Navigator, Spider, and Weka platforms, the study employs various techniques, including feature selection, numpy, panda's library, decision tree classifier, KNN classifier, and entropy, gini-index, and confusion matrix. Results underscore the decision tree classifier's efficacy in predicting outcomes from the UCI Cleveland heart dataset, emphasizing the urgent requirement for accurate predictive models to combat the alarming rise of cardiovascular diseases and prioritize timely intervention and management strategies.

13. A. A. Haruna (2019) et.al proposed An Improved C4.5 Data Mining Driven Algorithm for the Diagnosis of Coronary Artery Disease. Coronary artery disease (CAD) poses a persistent global health challenge, particularly in developed nations, representing a significant cause of mortality. This research introduces an advanced C4.5 data mining algorithm tailored for CAD diagnosis, surpassing the conventional C4.5 in performance metrics. The enhanced algorithm demonstrates an impressive overall accuracy of 97.23%, coupled with notable specificity (97.03%) and sensitivity (96.39%). The resulting decision tree, with twenty-seven leaves and forty-seven nodes, can be translated into production rules, enriching the knowledge base for an expert system in CAD diagnosis. This addresses a longstanding knowledge acquisition bottleneck in healthcare expert systems. Beyond CAD, the study underscores the versatility of data mining, with potential applications in diverse domains such as manufacturing, engineering, and education. The algorithm's success in CAD diagnosis highlights its potential adaptability to various datasets, emphasizing the far-reaching impact of advanced data mining techniques across industries and sectors.

14. V. Chandra kala (2019) et.al proposed Prediction of diseases with pathological characteristics classification using data mining. In the expansive realm of healthcare data, the integration of machine learning algorithms has revolutionized disease prediction, ushering in unprecedented advancements. From predicting epidemic outbreaks to fortifying healthcare data storage and security, machine learning applications in healthcare hold the promise of precise outcomes. A primary focus lies in leveraging machine learning to elevate patient care, streamlining disease identification and ensuring accurate diagnoses. This paper provides a comprehensive overview of implementing machine learning technologies for disease prediction, achieving a notable 70% accuracy in matching predicted infections and symptoms. It delves into the exploration of optimized machine learning algorithms tailored for diverse disease predictions, harnessing the contemporary era of machine learning to bolster early detection. The integration of these algorithms in healthcare presents a proactive approach to disease prediction, empowering medical professionals with efficient diagnostic tools and aiding in informed decision-making for enhanced patient well-being. Additionally, the paper introduces an analysis system that combines machine learning algorithms and the Internet of Things (IoT) for a comprehensive and advanced healthcare solution.

15. S. J. Pasha (2019) et.al proposed Bio inspired Ensemble Feature Selection (BEFS) Model with Machine Learning and Data Mining Algorithms for Disease Risk Prediction. In the realm of healthcare, the convergence of machine learning (ML) and data mining (DM) algorithms has gained substantial traction, particularly in disease risk prediction. This research introduces an innovative Bio-inspired Ensemble Feature Selection (BEFS) model, employing a combination of genetic algorithm (GA) and random forest algorithm, to discern pivotal features crucial for disease prediction. Applied to the Breast Cancer Wisconsin (Diagnostic) dataset, the BEFS model significantly enhances accuracy in conjunction with logistic regression (LR) and random forest (RF) ML algorithms, boasting an impressive 96.49% accuracy, 96% Area under Curve (AUC), and 98.11% sensitivity. Focusing on breast cancer, a leading cause of cancer-related deaths in women, underscores the imperative of early detection. The BEFS model's efficacy lies in reducing the dataset's features from thirty-two to six, offering a promising avenue for accurate and streamlined disease risk prediction. Moreover, the model's potential applicability to diverse disease datasets suggests its broader utility in advancing diagnostic capabilities across various medical domains.

3. PROPOSED METHODS, MERITS AND DEMERITS

Authors	Proposed Method	Merits	Demerits
B. V. Baiju (2019)	Disease Influence Measure Based Diabetic Prediction with Medical Data Set Using Data Mining	DIM-based diabetes prediction contributes to improved disease prediction approaches by improving accuracy through thorough pre-processing.	The complexity of the DIM technique may make it difficult to apply and interpret, necessitating careful assessment of computing power and human comprehension.
D. B. Mehta (2019)	Newfangled Approach for Early Detection and Prevention of Ischemic Heart Disease using Data Mining	Uses SVM to achieve high accuracy (97.91%), improving dependability for Ischemic Heart Disease (IHD) risk early identification.	The SVM algorithm's complexity may necessitate large computing resources, which in some circumstances may restrict real-time use.
G. Choudhary (2019)	Prediction of Cardiovascular Disease using Data Mining Technique	In a variety of age groups, the random tree method improves early prognosis for cardiovascular illnesses with its excellent precision and accuracy.	Comprehensive analysis may be hampered by limited dataset attributes; nevertheless, refinement through the addition of new features and the use of different data mining techniques is possible.
S. Cheng (2020)	Pathogenic Factors Mining Based on Medical Big Data Using Association Analysis Algorithm: A Case of Anorectic Diseases.	Better By removing irrelevant itemsets, the apriori approach improves efficiency and shortens the runtime for identifying illness association rules from large amounts of medical data.	Applying the algorithm in various medical contexts may present integration issues that need to be adjusted for best results.
R. A. Canessane (2019)	HUSP Mining Techniques to Detect Most Weighted Disease and Most Affected Diseases for the Healthcare Industry	Through efficient data mining, the HPS algorithm improves disease prioritization in healthcare, optimizing decision-making and raising the standard of service.	Potential issues with scalability or adaptability to other healthcare settings could occur; therefore broader applications would need to be carefully considered.
M. F. Kabir (2018)	Rule Discovery from Breast Cancer Risk Factors using Association Rule Mining	The response variable is twice in an examination for breast cancer, the twofold logit model is frequently used.	One of the most important tasks in data mining is rule mining since rules provide concise declarations of potentially important information that end users may easily comprehend.
Tafish (2018)	Breast Cancer Severity Degree Predication Using Data Mining Techniques in the Gaza Strip	SVM has recently been used in several applications, such as medical diagnosis, handwritten digit recognition, and text categorization.	The degree of the disease indicates greater prominence when there is a significant difference between the contaminated cells, making Poorly-diff the most problematic instance.
N. Ramkumar (2017)	Prediction of liver cancer using Conditional probability Bayes theorem	The Bayes theory is applied using the WEKA tool to determine the probability of predicting liver cancer.	There is little analysis done on the predictions made by data mining algorithms.
Ghasemi Z (2020)	Automated chagas disease vectors identification using data mining techniques	Different data balancing techniques may have an important effect on improving the results specially using the SVM technique.	Need for high-resolution images, their tremendous pre-processing.
C. N. Aher (2020)	Soft Computing based Approaches for Classifying Diseases using Medical Diagnosis Dataset	Classification methods that is highly helpful for evaluating the severity of a disease and diagnosing it quickly.	Research is focused on development; feature selection is an undeveloped field.
M. I. Qrenawi (2018)	Identification of Cardiovascular Diseases Risk Factors among Diabetes Patients Using Ontological Data Mining Techniques	Using ontology-driven data mining improves learning accuracy by 90% and provides important information for estimating the risk of cardiovascular disease.	Operational difficulties may arise from the potential complexity and resource-intensiveness of establishing and maintaining ontological data mining techniques.

R. Patra (2019)	Predictive Analysis of Rapid Spread of Heart Disease with Data Mining	Using the UCI Cleveland heart dataset, the study successfully applies the decision tree classifier and shows how well it predicts outcomes in the setting of cardiovascular disease.	The findings may not be as broadly applicable to other populations and datasets due to the reliance on a particular dataset (UCI Cleveland heart disease).
A. A. Haruna (2019)	An Improved C4.5 Data Mining Driven Algorithm for the Diagnosis of Coronary Artery Disease	The improved C4.5 algorithm has remarkable CAD diagnostic accuracy (97.23%), offering important information for healthcare expert systems.	Further validation may be necessary for generalizability to diverse datasets outside of CAD diagnosis in order to achieve wider application.
V. Chandra kala (2019)	Prediction of diseases with pathological characteristics classification using data mining	With 70% accuracy, machine learning improves disease prediction, enabling early detection and proactive patient management.	The study might go into greater detail about the particular difficulties and constraints that arise when putting machine learning algorithms into practice.
S. J. Pasha (2019)	Bio inspired Ensemble Feature Selection (BEFS) Model with Machine Learning and Data Mining Algorithms for Disease Risk Prediction.	The BEFS model improves the effectiveness of early detection by achieving high accuracy (96.49%) and sensitivity (98.11%) in breast cancer risk prediction.	Limited generalizability due to the study's exclusive emphasis on breast cancer; additional validation may be needed for application to other diseases.

CONCLUSION

This thorough analysis has explored a range of data mining techniques for Alzheimer's disease early detection and illuminated the variety of approaches used in present research. The report highlights the constantly changing landscape of efforts to find preclinical signs of Alzheimer's disease, ranging from machine learning algorithms to brain imaging approaches. A review of literature shows that significant progress has been made in using advanced computational methods for early diagnosis. But problems still exist, calling for further interdisciplinary cooperation and the incorporation of cutting-edge technologies. The combination of several methodologies may open the door to more precise and trustworthy early detection techniques as the field develops, providing promise for better therapies in the crucial early stages of Alzheimer's.

REFERENCES

[1]. V. Baiju and D. J. Aravindhar, "Disease Influence Measure Based Diabetic Prediction with Medical Data Set Using Data Mining," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), Chennai, India, 2019, pp. 1-6, doi: 10.1109/ICIICT1.2019.8741452.

[2]. B. Mehta and N. C. Varnagar, "Newfangled Approach for Early Detection and Prevention of Ischemic Heart Disease using Data Mining," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 1158-1162, doi: 10.1109/ICOEI.2019.8862544.

[3]. G. Choudhary and S. N. Singh, "Prediction of Cardiovascular Disease using Data Mining Technique," 2019 4th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2019, pp. 99-103, doi: 10.1109/ISCON47742.2019.9036256.

[4]. S. Cheng, J. Qiu, Z. Lai, C. Peng and K. Lu, "Pathogenic Factors Mining Based on Medical Big Data Using Association Analysis Algorithm: A Case of Anorectal Diseases," 2020 7th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS), Guangzhou, China, 2020, pp. 37-41, doi: 10.1109/ICCSS52145.2020.9336886.

[5]. R. A. Canessane, R. Dhanalakshmi, B. Pavithra, B. Sasikanth and C. Sandeep, "HUSP Mining Techniques to Detect Most Weighted Disease and Most Affected Diseases for the Healthcare Industry," 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2019, pp. 25-32, doi: 10.1109/ICONSTEM.2019.8918784.

[6]. M. I. Qrenawi and W. Al Sarraj, "Identification of Cardiovascular Diseases Risk Factors among Diabetes Patients Using Ontological Data Mining Techniques," 2018 International Conference on Promising Electronic Technologies (ICPET), Deir El-Balah, Palestine, 2018, pp. 129-134, doi: 10.1109/ICPET.2018.00030.

[7]. R. Patra and B. Khuntia, "Predictive Analysis of Rapid Spread of Heart Disease with Data Mining," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2019, pp. 1-4, doi: 10.1109/ICECCT.2019.8869194.

[8]. Haruna, L. J. Muhammad, B. Z. Yahaya, E. J. Garba, N. D. Oye and L. T. Jung, "An Improved C4.5 Data Mining Driven Algorithm for the Diagnosis of Coronary Artery Disease," 2019 International Conference on Digitization (ICD), Sharjah, United Arab Emirates, 2019, pp. 48-52, doi: 10.1109/ICD47981.2019.9105844.

[9]. V. Chandra kala, L. Venkateswara kiran and P. Siva Prasad, "Prediction of diseases with pathological characteristics classification using data mining," 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), Vellore, India, 2019, pp. 1-5, doi: 10.1109/ViTECoN.2019.8899346.
[10]. S. J. Pasha and E. S. Mohamed, "Bio inspired Ensemble Feature Selection (BEFS) Model with Machine Learning and Data Mining Algorithms for Disease Risk Prediction," 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2019, pp. 1-6, doi: 10.1109/ICCUBEA47591.2019.9129304
[11]. N. Aher and A. K. Jena, "Soft Computing based Approaches for Classifying Diseases using Medical Diagnosis Dataset," 2020 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2020, pp. 77-81, doi: 10.1109/ESCI48226.2020.9167518.
[12]. Kabir MF, Ludwig SA, Abdullah AS. Rule discovery from breast cancer risk factors using association rule mining. In2018 IEEE International Conference on Big Data (Big Data) 2018 Dec 10 (pp. 2433-2441). IEEE.
[13]. Tafish MH, El-Halees AM. Breast cancer severity degree predication using data mining techniques in the gaza strip. In2018 International Conference on Promising Electronic Technologies (ICPET) 2018 Oct 3 (pp. 124-128). IEEE.
[14]. Ghasemi Z, Banitaan S, Al-Refai G. Automated chagas disease vectors identification using data mining techniques. In2020 IEEE international conference on electro information technology (EIT) 2020 Jul 31 (pp. 540-545). IEEE.
[15]. Ramkumar N, Prakash S, Kumar SA, Sangeetha K. Prediction of liver cancer using Conditional probability Bayes theorem. In2017 International Conference on Computer Communication and Informatics (ICCCI) 2017 Jan 5 (pp. 1-5). IEEE.