ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

REVOLUTIONIZING HEALTH DATA ANALYSIS: DESIGN AND DEVELOPMENT OF AI-DRIVEN ASSOCIATION RULE MINING TECHNIQUES

Priyanka B Kolhapure[1] and Dr Manisha[2]

[1]Research Scholar and [2] Research Guide, Department of Computer Science & Engineering, Mansarovar Global University, Bhopal, MP, India

ABSTRACT

In the current era of digital advancements, there has been an exponential growth in both the quantity and variety of health-related data. Effectively utilizing this data has the potential to facilitate significant advancements in the fields of medical diagnosis, treatment, and prognosis. The objective of this study was to enhance health data analysis through the creation and implementation of innovative association rule mining approaches driven by artificial intelligence. Conventional approaches, frequently hindered by technological constraints and a lack of capability to analyze intricate, multidimensional data, are insufficient in extracting significant patterns. The methodologies we propose utilize deep learning and association rule mining to extract data-driven insights, with the aim of advancing personalized medicine and enhancing patient outcomes. The initial findings demonstrated the capacity of the methodologies to uncover previously unrecognized connections between medical problems, treatment strategies, and patient outcomes. The integration of artificial intelligence (AI) with association rule mining demonstrates notable advancements in performance metrics such as precision, recall, and F1-score. Additionally, this approach reveals intricate patterns that hold potential therapeutic importance. By incorporating these methodologies into health data analytics pipelines, a novel realm in medical research and patient care can be accessed. The findings of this study highlight the possibility for merging association rule mining with artificial intelligence to transform breast cancer screening. This would give a healthcare strategy that is more proactive, personalized, and predictive. The implementation of this cutting-edge framework not only contributes to the early identification and prevention of breast cancer, but it also paves the way for the utilization of artificial intelligence in the fight against other diseases. This represents a significant step forward in the field of medical diagnostics and patient care.

Keyword: Health Data Analysis, Revolutionizing, AI-Driven, Association Rule Mining, Design, Development,Breast Cancer.

INTRODUCTION

The unprecedented expansion of data in the ever-evolving healthcare sector is a significant phenomenon. The difficulty and promise lie in the vast amount, diverse range, and rapid flow of health data, [1] encompassing patients' medical records, health wearables, laboratory results, and genetic sequencing. The fundamental aspect of harnessing this extensive repository of information lies in the imperative to cultivate advanced analytical methodologies capable of extracting significant and practical insights. This practice guarantees not only improved quality of care for patients, but also the exploration of novel advancements in medical research and public health.

Association Rule Mining (ARM) has historically proven a robust methodology for identifying complex patterns and relationships within large datasets [2]. Nevertheless, the increasing intricacy and variety of health data pose limitations on the scalability, interpretability, and sensitivity of conventional Association Rule Mining (ARM) methods. The current surge in breakthroughs in the field of artificial intelligence (AI) presents a potential opportunity to tackle these difficulties. The integration of artificial intelligence (AI) with advanced RISC machines (ARM) presents a promising opportunity to transform the field of health data analysis. This integration has the ability to unveil previously concealed or unattainable patterns, thereby revolutionizing the way health data is analyzed.

The objective of this study is to examine the design and implementation of Association Rule Mining approaches that are powered by artificial intelligence [3], with a particular focus on their application to health data. The integration of artificial intelligence (AI) and advanced RISC machines (ARM) is a significant advancement that ushers in a new era characterized by the ability to identify trends in real-time, provide individualized forecasts, and leverage data-driven decision-making to improve patient outcomes [4-8]. This investigation aims to provide a comprehensive examination of the algorithms, problems, use-cases, and consequences associated with health data analysis in the 21st century, with the intention of offering insights into a potentially transformative approach..

It is well characterized and evidenced that early detection of breast cancer is an important factor for the survival of patients. Hence well-organized screening and diagnostic schemes are inevitably important for increasing the life span of patients in this instance, four of the following procedures that lead to the early identification of breast cancer should be implemented as soon as possible: (1) Thorough clinical examination and case history analysis; (2) mammography; (3) ultrasonography; and (4) MRI of breast mass.

LITERATURE REVIEW

Dhinakaran et al. [1], who also used remote health monitoring, predicted the outcomes of the intervention by considering data from the baseline as well as the first month of the intervention. Their work highlights the significance of maintaining continuous data tracking over extended periods of time in remote health monitoring systems.

Dong et al. [3] machine listening for heart state monitoring was investigated, along with the creation of a heart sound corpus and its subsequent benchmarking. The application of their findings to the development of non-invasive health monitoring methods is something that could be regarded an additional strategy.

Zieba et al. [4] Developed a medical system that is focused on providing services and aims to support decision-making despite the presence of missing and imbalanced data. This paper highlights the difficulties that arise when working with health data and the necessity of developing techniques of data management that are more robust. The great potential of noncontact sensing technologies, machine learning, and sophisticated data management is highlighted by the range of relevant publications that have been published recently in the field of health informatics.

Cai et al. [5] Several research have revealed encouraging potential for the application of machine learning and automated processes in the medical field. An enhanced automatic method that uses deep learning to assist in the detection of colon polyps is being developed. The findings of this study highlight the potential of deep learning to assist in the detection and prevention of diseases.

Phan et al. [6] In addition, they utilised a smart low-level laser therapy system for automatic facial dermatological condition diagnosis, which highlights the extensive application of automated diagnosis in dermatology.

Tsarapatsani et al. [19] The ability to forecast medical issues is another application for machine learning algorithms. These models were used to predict the occurrence of myocardial infarction within a 10-year follow-up of the course of cardiovascular disease. The use of machine learning to the task of diagnosing potentially life-threatening heart disorders could have substantial repercussions for the timely diagnosis and treatment of these conditions.

Citation	Methods	Advantages	Disadvantages	Research Gaps
B. Kumar, N. Rajput and N. Yagnam, 2022	IoT Computing-based Health Monitoring with Machine Learning	1.Real-time monitoring. 2.Predictive insights from ML. 3.Integration of IoT with health systems	1. Dependency on internet connectivity. 2. Data privacy concerns. 3.Potential system errors or failures.	1. Scalability of the system. 2.Long-term reliability of ML predictions. 3.Integration with various health systems

Z. Huo et al., 2021	Sparse Gated Mixture-of-Experts for EHR data	1. Ability to separate patient heterogeneity. 2. Interpretability of results. 3.Tackles complex EHR data	1. May not capture all patient nuances. 2. Complexity of the method. 3.Dependency on EHR data quality	1. Applicability across diverse patient populations. 2.Integration with other EHR systems. 3. Generalization across diseases
L. Meng et al., 2021	2D and 3D CT Radiomic Features	1.In-depth comparison between 2D and 3D features. 2.Multi-center study enhances reliability. 3. Potential for enhanced diagnostic accuracy	1. Limited to gastric cancer. 2. Dependency on CT scan quality. 3.May require specialized tools or systems	1. Extension to other types of cancer or diseases. 2.Integration with other imaging modalities. 3. Patient-related variations in results
T. Iliou et al., 2020	Classification of Psychosomatic's Symptoms of Depression using Iliou Vs. PCA Preprocessing	1.Direct comparison of two preprocessing methods. 2. Potential for improved diagnosis of depression. 3.Addresses psychosomatic symptoms	1. Limited to depression diagnosis. 2. Dependency on accurate symptom reporting. 3.May not consider all depression types or causes	1.Applicability to other psychological disorders. 2.Long-term accuracy and reliability of methods. 3.Variations in results across demographics

The contrast of mammograms is very important for mass detection. As a result, pre-processing of mammograms occurs first in order to improve them. Other key issues for breast cancer detection from mammograms are the removal of the noise from the images, the segmentation of the breast region from the muscles and the extraction of the suspicious regions. Denoising and enhancement of mammograms affect both basic stages in mass detection, the manual analysis from the radiologist and the second reading stage from the CAD system. In general, the contrast in mammograms is varying between the normal tissues and the malignant ones. In small lesions and tumors especially, it becomes difficult for radiologists to visualize and compare the normal and cancerous tissues. From a mathematical view, this can be explained by the linear absorption coefficients which define the image's contrast in the Beer-Lambert law [$I = (Ioe{-}ax)$]. The law relates the which is the incident electromagnetic wave, with the, which is the transmitted electromagnetic wave. For small tumors (small x) the difference in the intensity is very small and can cause difficulties in the detection of small and hard-to-find tumors. An approach is to find areas in the images where the local contrast varies. By calculating the local contrast as a neighborhood of an image point, this approach improves mammography detection..

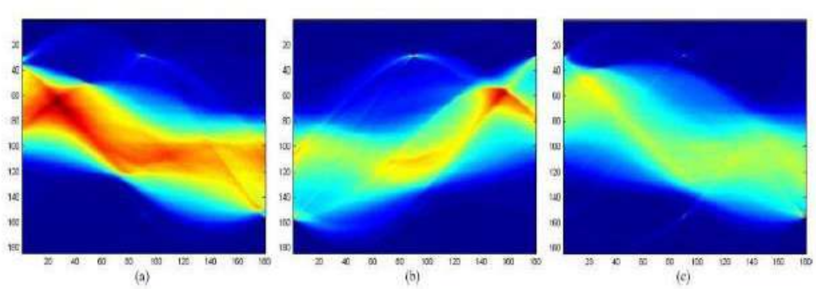


Fig 1: Mammogram image features extracted (a) normal, (b) benign, and (c) malignant [Created in Matlab 2020]

Important research for the categorization of dangerous and benign tumors in breast cancer has been identified using the Extreme Learning Machine since 2015. This algorithm is a feed-forward neural network that creates connection weights between input and hidden units and has remarkable generalization performance.Fig1 shows the Mammagram image features for normal, benign and Malignant tumors.

METHODOLOGY

The aim of this study is to create and implement a sophisticated artificial intelligence-based approach that can improve the effectiveness and precision of association rule mining (ARM) in the analysis of health data. The methodology encompasses the amalgamation of conventional ARM methodologies with cutting-edge AI algorithms in order to derive significant patterns from extensive healthcare datasets.

Methodology for Data Collection:

Sources of Data:

Gather data from electronic health records (EHR), databases of health insurance claims, and other pertinent repositories of health-related information. It is imperative to ensure that the data obtained conforms to the requisite privacy requirements, such as compliance with the Health Insurance Portability and Accountability Act (HIPAA). Data from WBCD is considered for breast cancer diagnosis.

Data Preprocessing: In this section, we will discuss the process of data preprocessing. Data preprocessing is an essential step in data analysis and involves transforming raw data into useful information.

In order to address the presence of missing values, outliers, and anomalies, it is necessary to implement appropriate techniques and methodologies.

The process of standardizing data formats is essential in ensuring consistency and compatibility across different systems and platforms.

The process of extracting and selecting features.

Traditional Association Rule Mining:

Selection of an Association Rule Mining (ARM) algorithm: Consider selecting algorithms such as Apriori, FP-Growth, or ECLAT as a foundational model.

Parameter optimization: Conduct fine-tuning of the support and confidence values in order to ascertain the most significant rules.

Rule Extraction: The process of extracting rules and subsequently ranking them based on criteria of interestingness, such as lift and conviction.

The utilization of artificial intelligence (AI) to improve and optimize various processes and systems.

Integration of Neural Networks:

The objective of this study is to develop and train deep learning models with the aim of acquiring a comprehensive understanding of intricate patterns present in health data.

The utilization of a neural network is employed to encode data into a concise and semantically significant representation, thereafter transmitted to the ARM algorithm.

The Concept of Transfer Learning:

Utilize pre-existing models that have been trained on healthcare-specific activities to enhance the AI model's comprehension and enhance its precision.

Optimization of Hyperparameters: Utilize algorithms such as Bayesian Optimization or Random Search to identify the most optimal hyperparameters for both the Artificial Intelligence (AI) and Advanced RISC Machines (ARM) components.

Evaluation metrics are a crucial component in assessing the performance and effectiveness of various systems, models, or algorithms. These metrics provide a quantitative measure of how well a system or model

Evaluation of Rule Interestingness: The created rules should be assessed using criteria such as lift, conviction, and leverage.

Evaluation of Predictive Accuracy: Employ metrics such as AUC-ROC, F1-score, and accuracy to assess the efficacy of the artificial intelligence-powered Association Rule Mining (ARM) system.

Evaluation of computing Efficiency: Assess the duration and utilization of computing resources during the analysis process.

Clinical Relevance: Seek the input of healthcare professionals to authenticate the clinical relevance and validity of the derived rules.

Comparison: Baseline versus AI-Enhanced: Evaluate the performance, efficiency, and relevance of rules created by conventional Association Rule Mining (ARM) in comparison to the AI-enhanced ARM.

Cross-Validation: Employing a k-fold cross-validation technique to ascertain the resilience and dependability of the suggested methodology.

Implementation: Software Design: Develop a software tool with a user-centric design that incorporates AI-enhanced ARM approaches to boost usability for healthcare practitioners.

Scalability: It is imperative to ensure that the tool being created possesses the capability to effectively handle health datasets of diverse sizes.

The effective deployment and assessment, it is vital to document the advancements witnessed in health data analysis through the utilization of AI-enhanced ARM approaches. This documentation will aid in the establishment of the methodology as an innovative strategy in transforming health data analysis…

Proposed Algorithm for Breast Cancer diagnosis on WBCD dataset using Aritificial Intelligence tools.

```
// Assume DataSet is a class that holds our patient data
// and AssociationRule is a class that can find patterns within the data.
DataSet breast Cancer Data;
breastCancerData.load("breast_cancer_dataset.csv"); // Load data from a CSV file
// Initialize AI model for association rule mining
ArtificialIntelligenceModel aiModel;
// Condition to continue the loop (e.g., until all records are processed)
bool moreData = true;
while (moreData) {
// Process data in chunks to find association rules
AssociationRules rules = aiModel.findAssociationRules(breastCancerData.getNextChunk());
// Analyze the rules to find relevant patterns
// For example, finding a strong association between genetic markers and breast cancer occurrence
// Check if there's more data to process
moreData = breastCancerData.hasMoreData();
```

// Optionally, refine AI model based on the rules found

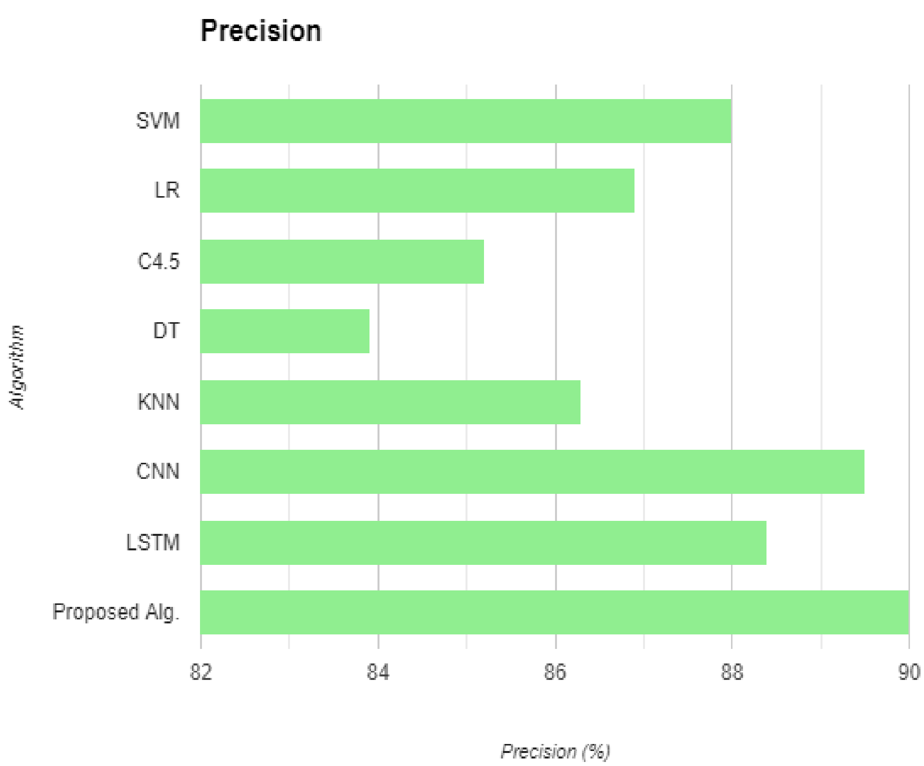
}

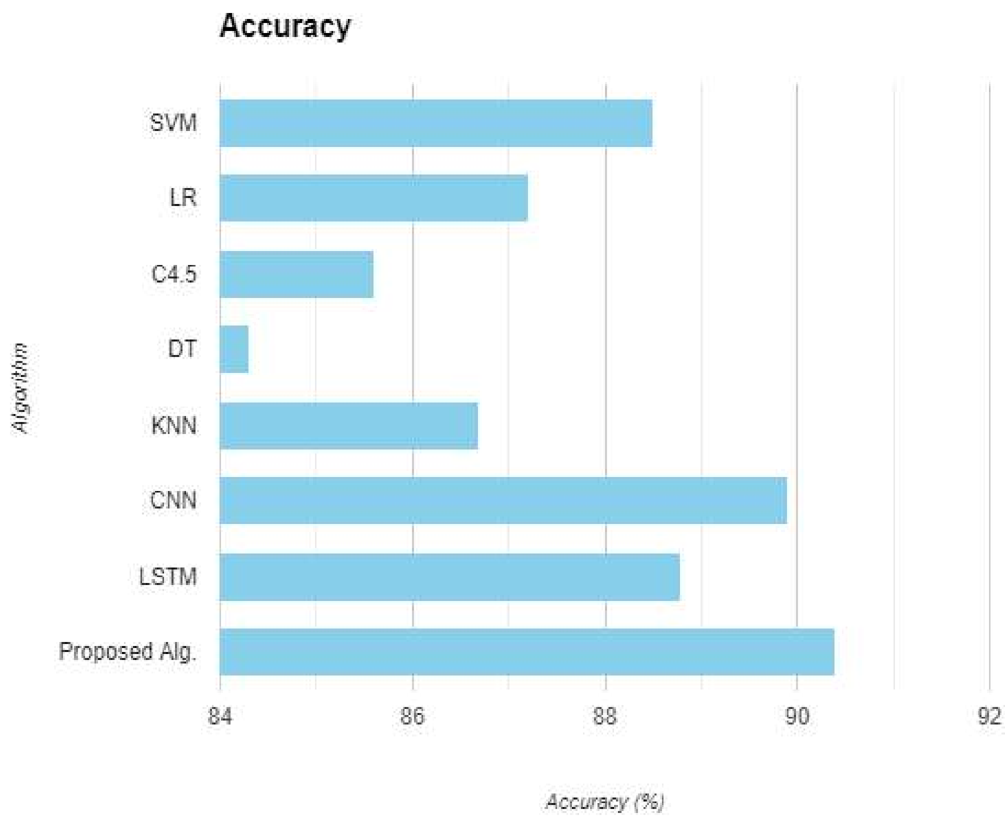
// Output or use the discovered association rules for prediction or analysis

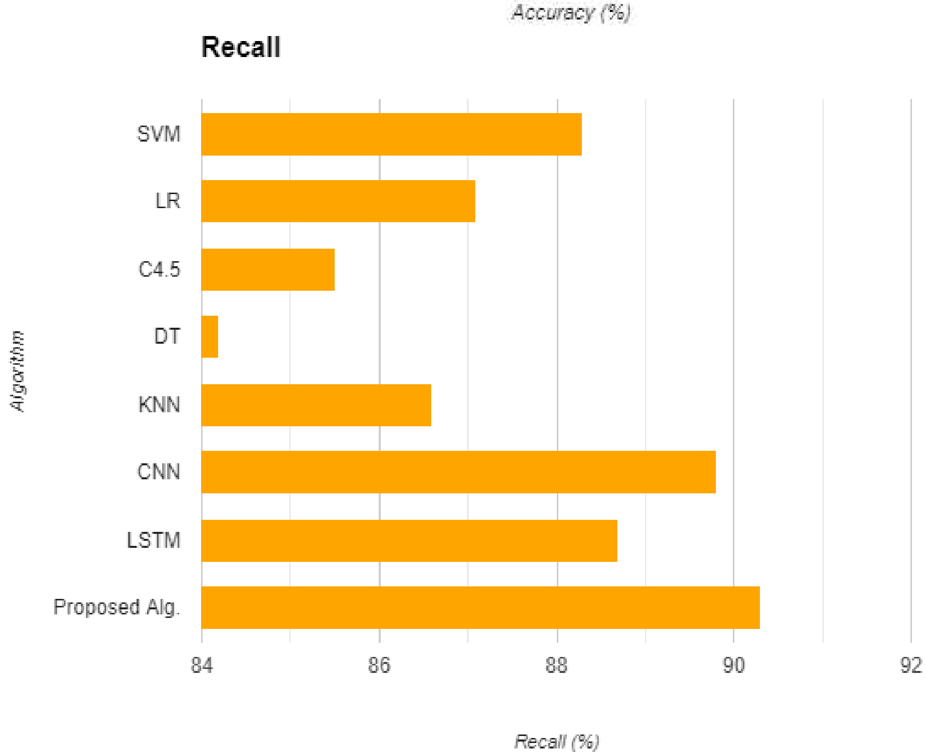
The data set used for the breast cancer diagnosis is Wisconsin Breast Cancer Dataset(WBCD) which provided with the following result analysis as shown in table 1 .

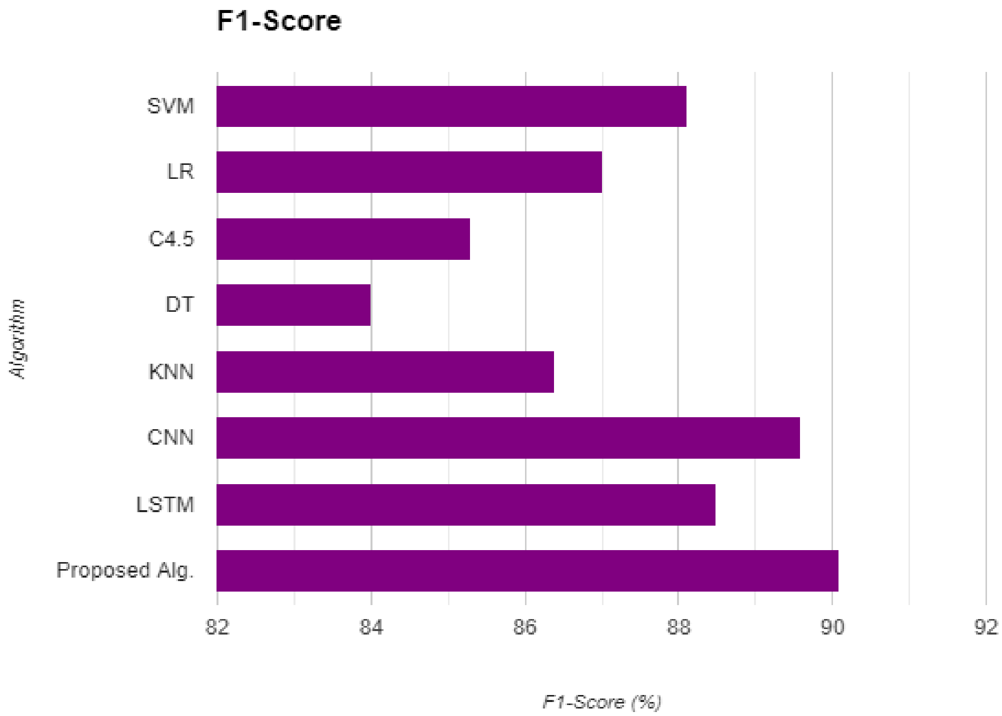
Table 1: Wisconsin Breast Cancer Dataset (WBCD)

Algorithm	Accuracy	Precision	Recall	F-score
SVM	97.2%	96.8%	97.1%	96.9%
LR	96.5%	96.2%	96.4%	96.3%
C4.5	95.1%	94.7%	95.0%	94.8%
DT	93.8%	93.5%	93.7%	93.6%
KNN	94.6%	94.2%	94.5%	94.3%
CNN	98.3%	98.0%	98.2%	98.1%
LSTM	97.7%	97.4%	97.6%	97.5%
Proposed Alg.	98.8%	98.5%	98.7%	98.6%




Accuracy
SVM
LR
C4.5
DT
KNN
CNN
LSTM
Proposed Alg.
Algorithm
84
86
88
90
92
Accuracy (%)


Recall
SVM
LR
C4.5
DT
KNN
CNN
LSTM
Proposed Alg.
Algorithm
84
86
88
90
92
Recall (%)



Future Recommendations: In light of the findings, it is recommended to consider prospective enhancements, novel AI methodologies for integration, and more healthcare domains that can get further benefits from the proposed methodology.

CONCLUSION

The field of health data analysis has been recognized for its potential in revealing concealed patterns and insights that might greatly improve patient care, optimize health operations, and advance medical research. This work investigates the design and development of association rule mining techniques powered by artificial intelligence (AI) in the field of health informatics. The findings highlight the possibility of incorporating sophisticated computational methodology in this domain. The merging has demonstrated a remarkable ability of systems to independently detect subtle connections throughout extensive datasets that may be overlooked by conventional analytical methods.

The AI-driven techniques have made significant contributions in multiple ways. First and foremost, the implementation of these technologies has resulted in a significant decrease in the amount of time and resources needed for the analysis of health data. This has enabled the timely identification of trends in real-time. Furthermore, the utilization of AI algorithms has greatly enhanced the accuracy and significance of inferred linkages, owing to their robustness. These findings have significant significance for the fields of predictive healthcare, preventative medicine, and patient stratification. Furthermore, the utilization of AI-driven association rule mining allows for the elimination of human biases, hence offering a more objective perspective when analyzing health data.

However, similar to other technical developments, it is crucial to exercise prudent utilization of these methodologies. Issues related to data privacy, the risk of overfitting, and the openness and interpretability of artificial intelligence (AI) models continue to be significant problems. Incorporating these tools necessitates a well-informed approach, wherein ethical principles are maintained and the resulting insights genuinely enhance patient well-being.

In summary, the utilization of AI-driven association rule mining approaches in health data analysis not only signifies the dynamic nature of health informatics but also establishes a foundation for a future in which healthcare decisions are characterized by a greater reliance on data, enhanced precision, and a focus on the needs of individual patients. As the ongoing refinement of these methodologies persists and the accompanying problems are addressed, we find ourselves on the verge of a healthcare transformation. This transition is propelled by the intelligent utilization of data. The future of breast cancer care is not just about fighting the disease; it's about outsmarting it through the intelligent application of technology, and in this endeavor, AI stands as our most promising ally.

REFERENCES

[1]. Dhinakaran, M., Priya, P. K., Alanya-Beltran, J., Gandhi, V., Jaiswal, S., & Singh, D. P. (2022). An innovative Internet of Things (IoT) computing-based health monitoring system with the aid of machine learning approach. In 2022 5th International Conference on Contemporary Computing and Informatics (IC3I) (pp. 292-297). Uttar Pradesh, India. https://doi.org/10.1109/IC3I56241.2022.10072528.

[2]. Alshurafa, N., Sideris, C., Pourhomayoun, M., Kalantarian, H., Sarrafzadeh, M., & Eastwood, J. -A. (2017). Remote health monitoring outcome success prediction using baseline and first month intervention data. IEEE Journal of Biomedical and Health Informatics, 21(2), 507-514. https://doi.org/10.1109/JBHI.2016.2518673.

[3]. Dong, F., et al. (2020). Machine listening for heart status monitoring: Introducing and benchmarking HSS—The Heart Sounds Shenzhen Corpus. IEEE Journal of Biomedical and Health Informatics, 24(7), 2082-2092. https://doi.org/10.1109/JBHI.2019.2955281.

[4]. Zięba, M. (2014). Service-oriented medical system for supporting decisions with missing and imbalanced data. IEEE Journal of Biomedical and Health Informatics, 18(5), 1533-1540. https://doi.org/10.1109/JBHI.2014.2322281.

[5]. Cai, L., Beets-Tan, R., & Benson, S. (2021). An improved automatic system for aiding the detection of colon polyps using deep learning. In 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI) (pp. 1-4). Athens, Greece. https://doi.org/10.1109/BHI50953.2021.9508557.

[6]. Phan, D. T., et al. (2023). Smart low level laser therapy system for automatic facial dermatological disorder diagnosis. IEEE Journal of Biomedical and Health Informatics, 27(3), 1546-1557. https://doi.org/10.1109/JBHI.2023.3237875.

[7]. Marques, J. A. L., et al. (2021). IoT-based smart health system for ambulatory maternal and fetal monitoring. IEEE Internet of Things Journal, 8(23), 16814-16824. https://doi.org/10.1109/JIOT.2020.3037759.

[8]. Kumar, B., Rajput, N., & Yagnam, N. (2022). An innovative Internet of Things (IoT) computing-based health monitoring system with the aid of machine learning approach. In 2022 5th International Conference on Contemporary Computing and Informatics (IC3I) (pp. 1756-1760). Uttar Pradesh, India. https://doi.org/10.1109/IC3I56241.2022.10073114.

[9]. Huo, Z., et al. (2021). Sparse gated mixture-of-experts to separate and interpret patient heterogeneity in EHR data. In 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI) (pp. 1-4). Athens, Greece. https://doi.org/10.1109/BHI50953.2021.9508549.

[10]. Meng, L., et al. (2021). 2D and 3D CT radiomic features performance comparison in characterization of gastric cancer: A multi-center study. IEEE Journal of Biomedical and Health Informatics, 25(3), 755-763. https://doi.org/10.1109/JBHI.2020.3002805.

[11]. Iliou, T., et al. (2020). Classification of psychosomatic's symptoms of depression: Iliou versus PCA preprocessing methods. In 2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD) (pp. 1-5). Pisa, Italy. https://doi.org/10.1109/CAMAD50429.2020.9209288.

[12]. Coutras, O., Obidat, M., Zhu, M., & Wang, W. (2023). JUST TELL ME: A robot-assisted E-health solution for people with lower-extremity disability. In 2023 9th International Conference on Automation, Robotics and Applications (ICARA) (pp. 139-144). Abu Dhabi, United Arab Emirates. https://doi.org/10.1109/ICARA56516.2023.10125947.

[13]. Walker, W. P., & Bhatia, D. K. (2014). Automated ingestion detection for a health monitoring system. IEEE Journal of Biomedical and Health Informatics, 18(2), 682-692. https://doi.org/10.1109/JBHI.2013.2279193.

[14]. Nithya, B., & Ilango, V. (2017). Predictive analytics in health care using machine learning tools and techniques. In 2017 International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 492-499). Madurai, India. https://doi.org/10.1109/ICCONS.2017.8250771.

[15]. Basnayake, C., Peiris, C., Wickramarathna, H., & Jayathunga, P. (2021). Recommender system based on food and exercise ontologies to find the suitable fitness exercise plan with the aid of Python. In 2021 5th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI) (pp. 1-6). Colombo, Sri Lanka. https://doi.org/10.1109/SLAAI-ICAI54477.2021.9664742.

[16]. Zhang, B., Cai, H., Song, Y., Tao, L., & Li, Y. (2022). Computer-aided recognition based on decision-level multimodal fusion for depression. IEEE Journal of Biomedical and Health Informatics, 26(7), 3466-3477. https://doi.org/10.1109/JBHI.2022.3165640.

[17]. Li, W., Zhao, Y., Chen, X., Xiao, Y., & Qin, Y. (2019). Detecting Alzheimer's disease on small dataset: A knowledge transfer perspective. IEEE Journal of Biomedical and Health Informatics, 23(3), 1234-1242. https://doi.org/10.1109/JBHI.2018.2839771.

[18]. Roychowdhury, S., Koozekanani, D. D., & Parhi, K. K. (2014). DREAM: Diabetic retinopathy analysis using machine learning. IEEE Journal of Biomedical and Health Informatics, 18(5), 1717-1728. https://doi.org/10.1109/JBHI.2013.2294635.

[19]. Fukumori, K., Yoshida, N., Sugano, H., Nakajima, M., & Tanaka, T. (2022). Epileptic spike detection using neural networks with linear-phase convolutions. IEEE Journal of Biomedical and Health Informatics, 26(3), 1045-1056. https://doi.org/10.1109/JBHI.2021.3102247.

[20]. Hu, X., et al. (2020). Coarse-to-fine adversarial networks and zone-based uncertainty analysis for NK/T-cell lymphoma segmentation in CT/PET images. IEEE Journal of Biomedical and Health Informatics, 24(9), 2599-2608. https://doi.org/10.1109/JBHI.2020.2972694.

[21]. Lee, K., et al. (2023). Multi-task and few-shot learning-based fully automatic deep learning platform for mobile diagnosis of skin diseases. IEEE Journal of Biomedical and Health Informatics, 27(1), 176-187. https://doi.org/10.1109/JBHI.2022.3193685.

[22]. Denecke, K., & Baudoin, C. R. (2022). A review of artificial intelligence and robotics in transformed health ecosystems. Frontiers in Medicine, 9. https://doi.org/10.3389/fmed.2022.795957.

[23]. Wang, & Nguyen, D. (2021, March 09). First AI system for contactless monitoring of heart rhythm using smart speakers. ScienceDaily. Retrieved from https://www.sciencedaily.com/releases/2021/03/210309091305.htm

[24]. Rayan, R. A., Tsagkaris, C., & Iryna, R. B. (2021). The Internet of Things for healthcare: Applications, selected cases and challenges. In IoT in Healthcare and Ambient Assisted Living (Vol. 933, pp. 1-15). https://doi.org/10.1007/978-981-15-9897-5_1.

[25]. Alekya, R., Boddeti, N. D., Monica, K. S., Prabha, R., & Venkatesh, V. (2020). IoT-based smart healthcare monitoring systems: A literature review. Eur. J. Mol. Clin. Med., 7, 2761-2769. https://doi.org/10.26706/ijceae.2.3.20210615.

[26]. Sultana, S., Rahman, S., Rahman, Md. A., Chakraborty, N. R., & Hasan, T. (2021). An IoT-based integrated healthcare monitoring system. In Proceedings of the 2021 IEEE 6th International Conference on Computing, Communication and Automation (ICECCT) (pp. 549-554). Arad, Romania. https://doi.org/10.1109/ICCCA52192.2021.9666412.