Nandhakumar. P. V.[1],

Dr. M. Elamparithi[2],

Dr. V. Anuratha[3]

A Comprehensive Survey on Early Detection and Monitoring of Disease Using Machine Learning

Journal of Electrical Systems

Abstract- Over recent years, forecasted analytics within medical care has assumed a critical role in life-saving interventions. In medical care, super smart tech that figures out tricky connections in data and turns them into useful stuff for making forecasts is advancing really fast. Consequently, artificial intelligence is effecting rapid transformations within the medical care industry. Consequently, there has been the advent of diagnostic and prognostic systems predicated upon ML and DL methodologies, leveraging both clinical datasets and medical imagery for disease identification and anticipation. These systems are super helpful clinically. In the medical care sector, forecasted analytics is a crucial requirement. This has the potential to markedly enhance the precision of disease forecasting, thereby facilitating life-saving interventions contingent upon prompt and accurate prognostication; Conversely, erroneous forecasting has the potential to imperil patient lives. So basically, we need to check out diseases super accurately and forecast as well as we can. Therefore, the deployment of dependable and efficacious methodologies for forecasted analytics within medical care contexts is of paramount importance. Thus, the purpose of this study is to provide a thorough overview of the current ML and DL techniques used in medical care forecasting and to pinpoint the inherent challenges associated with using these techniques in the medical care industry.

Keywords- earlier prediction, machine learning, data driven, deep learning, disease prognosis

1. Introduction

Human existence changes daily, but the aggregate health status of successive generations exhibits a trajectory of either progressive improvement or incremental decline. Life is full of uncertainties. Sometimes run into tons of people dealing with super serious, life-threatening health problems as a result of diseases being discovered too late [1]. Over 50 million adults worldwide would be impacted by chronic liver disease. However, the illness can be prevented if it is identified early. Leveraging machine learning methodologies for disease prognostication facilitates the early identification of prevalent pathological conditions [2]. Right now, people aren't focusing on health as a top priority, and that's causing a bunch of problems. Ignoring recurrent long-term discomfort might have detrimental effects on one's health, even though many people are too busy or have limited time to visit a doctor, and others cannot afford to do so. Medical personnel and researchers are working nonstop to lower disease-related mortality since infections are a worldwide issue [3]. Due to the growing amount of health information from a variety of disparate and conflicting sources, in the medical field, predictive analytic patterns have grown in significance in recent years.

The endeavour of society as a whole to guarantee, supply, finance, and advance health is known as medical care. The ideal of well-being and the avoidance of illness and disability underwent a dramatic change in the twentieth century [4]. The provision of medical care services involves coordinated public or private initiatives to help people restore their health and avoid illness and disability. Standardised guidelines that aid in assessing behaviours or circumstances that influence decision-making can be used to characterise health care. The medical care system is multifaceted. Diagnosing and treating diseases or disabilities is the fundamental objective of health care [5]. Physicians and nurses, medical facilities, and a financing organisation to support the organisation are the first two essential elements for support of a medical care system. The massive volume of data generated by medical care services on a daily basis makes it more difficult to handle and evaluate in traditional methods [6]. This data can be appropriately processed to produce useful insights using deep computing with automated learning. Medical

[1]Research Scholar,
[2,3]Associate Professor,
Department of Computer Science, Kamalam College of Arts and Science, Anthiyur, Affiliated to Bharathiar University, Coimbatore, Tamil Nadu, India.
pvnandhakumar175@gmail.com, profelamparithi@gmail.com, profanuratha@gmail.com.

care information can also be supplemented by genomes, a range of data sources, including data from multiple domains, including environment, social media, healthcare, and more. Prognosis, treatment, treatment, and medical procedures are the four key medical deployment domains that can profit from leveraging machine learning for improvement, as explained in the section that follows. The increasing interest in forecasted analytics methods to enhance medical care is demonstrated by a commitment to advancing innovation through sustained investment in cutting-edge technologies on ML and DL approaches to improve individual health through the forecasting of future events [7]. In the past, clinical forecasting algorithms helped identify those who were more likely to get sick. Based on specific patient characteristics, these algorithms for forecasting are used to advise clients and make medical decisions. Medical analytics that are forecasted are critical business needs [8]. It can significantly affect how accurately diseases are forecasted, perhaps saving lives when done correctly and on time, but can potentially jeopardize patients' lives when done incorrectly. As a result, infections need to be accurately estimated and anticipated [9]. As a result, for medical care forecasted analysis, reliable and effective methods are required. This article aims to provide a thorough analysis of popular ML and DL techniques used in medical care forecasting, as well as to recognise the inherent difficulties in applying these strategies in the medical care industry.

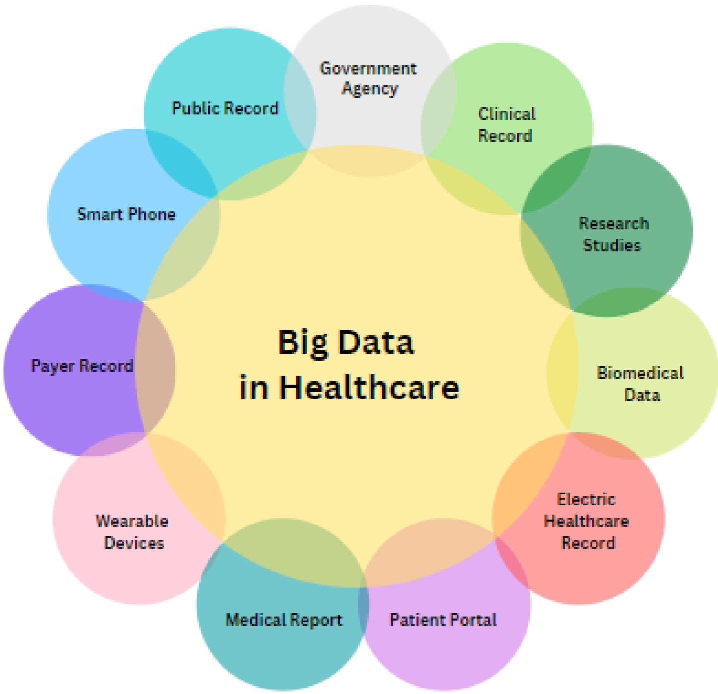

Fig 1. Big data sources from healthcare record

2. Background analysis

Machine learning (ML) is an area of this type of AI whose objective is to develop algorithms that learn from data patterns and improve from data, and make predictions without human intervention. ML employs statistical and computational methods for identifying relationships in data and building predictive tools to handle those relationships [10]. ML is commonly connected with extracting insights from data, identifying patterns, and applying advanced learning techniques. These domains tend to blend without sharp boundaries, and they sometimes overlap. Deep learning is widely regarded as a relatively new branch of machine learning that relies heavily on computational methods and enormous volumes of uncover intricate patterns and relationships in data. This section illustrates the classification of ML algorithms into three distinct categories: Learning with labelled data, learning without labels, and learning through trial-and-error [11].

2.1. Disease forecasting analytics with machine learning

The research selected for this study is concerned with the deployment of ML and DL algorithms in health care forecasting. Our review includes empirical research and literature reviews related to the subject. To forecast patients diagnosed with diabetes, the researchers used a framework method for developing and evaluating machine learning classification models. The PIDD was used to implement the machine learning method. The model achieves an accuracy rate of 83% in its forecasts. The way the implementation is carried out results show that logistic regression outperformed other ML algorithms. Furthermore, only structured data was considered, while unstructured data was not taken into account. The pattern should also be deployed in various healthcare settings, including heart disease management and COVID-19 care. Further variables, including diabetes family history, tobacco use, and lack of physical activity, should be considered for diabetes forecasting. The researchers developed a diabetes diagnosis system that employs two information sets and four forecasting patterns. The accuracy rate for the SVM algorithm was 83%. Some components of this work should be improved, such as utilizing a deep learning approach to forecast techniques for diabetes [12], which could lead to better outcomes; moreover, the pattern could be tried in other clinical areas, such as datasets for applications like heart disease management and COVID-19 forecasting information sets.

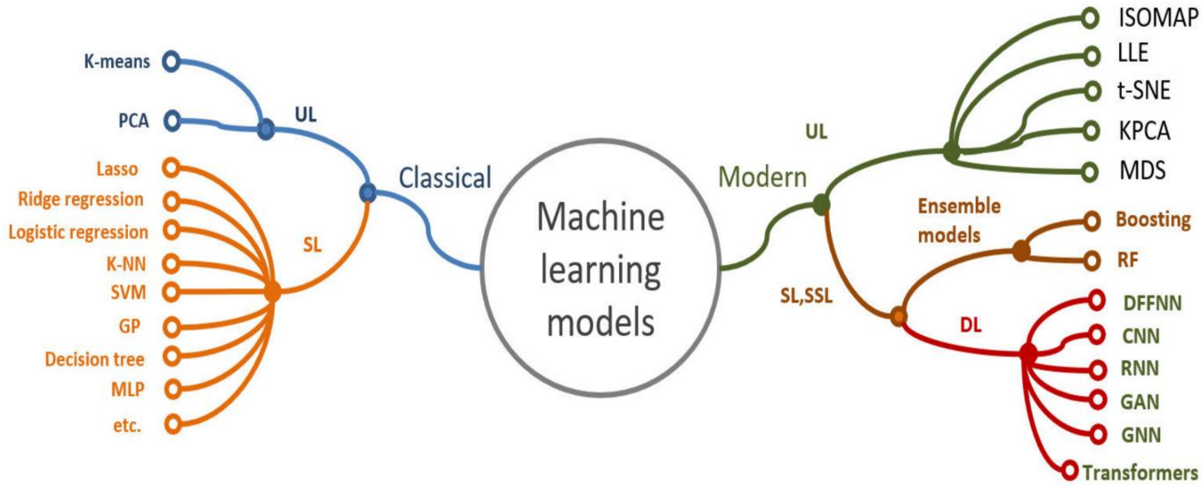

Fig 2. ML models and its evolutions

The researchers suggested three machine learning approaches to evaluate COVID-19 utilizing publicly available datasets from Mexico and Brazil. The proposed model predicts rescue emergency response and death outcomes based solely on COVID-19 patients' geographical, social, and economic factors, and geographical determinants, as well as patient-specific data, including clinical risk factors, patients' medical history and risk profile, and demographics. The model's performance on the Mexico dataset yielded 93% correctness and an F1 score of 79%. In contrast, the Brazilian model achieved 69% correctness and an F1 value of 75% on the dataset. The three ML frameworks were tested, and the findings revealed that LR proved to be the most suitable technique for processing the information [13]. The researchers should be worried about the use of identity verification and the management of private data of the generated information. The researchers suggested a new pattern for forecasting and applying network approaches and machine learning to study type 2 diabetes. To forecast the likelihood of type 2 diabetes, medical care information for a total sample of records of 1,028 type 2 diabetes patients and 1,028 non-diabetic individuals was extracted controls were obtained from de-identified data sources [14]. The patterns prove effective based on the experimental results, with an AUC ranging from 0.79 to 0.91. The RF pattern demonstrates superior correctness compared to the other models. This research depends solely on a dataset of inpatient admission and release information from patients insured by one company. Hospital visits outside the network and third-party insurance information are unavailable to consumers with multiple insurance carriers. The researchers suggested a medical care management system for patients can use to make doctor visits and confirm medications. It supports machine learning in diagnosing illnesses and choosing medications. Health datasets covering diabetes, heart disease, chronic kidney disease, and liver disorders, among other chronic conditions, are subjected to machine learning patterns [14]. According to the findings, the heart dataset results showed that logistic regression achieved the highest accuracy of 98.5% out of all the other patterns. However, the DT classifier performed the worst with the lowest accuracy, coming in at 92%. Among all the logistic regression patterns, the model performed best on the liver dataset with an accuracy of 75.17%. The LR, RF, and Gaussian NB all showed strong performance in the chronic renal illness dataset, with a correctness of 1. The dependability of the patterns should be tested using k-fold CV to confirm the accuracy of 100%. Random Forest achieved a maximum correctness of 84% in the diabetes dataset. A hospital directory should be included by the writers so that different clinics and hospitals may be accessed via a single gateway. Furthermore, visual information sets could be added to enable the use of DL for illness identification and visual processing of reports [15].

Table 1: Comparison of various existing approaches with research findings

Disease	Adopted techniques		Highest accuracy
	ML	**DL**	
Liver	Algorithms like Logistic Regression, K-NN, DT, SVM, NB, and RF.	--	LR outperformed other patterns with an accuracy of 75%.
Diabetes	Machine learning patterns including LR, K-NN, SVM, RF, NB, DT, and ensemble methods.	The patterns used included ANN, CNNs, DNN, LSTM, Conv LSTM, NNs, along with SVR and ARX, with parameter optimization techniques applied.	The deep learning model demonstrated outstanding performance with 98% accuracy.
Heart	Various machine learning patterns were explored, including LR, KNN, RF, DT, NB, SVM, and K-means clustering.	The study employed Neural Networks, J4.8 DT, NB, CSO-LSTM, and CNN.	CSO-LSTM outperformed other patterns with an accuracy of 96.16%.
Multiple Disease Detection	The study employed DT, RF, LR, and NB algorithms.	DNN, MLPs	LR outperformed other patterns with a 98.5% accuracy rate on the heart dataset.
COVID-19	The study employed DT, RF, LR, NB, and K-means algorithms.	Various deep learning patterns were used, including CNN, CSO-LSTM, RNN, MobileNetV2, ResNetV2, VGG19, DenseNet201, InceptionV3, Xception, and RNN-based LSTMs.	LR achieved an impressive 98.5% accuracy, surpassing other patterns.

The diabetes dataset was classified by the researchers who employed SVM and NB algorithms, utilizing relevant feature selection. Analysis was conducted on the PIDD dataset retrieved from the UCI Repository. The model evaluation utilized k-fold CV as implemented by the researchers for training and testing; the SVM classifier outperformed the NB method, producing accurate forecasting in approximately 91% of cases; however, the researchers acknowledge the limitation of their dataset and suggest extending it to more recent data, which will include more data attributes and records. The K-means algorithm is an unsupervised technique for clustering. It groups similar data points into meaningful groups introduced by the researchers to detect early stages of heart disease utilizing the UCI dataset for heart disease analysis. One method for feature reduction is PCA [16]. The method's results show a 94.06% accuracy rate for early heart disease forecasting. It is recommended that the researchers employ multiple algorithms and datasets while implementing the suggested method. Logistic regression was used as the classification method; the researchers created a diabetes categorisation model based on forecasted analytics. There are 128 cases for testing and 459 patients for training in this dataset. Logistic regression had a forecasting accuracy of 92%. The primary drawback of this study is that it cannot be verified because the researchers did not compare the pattern with alternative classification algorithms for diabetes detection [17].

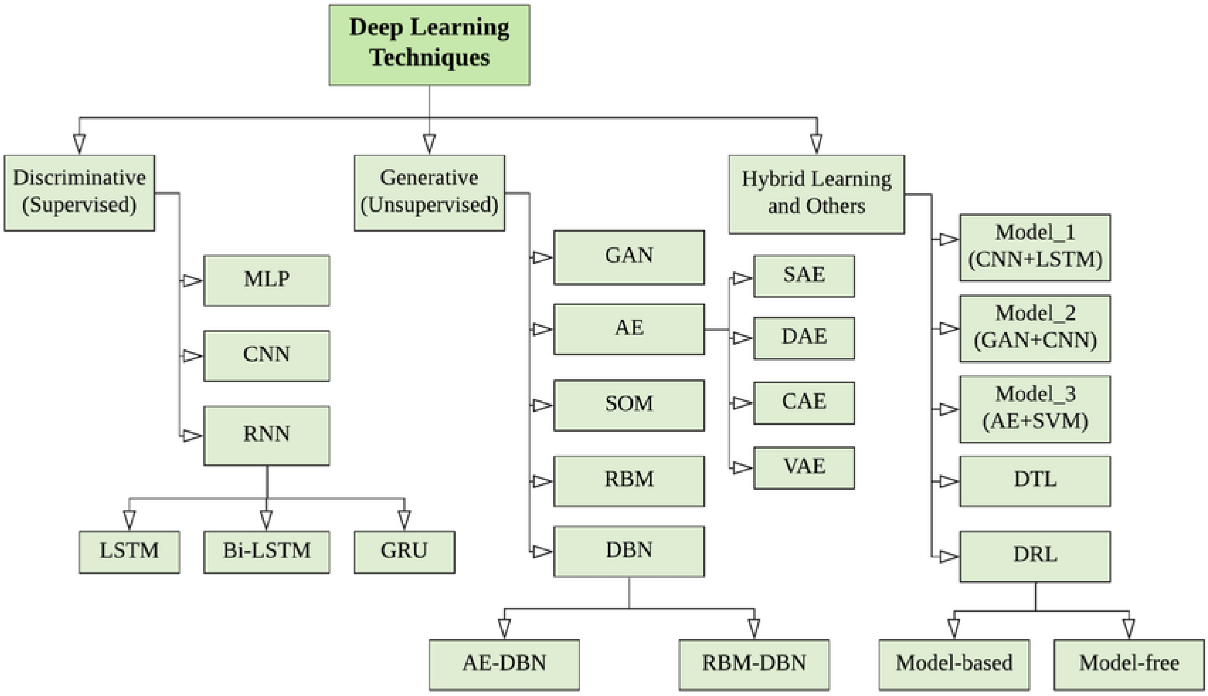

Fig 3. DL approaches and its workflow

Employing ML algorithms, the researchers created a forecasting pattern that examines the user's symptoms and predicts the illness. By enabling medical practitioners to anticipate diseases and illnesses in their early stages, this study aimed to address health-related issues. The dataset included 4,920 a sample of 4920 patient records, analysed for 41 identified health conditions. All ailments make up the dataset. 41 diseases served in all were used as dependent variables. The algorithms achieved an accuracy score of 95% was attained by every algorithm. When the full set of all 132 symptoms from the original dataset was included, instead of being evaluated, rather than just 95 symptoms, the scientists observed that significant overfitting took place. In other words, the tree fails to classify fresh data because it seems to remember the dataset it was trained and supplied. Consequently, just 95 symptoms were evaluated during the data-cleaning process involved evaluating 95 symptoms, and the most relevant ones were selected. The researchers developed a decision-making system that provides will provide automated forecasts forecasting's on the state of the patient's heart health and helps practitioners forecast cardiac problems with precise classification using a simplified approach. They implemented four methods are K-NN, RF, DT, and NB that were utilised in the Cleveland Heart Disease dataset [18]. The accuracy varies depending on the classification technique. When they used the K-NN method, with KNN achieving the highest accuracy, a correlation factor of nearly 94%, they achieved the highest accuracy. The researcher's method should expand this method to predict be expanded by the researchers to forecast other diseases by utilising many datasets. The Cleveland dataset, which comprised 303 cases and 76 attributes, was utilised by the researchers to test three distinct classification strategies: NB, SVM, and DT, in addition to K-NN. Out of the 76 attributes, only 14 were tested [19]. To eliminate noise, the data was pre-processed. K-NN achieved the highest accuracy at 90.79%. To improve, out of these seventy-six traits, only fourteen will be tested. To get rid of noisy data, the writers pre-processed the data. At 90.79 percent, the K-NN had the highest accuracy. To increase the precision of early heart disease forecasting, the researchers must employ more advanced models to identify patterns. Using a cardiovascular dataset that was then categorised using supervised machine learning techniques, the researchers suggested a model to forecast heart disease. The findings showed that the DT classification model shows, with an accuracy of 73%, the DT classification model outperforms other algorithms in forecasting cardiovascular diseases.

The researchers emphasised that combining a better sickness forecasting model can be produced by using ensemble ML techniques with the CVD dataset, could produce a more accurate disease prediction model [20].

Table 2. Comparison with various ML approaches, datasets, research findings and methodology

Disease	Approaches	Dataset	Findings	Methodology
Diabetes	Logistic regression, KNN, SVM, and RF	PIDD	Logistic regression outperformed other patterns with 83% accuracy; incorporating additional factors like family history, smoking habits, and physical activity could further enhance diabetes forecasting.	Machine learning framework for diabetes forecasting model development and evaluation
	RF, SVM, NB, and DT	Data sourced from Frankfurt Hospital in Germany and the PIDD dataset.	SVM emerged as the top-performing model with 83% accuracy.	Diabetes forecasting system using machine learning.
	Logistic regression, SVM, NB, KNN, DT, RF.	A large dataset of 18.7 million hospital admission records (1995-2018) from 124,000 de-identified patients in the Australian CBHS health funds dataset.	DL techniques have the potential to enhance diabetes forecasting accuracy.	Type 2 diabetes forecasting model using current year data to forecast next year's risk.
	A hybrid approach combining LR, RF, and SVM.	Diabetes forecasting through machine learning techniques	Among the patterns tested, RF attained an achieved the highest accuracy of 84.95%, the highest among the models.	A popular choice for binary classification problems like diabetes forecasting.
COVID-19	CNN	Two publicly available datasets from the US National Institutes of Health (NIH).	The performance of non-Haar wavelet functions warrants further examination.	CNNs with wavelet transforms for COVID-19 data analysis under constraints.
	CNN	2482 CT images from patients, publicly available for classification studies.	With 96.16% accuracy, the CNN model demonstrated strong performance.	CNN-based approach for identifying COVID-19 cases.
	CNN and RNN	A dataset of 657 chest X-ray images used to train for training a	VGG19 outperformed other patterns,	Detecting diseases in patients flagged as

		deep learning model for detecting to diagnose COVID-19.	achieving the highest success rate.	COVID-19 suspects based on X-rays.
	CNN	X-ray dataset: A dataset consisting of 100 images from 70 COVID-19 patients and 1431 images from 1008 non-COVID pneumonia patients' cases.	Model performance: 90% sensitivity, 87.84% specificity vs. 96% sensitivity, 70.65% specificity.	Introducing a deep learning-based anomaly detection framework.
	A range of CNN patterns, including MobileNetV2, ResNetV2, VGG19, DenseNet201, and InceptionV3.	A compact dataset consisting of 50 photos.	VGG19 and DenseNet demonstrated strong performance with F1f1-scores attained were of 89% and 91%, respectively. While InceptionV3 struggled at 67%.	The COVIDX-Net framework for automatic detection of COVID-19 from chest radiographs X-rays.
	CNN	A medical imaging dataset with 408 COVID-19, 4273 pneumonia, and 1590 healthy X-ray images was utilized for classification. A total of 6271 people are represented in the dataset.	The researchers may face limitations in generalizing these results, particularly when adopting such a model in real-world settings.	A novel approach for primary COVID-19 screening detection based on radiological analysis of chest radiographs.
Heart disease	K-NN, RF, DT, and NB	Cleveland Heart Disease dataset	KNN demonstrated exceptional performance, reaching an accuracy of 94%.	A decision support system for classification tasks using automated forecasting.
	Classifiers like NB, DT, KNN, SVM	Cleveland dataset	KNN demonstrated exceptional performance, reaching an accuracy of 90.79%.	Forecasting risk factors for heart disease with a machine learning model.
	Machine learning algorithms, including DT, NB, LR, RF,	Cardiovascular dataset	With 73% accuracy, the model demonstrated the effectiveness of ensemble machine learning approaches.	Applying machine learning techniques for cardiovascular risk forecasting.

	SVM, and KNN			
	Using Scikit-Learn's logistic regression in machine learning	The Framingham Heart Study dataset includes variables such as age, gender, sex, chest pain, slope, and cardiovascular outcome.	Achieved 87% accuracy, but further optimization is needed to reduce time complexity.	A framework for boosting forecasting accuracy in machine learning patterns.

The researchers used a medical care dataset to use a logistic regression to identify the presence or absence of heart sickness in patients in an effort to boost accuracy in cardiovascular disease forecasting. The findings came from an ongoing landmark cardiovascular research project tracking residents of Framingham, Massachusetts. The pattern's forecasting accuracy was 87% [21]. The researchers acknowledge that additional data and methods would be beneficial machine learning patterns could enhance the pattern. To analyse the breast cancer dataset consists of 569 observations for breast cancer analysis columns, the author proposed a precise algorithm for breast cancer classification, motivated by the significant impact of the disease on Indian women [22]. In a similar vein, this study provided a novel approach to function selection using datasets related to lung cancer and heart disease. This selection process is based on a combination of SVM classification and genetic algorithms. The findings of the classifier are 78% for diabetes and 82% for lung cancer observed that there is no indication of the type, size, or source of the data used. The article sought to describe the benefits of utilising heart disease survival forecasting patterns with proven track records and a range of data mining techniques [23]. Based on the observations, the researchers suggested that DT and RF yield superior outcomes for datasets with few features, whereas logistic regression and NB yielded the best results when applied to a technique to a high-dimensional dataset. Because RF is an optimised learning method, it provides more precision than the DT classifier [24]. The pattern might be constructed in leveraging technologies like Map-Reduce, Apache Mahout, and HBase, according to the author, and this work could be expanded to additional forecasted modelling methods. To determine the likelihood of being diagnosed with developing type 2 diabetes, the researchers developed a forecasted analytics framework for chronic disease proven effective in the Australian medical care setting using DM and network analysis techniques [25]. Using a year's Australian private medical care fund data, analysed with three forecasted modelling approaches (regression, parameter optimisation, and DT). The forecasting is accurate between 82 and 87 percent of the time. The information set is based on the discharge and admission summaries for hospital patients. Consequently, it excludes data on visits to general physicians and subsequent diagnoses.

3. Research challenges

Notwithstanding the considerable progress achieved in ML and DL in recent years, they require further advancement to address the fundamental challenges affecting medical care systems effectively. This article discusses some of the issues related to adopting ML and DL algorithms for medical care forecasted modelling. Biomedical data streams pose a significant issue that requires attention [26]. A substantial volume of new medical data is emerging at a quick pace, and the medical care industry as a whole is changing swiftly. Health indicators such as blood pressure, oxygen levels, and glucose are important real-time biological indicators. Even though numerous DL architectural variations have tried to solve this issue, there are still a lot of obstacles to overcome before efficient analyses of quickly changing, enormous volumes of broadcasting information can be carried out. Content delivery includes issues with computational complexity, memory usage optimization, feature selection, and missing data imputation [27]. The intricacy of the medical care industry presents another difficulty for ML and DL. Compared to other fields, biomedical research and medical care pose more complex issues. We still don't fully understand the causes, spread, and treatments of many of these remarkably varied illnesses. Because there aren't always enough patients, it is challenging to get enough data. However, there might be a way to resolve this problem [28]. Due to the tiny patient population, thorough patient characterisation, creative information processing, and the inclusion of new information sets are required. To extract patient data, each dataset can be processed separately using tailored deep learning techniques, with findings combined into a comprehensive model [29] – [30].

4. Conclusion

The delivery of traditional medical care services may be altered by the deployment of ML and DL methodologies for forecasted analytics in medical care. Medical care data is thought to be the most important element that supports the integration of ML and DL technologies within medical care delivery systems. The goal of this research is to conduct a comprehensive examination of prominent ML and DL methodologies used in medical care forecasting analytics. It also covered the difficulties and barriers of using ML and DL algorithms in the medical field. Additionally, each methodology was well addressed. According to the studies analysed, AI approaches play an important role in effectively identifying illness as well as anticipating and analysing medical care data aggregated through the linkage of extensive clinical datasets and reconstructing a client's history utilising this data. This study improves the investigation of the deployment of ML and DL methodologies in medical care forecasted analytics, as well as contributing to a useful starting point for academics and researchers looking to explore this topic further.

References

[1] Teo ZL, Jin L, Liu N, Li S, Miao D, Zhang X, Ting DSW. Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture. Cell Rep Med 2024; 5(2).

[2] Lee HY, Yoon YJ, Choi YJ, Ham YH. Factors associated with Korean American women’s Health-Related internet use: findings from Andersen’s behavioral model. J Immigr Minor Health. 2024; 26(1):124–32.

[3] Babitsch B, Gohl D, Von Lengerke T. Re-revisiting Andersen’s behavioral model of health services use: a systematic review of studies from 1998–2011. GMS Psycho Soc Med. 2012; 9:1-15.

[4] Genç G, Kurutkan MN. Eşitsizlik Bağlamında Karşılanmayan Sağlık İhtiyacı: Türkiye Sağlık Araştırması Verilerinden Kanıtlar. SDÜ Sağlık Yönetimi Dergisi. 2021; 3(1):34–51.

[5] Tsoni R, Kaldis V, Kapogianni I, Sakagianni A, Feretzakis G, Verykios VS. A machine learning pipeline using KNIME to predict hospital admission in the MIMIC-IV database. In: 2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA). IEEE. 2023; 1–6.

[6] Zhang Y, Kreif N, Gc VS, Manca A. Machine learning methods to estimate individualized treatment effects for use in health technology assessment. Med Decis Making. 2024; 44(7):756–69.

[7] Moulaei K, Shanbehzadeh M, Mohammadi-Taghiabad Z, Kazemi-Arpanahi H. Comparing machine learning algorithms for predicting COVID-19 mortality. BMC Med Inf Decis Mak 2022;22(2).

[8] Samuel O, Zewotir T, North D. Application of machine learning methods for predicting under-five mortality: analysis of Nigerian demographic health survey 2018 dataset. BMC Med Inf Decis Mak. 2024;24(1):86.

[9] Sidey-Gibbons JA, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol. 2019;19:1–18.

[10] Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. Medicina. 2020;56(9):455.

[11] Tesfaye SH, Seboka BT, Sisay D. (2024). Application of machine learning methods for predicting childhood anaemia: analysis of Ethiopian demographic health survey of 2016. PLoS ONE 19(4):e0300172.

[12] Feretzakis G, Karlis G, Loupelis E, Kalles D, Chatzikyriakou R, Trakas N, Kaldis V. Using machine learning techniques to predict hospital admission at the emergency department. J Crit Care Med. 2022;8(2):107–16.

[13] Nusinovici S, Tham YC, Yan MYC, Ting DSW, Li J, Sabanayagam C, Cheng CY. Logistic regression was as good as machine learning for predicting major chronic diseases. J Clin Epidemiol. 2020;122:56–69.

[14] Bansal M, Goyal A, Choudhary A. A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. Decis Analytics J. 2022;3:100071.

[15] Begum N, Rahman MM, Omar Faruk M. (2024). Machine learning prediction of nutritional status among pregnant women in Bangladesh: evidence from Bangladesh demographic and health survey 2017–18. PLoS ONE, 19(5):e0304389.

[19] Kim NH, Kim M, Han JS, Sohn H, Oh B, Lee JW, Ahn S. Machine-learning model for predicting depression in second-hand smokers in cross-sectional data using the Korea National health and nutrition examination survey. Digit Health. 2024a;10:20552076241257046.

[20] Kalayou MH, Kassaw AAK, Shiferaw KB. Empowering child health: Harnessing machine learning to predict acute respiratory infections in Ethiopian under-fives using demographic and health survey insights. BMC Infect Dis. 2024;24(1):338.

[21] Alie MS, Negesse Y. Machine learning prediction of adolescent HIV testing services in Ethiopia. Front Public Health. 2024;12:1341279.

[22] Kim J, Jeong K, Lee S, Baek Y. Machine-learning model predicting quality of life using multifaceted lifestyles in middle-aged South Korean adults: a crosssectional study. BMC Public Health. 2024b;24(1):159.

[23] Carmona P, Climent F, Momparler A. Predicting failure in the U.S. Banking sector: an extreme gradient boosting approach. Int Rev Econ Finance. 2019;61:304–23.

[24] Asselman A, Khaldi M, Aammou S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interact Learn Environ. 2023;31(6):3360–79.

[25] Turjo EA, Rahman MH. Assessing risk factors for malnutrition among women in Bangladesh and forecasting malnutrition using machine learning approaches. BMC Nutr. 2024;10(1):22.

[26] Yin C, Cao J, Sun B. Examining non-linear associations between population density and waist-hip ratio: an application of gradient boosting decision trees. Cities. 2020;107:102899.

[27] Peretz O, Koren M, Koren O. Naive Bayes classifier–An ensemble procedure for recall and precision enrichment. Eng Appl Artif Intell. 2024;136:108972.

[28] Rachmat R, Yusuf M, Basmar MF, Suherwin S. Analysis of algorithms and data processing efficiency in movie recommendation systems based on machine learning. Jurnal Mandiri IT. 2024;13(2):273–9.

[29] Placiszewski K, Wierzba W, Ostrowski J, Pinkas J, Jankowski M. (2023). Factors associated with the use of public eHealth services in Poland-a 2022 nationwide cross-sectional survey. Ann Agric Environ Med 30(1).

[30] Ye BZ, Wang XY, Wang YF, Liu NN, Xie M, Gao X, Liang Y. Impact of tobacco smoking on health care utilization and medical costs in chronic obstructive pulmonary disease, coronary heart disease and diabetes. Curr Med Sci. 2022;42(2):304–16.