Paper ID: 159

A Review of Methods Used for Health Prediction and Monitoring Utilizing Electronic Medical Records (EMR) System

SP Jayasekara[1#], and LP Kalansooriya[2]

[1,2]*Faculty of Computing, General Sir John Kotelawala Defence University, Sri Lanka*
[#] 37-cs-0009@kdu.ac.lk

Abstract— *With the global population steadily increasing, there is a growing demand for electronic medical records due to the substantial amount of information generated within hospitals. Handling these records physically can prove challenging. Electronic medical records (EMRs) have had a profound impact on the healthcare sector by digitizing hospital records, thereby enhancing patient care. By enabling electronic entry, maintenance, and storage of medical data over extended periods, EMRs contribute to improved patient care and safety. This review examines and compares various methods and techniques aimed at diagnosing and predicting diseases accurately through the use of EMRs. Additionally, it presents a comparative analysis of different approaches available for health prediction. Recent publications were studied to categorize these techniques into Deep Learning (DL) Methods, Machine Learning (ML) Methods, and Rule-Based Methods. Moreover, the review outlines the advantages and disadvantages associated with these diverse techniques and discusses their impact on the healthcare industry.*

Keywords— **Healthcare, Deep learning, Electronic Medical Records (EMR), Rule-based method, Disease diagnosis, Machine learning**

I. INTRODUCTION

One of the most critical responsibilities of medical institutions is managing patient data, and a patient file is an essential source of data since it enables the development of comprehensive healthcare strategies. It had been common practice for a long time to keep records on paper where medical offices, hospitals, and clinics frequently gathered files and kept patient history using a paper record system. However, paper medical records have a lot of drawbacks such as insufficient storage space, insufficient backups, inconsistency in the layout, and unclear audit trails. Due to technological advancements, electronic medical records were introduced to store patient data on computers or smart devices and overcome paper records' drawbacks.

EMRs are digitalized versions of paper charts in clinics and hospitals. Clinicians and doctors primarily use these EMRs to diagnose and treat patients and record information by and for the physicians in the hospital. It contains a patient's medical history, diagnoses, prescriptions, treatment schedules, vaccination dates, and lab and test results. These are stored in databases that enable doctors or clinicians to access patient information quickly, track vaccinations, follow patient health performance, and make informed judgments with proper understanding and confidence for the most complex multi-axial diseases, heart diseases, and cancers [4].

By computerizing patient information, there is also a significant change in how patient data are arranged and made available for applications that weren't previously possible with paper records. Thus, it shows that the main objective of an EMR is keeping an eye on the patient while improving healthcare quality. Even though EMR and EHR provide users and physicians with several advantages, several difficulties are connected to their implementation, such as computer downtime, computer professionals' limitations, a lack of user communication, security risks of confidentiality-leakage, etc which should be considered [6].

An accurate and timely diagnosis is the foundation of any successful treatment. Access to longitudinal data from a patient's EMR might be a valuable clinical resource that could be utilized to forecast future events or diagnoses [1]. A patient's status is thoroughly described in an EMR, and applying data-driven technologies to an EMR enables us to accurately predict and diagnose diseases. This can be made possible by making the raw EMR data into a machine learning representation or turning the data into meaningful information that can be algorithmically processed. It is a vital step in the predictive modeling of EMR data. There are different types of data-driven techniques used to accomplish prediction and diagnosis systems that medical professionals can employ to effectively forecast illnesses and enhance the health of their patients. This review aims to find the most accurate methods for diagnosing and predicting diseases by describing and comparing various methods and techniques used for health prediction and monitoring using electronic medical records (EMR).

This study discusses numerous disease diagnosis and prediction methods using electronic records, highlighting their benefits and drawbacks. It also discusses current trends and potential future developments and does a comparative comparison of the various methods.

The literature review of this paper explores the significance of Electronic Medical Records (EMR) data in monitoring patient health and advancing data-driven decision-making. It delves into the growing interest in employing computer-assisted methods for disease diagnostics based on EHR data, categorizing these methods into distinct approaches. Machine Learning (ML) methods, encompassing Bayesian, Support Vector Machine (SVM), and decision tree techniques, are discussed, along with the challenges of integrating raw EHR data into ML models due to complexity and limited healthcare data. Bayesian Networks are highlighted for their use in probabilistic medical ontology reasoning, aiding in disease diagnosis and prediction. Decision Trees are emphasized for their effectiveness in early identification of diseases like Diabetic Retinopathy and asthma. Additionally, rule-based heuristic techniques are explored for diagnosing colorectal cancer and lupus. Finally, Deep Learning methods, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Deep Belief Networks (DBN), and Autoencoders (AE).

The paper is structured into five sections. Section 2 discusses the current research on methods for illness diagnosis in Section 3, Methodology. Section 4 contains the discussion. Finally, Section 5 presents the conclusion of the review.

II. DISEASE DIAGNOSIS USING DATA DRIVEN MODELS

EHR data offers a practical way to monitor patient health data and improve decisions using data-driven technologies [8]. In contrast to clinical tests and other biological investigations, secondary data gathered from EHRs aren't aimed at proving a particular theory; their main objective is to track a patient's health over time. Researchers have created several methods for illness diagnostics with computer assistance.

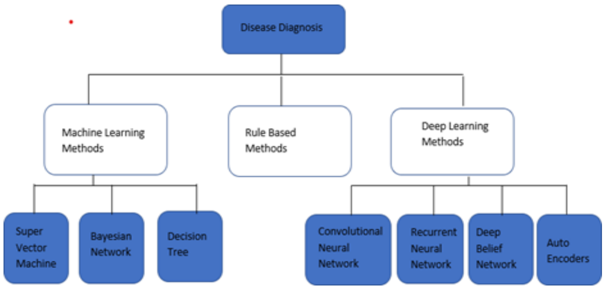

Figure 1. Breakdown of the techniques used in this review to diagnose diseases.

This paper discusses various electronic medical record-based methods for diagnosing diseases automatically. Depending on their technique, models have been grouped into different approaches to diagnosing diseases.

A. Machine Learning Methods

Health database systems based on electronic medical records (EMR) are most often created using machine learning methods for individuals who have had health examinations [7]. Machine learning methods can be categorized into different approaches, including Bayesian, Support Vector Machine (SVM), and decision tree methods. Each of these approaches represents a distinct category within the field of machine learning [34].

Many research studies have used EHR data for predictive modeling, which involves constructing a statistical model to predict a clinical outcome using machine learning. However, it is difficult to directly integrate raw EHR data into ML models for predictive modeling due to the complexity of EHR data [8]. Because a lack of data prevents machine learning from solving many healthcare issues.

i. Super Vector Machine (SVM)

For supervised classification, experts often use Support Vector Machines (SVM). SVM is based on labeled data, and Vapnik invented SVM [45]. A training dataset is used to find data from the input that has a structure similar to the output data when both the input and the output have already been supplied.

Getting a cancer diagnosis is crucial for prospective patients since early tumor identification and therapy can improve survival. In [9], a cancer diagnosis was performed using the SVM model using medical information retrieved straight from the EHR. As part of the proposed approach, SVM models for cancer classification were trained using medical records extracted from Electronic Health Records (EHRs). These SVM models played a significant role in the classification process of cancer based on the analyzed medical data. After being trained on 400 pieces of data for each cancer and employing 100 pieces of health information for each cancer, the algorithm has shown a predictive accuracy of 86.2% for ten different forms of cancer and 97.33% for three different types of cancer.

An SVM-based technique was used [10] for significant cohort research to diagnose contralateral breast cancer. Characteristics based on pathology reports for every area of breast cancer and narrative text in progress notes were used to derive features, Zeng et al. [10] designed and put into practice a novel methodology. The suggested strategy for identifying contralateral occurrences in the notes uses medical ideas and how they are combined. SVM and derived characteristics are used to detect contralateral cancers. During the validation process, the area under the curve (AUC) for the model was determined to be 0.93, indicating its high accuracy in predicting outcomes. In the test set, the AUC was slightly lower at 0.89, indicating a slightly reduced but still reliable performance. This strategy of feature development is advantageous due to its simplicity and can be applied to different occurrences of breast cancer as well as to identify various other diseases.

In order to identify Rheumatoid Arthritis (RA) patients, a computer learning technique called Support Vector Machine (SVM) has been employed. This technique utilizes a set of naïve and expert-defined Electronic Health Record (EHR) characteristics for the identification process [12]. This method uses Natural language processing (NLP) concepts, pharmaceutical exposures, and billing codes. The SVM methodology was trained using both expert-defined and naive data. The accuracy and recall scores were 0.94 and 0.87, respectively, as opposed to 0.75 and 0.51 for deterministic approaches. In this study, a dataset of 10,000 patients was employed. The test findings divided the patients into three groups: potential RA, definite RA, and not RA.

ii. Bayesian Network (BN)

A probabilistic graphical framework called a Bayesian network is utilized to represent a group of variables and their conditional interactions. This graphical model employs a directed acyclic graph (DAG) to illustrate the relationships among the variables and their dependencies. Naive Bayes (NB) and Bayesian Networks (BN) are both probabilistic algorithms that perform effectively with various characteristics [14].

Building Clinical Bayesian Networks (CBN) for probabilistic medical ontologies reasoning is described in [13] to directly learn the entire ontology and high-quality Bayesian topology from EMRs. More than 10,000 patient records analyzed for medical entity connections have used the K2 greedy method and Odds Ratio (OR value) computation to create a Bayesian topology automatically. The study demonstrates that medical information can generate high- quality health topology and ontology directly and automatically. A clinical Bayesian network has been developed using the study's probability distribution between illness and other parameters. With 1712 test samples, an accuracy of 64.83% was produced by the Naïve Bayesian network, while the Basic Bayesian network produced 68.45%.

In a study by Sakai et al. (15), they evaluated the diagnostic performance of a Bayesian network in comparison to the NB model, an artificial neural network (ANN), and a logistic regression model in order to identify instances of appendicitis. 169 people who were thought to have acute appendicitis were included in the dataset for the study. The performance of the proposed model was assessed using logistic regression and neural network metrics. Compared to other diagnostic models examined in this research, this model had the lowest error rate and produced the most trustworthy findings, detecting that 50.9% of patients (86 out of 169) had appendicitis.

The Naïve Bayes method was employed in Al-Aidaroos et al. [16] review of medical data mining to classify medical data and diagnoses such as primary tumors, hepatic issues, and breast or lung cancer. Using 15 datasets, the proposed NB strategy was empirically compared with five other approaches to show its superiority. The findings indicated that NB performed better than others regarding medical categorization. Deep learning ideas can produce superior segmentation results with the proposed approach. The report states that future research will combine NB and different methodologies.

Kazmierska and Malicki researched the Bayesian classifier, which is used to assess whether cancer is progressing or relapsing [17]. This study analyzed data from 142 individuals who had radiation therapy for brain tumors between 2000 and 2005. For training, 96 binary attributes were selected. As a result of the proposed model, the likelihood of having a cancer relapse has been determined as well as the likelihood of not having one. The proposed method received scores of 0.84, 0.87, and 0.80 for accuracy, specificity, and sensitivity, respectively.

iii. Decision Tree

EHR data can accelerate and simplify the early identification of Diabetic Retinopathy (DR). Five machine- learning techniques are used in [18] to identify diabetic retinopathy using electronic health record data. Records from 301 Chinese hospitals were compiled into a sizable retinal dataset. To increase the accuracy of DR illness diagnosis, preprocessing techniques such as label binarization, value normalization, and standard acceleration are carried out. According to the experimental findings, the machine learning model's Random Forest (RF) can achieve an accuracy level of 92% while performing well. Due to its low cost, low threshold, and excellent diagnosis accuracy, the suggested approach has an advantage over current DR diagnostic methods.

The primary objective of the study conducted by Lungu et al. [19] was to investigate whether machine learning techniques could enhance the diagnostic precision of Magnetic Resonance Imaging (MRI) in detecting pulmonary hypertension (PH). This was accomplished by employing computational modeling approaches and image-based metrics. MRI as well as the Right Heart Catheterization (RHC) were used to identify PH using a decision tree method [19]. Seventy-two individuals with potential PH underwent MRI and RHC, and 57 of these patients were found to have the condition, while 15 samples were determined to be PH- free. As a result of the proposed algorithm, 92% of the PH cases were correctly identified, while 4% were misclassified. If the findings of this study are as anticipated, RHC may not be required when PH is suspected.

In [20], the decision tree is used in the first phase to diagnose asthma, and the fuzzy system is utilized in the second phase to assess the level of asthma management. Dry cough, sore throat, sneezing, and other symptoms have been used to diagnose asthma, whereas breathlessness and other daytime symptoms have been used to measure the control level. In this study, the information was gathered through the patients' responses to questionnaires. Diagnoses of asthmatic patients were made using a decision tree classifier, which had accuracy and kappa coefficients of 0.90 and 0.783, respectively.

B. Rule Based Method

In [22], the diagnosis of colorectal cancer was made using a rule-based heuristic technique. Machine learning and rule- based methods' effectiveness was evaluated for each phase. The algorithm identified concepts at the document level with an F-measure of 0.996 as well as detected cases at the patient level with 0.93 for the F-measure using the manually examined data set of 300 potential Colorectal cancer patients. In the work by Breischneider et al. [23], in this study, rule- based grammar was used to obtain textual information from records of patients with mamma carcinoma. Based on recovered textual fragments, seven essential criteria were listed to construct the therapeutic suggestion. The mammography use case was used to assess the proposed system. With an accuracy of 0.69, a textual feature extraction approach based on rule-based decision support, information extraction, and semantic modeling was employed to determine the lymph node status.

In an EHR dataset with 400 records, Jorge et al. [24] used rule-based approaches to identify lupus patients. Natural language processing was used to extract the narrative and codified data from the training set of data (NLP). Based on penalized logistic regression, the author classified systemic lupus erythematosus (SLE) as either definite or probable. The machine learning code utilized in this work for definite SLE showed a 90% positive predictive value, with a specificity of 97%. According to the best rule-based method (ICD-9 code), the specificity and sensitivity were respectively 86% and 84% and 60 % and 69 % for definite and definite/probable SLE.

C. Deep Learning Methods

Deep neural networks, including autoencoders (AE), Convolutional Neural Networks (CNN), Deep Belief Networks (DBN), Recurrent Neural Networks (RNN), and other similar architectures, are considered the most effective machine learning techniques in the biomedical sector [25]. These networks form the foundation of deep learning and have shown remarkable effectiveness in various biomedical applications. Various deep learning methods used on electronic medical records are examined in this review to apply them to clinical tasks. Their benefits are discussed in practice and potential future applications.

i. Convolutional Neural Networks

A method for unsupervised deep feature learning that Miotto et al. introduced in [21]. Using clinical notes as the input, they drove patient representation in their predictive modeling technique. By identifying hierarchical regularities and relationships in clinical notes, 700,000 individuals from the Mount Sinai dataset were used. The study encompassed a broad range of clinical areas and chronological periods, involving a total of 76,214 test individuals, representing 78 distinct diseases. The study's findings surpassed approaches that relied on a representation derived from basic medical information, where accurate and F-score forecasts improved by 92.9% and 18.1%, respectively. When produced patient representations are included in DL approaches, clinical prediction can be improved. This study can use the laboratory findings to improve the quality of its model.

Multiple illnesses have been evaluated using the disease prediction model built on EMRs [26]. The Convolutional Neural Network (CNN) has been used to characterize the suggested strategy for multiple illness prediction. This approach was tested on 4298 patients with a brain infection, coronary heart disease, and pulmonary infection. In a dataset for cerebral infections, the CNN algorithm, the accuracy was 96.5% and the F1-measure score was 96.6%.

ii. Recurrent Neural Networks

Recurrent Neural Networks (RNN) were specifically designed to process sequential inputs, such as language data. The present state of an RNN implicitly incorporates knowledge about the whole history of the series since RNNs process, a series of inputs that transmits the concealed value of every input unit to the next input unit, one item at a time. Doctor AI [27], which was over eight years, performed over 260K time-stamped analyses on individuals' electronic health records longitudinally, which is one RNN-inspiring technique. Doctor AI surpassed multiple baselines and scored 79.58% on a sizable real-world EHR dataset.

Wu et al. [28] presented a novel approach for categorizing pediatric asthma by utilizing event sequences and their corresponding characteristics. The findings of this study show that including a timestamp in an RNN model enhances the categorization of individuals without asthma rather than those who have it.

iii. Deep Belief Network

A Deep Belief Network (DBN) has been used to diagnose Parkinson's disease (PD) using speech sounds collected from the UCI repository [30]. A range of healthy and sick voices was used to train the suggested approach, using DBN as a data source, and the features were extracted. According to the proposed method, the PD consists of one output layer and two stacked limited Boltzmann machines. Parkinson's disease was diagnosed with 94% accuracy using the recommended approach.

DBN has been used [31] to diagnose Attention Deficit Hyperactivity Disorder (ADHD), which is one of the most common diseases. The network was built and trained using a greedy methodology according to the recommended strategy. The Global Competitions ADHD-200 has provided the two training and testing datasets. This study has used samples from the Neuroimaging (NI) and New York University (NYU) databases for training and testing, respectively. These findings show that they attain cutting-edge accuracy of 0.6368 on the NYU dataset and 0.6983 on the NI dataset.

iv. Auto Encoders

In a study, researchers employed auto-encoders to forecast a particular group of diagnoses [29]. For the detection and classification of softmax, stacked autoencoder, and cervical cancer classification algorithms have been utilized [32]. To train and test the approach, the UCI dataset with 30 characteristics, four targets, and 668 samples was used. A training set made up 70% of the dataset, while a test set made up 30%. Four target variables were applied to the suggested model, and the efficacy of its categorization was evaluated. This comparison produced a 0.978 accurate classification rate. Due to the dimensionality reduction of the samples, this model's training takes far too much time. In the future, advanced methods could be used to reduce the training time of the model. Hwang et al. [33] examined the efficacy of

missing value prediction, conventional networks, and generative adversarial networks (GANs) methods combined for illness prediction [33]. With a specificity of 0.99, along with a sensitivity of 0.95, also with an accuracy of 0.98, the stacked autoencoder (missing value forecasting technique) and auxiliary classifier GANs (AC-GANs: illness prediction) have shown excellent results. In this work, AE fills in the gaps left by the GAN generic model. The use of GAN to fill in the missing data is one of this work's future directions.

III. METHODOLOGY

An efficient and effective way to obtain requirements is through document analysis, which involves reviewing current system documentation and acquiring data. In the study, recent research articles were used from Google Scholar and other research archives. Keywords related to the topic were searched to find existing research articles.

Several research articles on EMR systems and methods used for predicting systems were reviewed and analyzed. The research article categorizes the methods used to track and predict health into three categories: Different approaches were employed in the study, including the utilization of Rule-Based Methods, Machine Learning (ML) Methods, and Deep Learning (DL) Methods.

To simplify data analysis, the literature review was summarized into tables. Tables provide an overview of methods used, the disease addressed in the paper, performance measures such as accuracy and F-measures (a measure of test accuracy), and the paper's objectives. Comparing the effectiveness and accuracy of different methods can be done using the tables.

SVM encompasses both nonlinear and linear regression techniques, making it a powerful and advanced technology. It is capable of performing multiclass and binary classification processes, which are essential components of Data Mining using SVM. SVM classifiers are among the most accurate algorithms for predicting future events due to their high classification accuracy. Data mining techniques like SVMs, used for predicting and classifying data, have been used in various fields, including identifying and predicting health.

Researchers commonly use Support Vector Machines (SVMs) for supervised classification. Different types of diseases, such as breast cancer and rheumatoid arthritis, were identified using these SVM methods. With the use of SVM, Zhang et al. [9] cancer classification from EHRs was 97.33% accurate. Using data from 1063 breast cancer patients, Zeng et al. [10] conducted a validation study on a system designed to detect contralateral breast cancer. When utilizing extracted features in combination with pathology reports, the system achieved a high area under the ROC curve (AUC) of 93%. Another evaluation involved the identification of RA phenotypes using Support Vector Machines (SVMs) in a Naïve Electronic Health Record (EHR) with a sample size of 376 patients. The system achieved an F-Measure score of 88.6% [12].

Table 1. Summary of the SVM Method

Methods	Focused disease	Performanc eMeasures	Dataset	Objective
SVM -RBF [9]	Cancer	Accuracy- 97.33%	Employed 100 pieces of health information for each cancer and trained on 400 pieces of data for each cancer.	Using SVMs to classify cancer from EHRs.
SVM [10]	Breast Cancer	Testing- 89% AUC Validation- 93%	A total of 1063 women with breast cancer.	Analyzing pathology reports and extracted features to identify contralateral 1 breast cancer.
SVM [12]	Rheumatoi dArthritis	Precision- 96.8% F-Measure- 88.6% Recall- 87% AUC-96.6%	In total, 376 patients (185 with RA and 191,not RA).	SVM-based phenotyping of RA in the NaïveEHR.

Naïve Bayes (NB) and Bayesian Networks (BN)are probabilistic algorithms that utilize many features in an elegant manner [51]. With Bayesian clustering, we can accommodate patients' varying data availability by incorporating established biomarker operating characteristics. Since EHR data are primarily collected for clinical treatment and health system administration, they lack several desirable characteristics of research data; however, Bayesian joint modeling has the advantage that it incorporates phenotypic uncertainty into future association analyses, yielding accurate uncertainty estimates.

In contrast to Bayesian networks, NB classifiers don't require dependency networks and are better at handling high- dimensional features. This study discusses four papers using Bayesian methods to predict diseases such as cancer, appendicitis, hepatitis, and brain tumors. Shen et al. [13] tested the accuracy of Naive Bayes and Bayesian Networks in predicting cancer and achieved 64.83% and 64.83% accuracy, respectively. By using the Bayesian network, Sakai et al. [15] conducted a diagnostic prediction of acute appendicitis. In a study by Aidaroos et al. [16], Cancer and hepatitis, as well as liver disorders, were all classified using NB with an accuracy of 97.43%. Bayesian networks were also used to optimize treatment decisions for a brain tumor with 84% accuracy.

The table below lists a few Bayesian method-based systems, and research articles are used to note how accurate the results were when using SVM to predict diseases.

Table 2. The Summary of the Bayesian Method

Methods	Focuse d Diseas e(s)	Performanc e Measures	Dataset	Objective
NB, BN [13]	cancer	NB Accuracy- **64.83%** **BN Accuracy- 68.45%**	Records of 10,000 identified patients.	An Automatic Bayesian topology generation using the K2 greedy method and odds ratios (OR values).
Bayesia n Networ k [15]	Appendicitis	-	A database contains 169 people whomay have acute appendicit is.	An algorithm for predicting acute appendicitis using Bayesian networks.
NB, LR, DT, an dNN [16]	Multiple diseases, including cancer, hepatitis, andliver disorders	Accuracy- 97.43% AUC- 99%	Various illnesses are illustrated in 15datasets from the UCI library.	LR, NB, NN, and DT classification ofmedical data.
Bayesia n Networ k [17]	Brain Tumor	The accuracy rate is 84% The sensitivity is 80% The specificity is 87%	142 patients with brain tumors.	Optimization of treatment decisions using the Naïve Bayesian Classifier.

Using decision trees was another method used for predicting diseases in research articles. A decision tree aids in creating a fair picture of the rewards and hazards related to each potential result. When contemplating EHRs, where uncertainty is prevalent, decision trees are highly helpful because they are especially beneficial when the results are unknown. A decision tree is an effective tool for decision- making. It offers a useful framework within which to consider options and investigate what might result from each.

Decision trees are used to categorize records, which are useful for challenges involving association and regression. By using a decision tree, advantages and disadvantages can be quickly visualized and identified. The diagnosis system for diabetic retinopathy developed by Sun and Zhang [18] achieved 86.82% accuracy. Based on MRI images, Lung et al. [19] were able to diagnose pulmonary hypertension with 92% accuracy using a decision tree. Using a decision tree and fuzzy system, asthma diagnosis and control levels were determined[20].

Table 3. The Summary of the Decision Tree Method

Methods	Focused Disease(s)	Performance Measures	Dataset	Objective
Decision Tree[18]	Diab etic Reti nopa thy	Accuracy- 86.6%	301 Chinese hospitals provided5057 records.	Five machine learning techniques are used with the EHR to diagnose DR.
Decision Tree[19]	Pulm onary hyper tensi on	Sensitivity is 97% Accuracy of 92% Specificity- 73%	Pulmonary hypertension i ssuspected in 72 patients.	Analyzing MRI images to diagnosepulmonary hypertension.
Decision Tree [20]	Asthma	Kappa- 78.32% Accuracy- 90%	30 of patients with asthma.	Using fuzzy logic and decision trees to diagnose and control asthma.

By using rule-based systems, we can retrieve features from electronic medical records quickly. For the extraction of data, rule-based systems are used since the most common kind of knowledge representation is if-then logic. Using rule- based systems, domain experts can express and rate their expertise. The decision-making process can then use that data. To determine the outcomes of rule-based or identically based systems, users must input specific attributes or facts, such as patient symptoms. It is difficult for someone without medical training to do this. A drawback of this method is the requirement for precise definitions of data properties.

Using the rule-based method, computer scientists identify rules and identify patterns associated with them. Xu et al.[22] used this method to identify colorectal cancer. Breischneider et al. [23] used automated breast cancer detection using rule-based grammar and achieved 90% accuracy. Using a rule-based algorithm and machine learning codified algorithm, Jorge et al. [24] identified Lupus patients from EMR.

T able 4: The Summary of The Rule-Based Method

Methods	Focused Disease(s)	Performance Measures	Dataset	Objective
Decision Tree[18]	Diabetic Retinopathy	Accuracy- 86.6%	301 Chinese hospitals provided 5057 records.	Five machine learning techniques are used with the EHR to diagnose DR.
Decision Tree[19]	Pulmonary hypertension	Sensitivity is 97% Accuracy of 92% Specificity- 73%	Pulmonary hype rtension is suspected in 72 patients.	Analyzing MRI images to diagnosepulmonary hypertension.
Decision Tree and Fuzzy system [20]	Asthma	Kappa- 78.32% Accuracy- 90%	30 of patients with asthma.	Using fuzzy logic and decision trees to diagnose and control asthma.

Deep learning (DL), also known as hierarchical learning, is a common modeling technique that applies numerous processing layers simultaneously to difficult data. Growing numbers of electronic health records (EHRs) are being analyzed using deep learning (DL). The use of DL on HER is increasing in research studies aimed at predicting individual health trajectories and risks. In deep learning, several types of neural networks are utilized, such as convolutional neural networks, recurrent neural networks, deep belief networks, and autoencoders.

Table 5: The Summary of The Deep Learning Method

Methods	Focused Disease(s)	Performance Measures	Dataset	Objectives
Unsupervised deep feature learning [21]	78 diseases	Accuracy- 92.9% F-score- 18.1%	The Data warehouse from MountSinai contains 700,00 patients.	Predictive models can be developed using patient representations from EHRs.
CNN and Framingham risk score [26]	Cerebral infraction (CI), PulmonaryInfarction (PI), And Coronary Heart(CH)	Accuracy CI-96.5% PI- 95.6% CH- 93.6%	From a Chinese hospital with a grade-A rating, 4298 individuals were evaluated.	Clinical notes based on a uniform modelfor assessing multiple diseases.
RNN[27]	Numerous diseases	Recall- 79.58%	260K patients.	Applied to longitudinally timestampedEHRs.
RNN[28]	Pediatric Asthma	Precision- 84.54% F-measure- 85.08% Recall- 85.65%	4000 patients from Physionet and 4013 patients from OlmstedCountry Birth Cohort.	RNN-based asthma classification inpediatrics.
DBN[30]	Parkinson's Disease	Accuracy- 94%	Data set on 31 Parkinson's patients.	DBN-based Parkinson's diseasediagnosis system.
DBN withgreedy Approach [31]	ADHD	NI- 69.83% Accuracy- NYU-63.68%	Neuroimaging-samples of 73 New York University-samples of263.	A greedy approach to the diagnosis ofADHD using DBN.
Stacked AE and Softmax classification [32]	Cervical cancer	Accuracy- 97.25%	668 samples from the UCI dataset.	Stack autoencoder and softmax classification for cervical cancerclassification and diagnosis.
Stacked AEand GAN [33]	Breast cancer	Sensitivity- 95.28% Accuracy- 98.05% Specificity- 99.47%	Breast cancer records are available for 569 cases, of which 212 are malignant and 357 are benign.	Generative Adversarial Networks(GAN) and stacked autoencoders for disease prediction from EHRs.

IV. DISCUSSION

Prescriptions are now kept in an electronic (digital) format known as EMR thanks to technological advancements. These digital records let doctors and other medical professionals quickly access patient information, remember when patients require checkups and vaccines, and keep track of patients' health performance. Modeling EMR patient data has undergone a wide range of developments because of the use of various forecasting methodologies in the EMR.

Researchers have employed a variety of approaches for the same condition and have obtained different levels of accuracy when health-predicting.

This review discusses different techniques for predicting cancer diseases, including SVM, Bayesian networks, rule- based methods, and stacked AE. Using SVM, Zhang et al. [9] classified cancer and achieved an accuracy of 97.33 %, while Zeng et al. [10] identified breast cancer with an accuracy of 93%. Using a Bayesian network, a similar cancer disease could be identified with 64.83% accuracy, while the same disease could be identified using Naive Bayes with 64.83% accuracy. Rule-based grammar was used to detect colorectal cancer [22], which earned an accuracy of 99.6%. Breast cancer was also detected using rule-based grammar [23], which achieved an accuracy of 90%. By combining AE and Softmax, Adem et al. of [32] classified cervical cancer with 97.25 percent accuracy. In this study, stacked AE and GAN [33] were used to predict breast cancer, and the accuracy rate was 98.05%.

To predict Asthma, different methods have been used. The Decision Tree and Fuzzy system [20] were used to diagnose and control asthma levels, and this system showed an accuracy of 90%. Wu et al. [28] used the RNN method to create a pediatric asthma prediction system with an accuracy f- a measure of 85.08%.

Even though there are many predictive models available, most of them are designed to predict single diseases without taking into account the many factors that can affect patients, for example, a cancer prediction system will only consider the symptoms of a patient to predict cancer and will not suggest other diseases based on these symptoms. However, several models have been developed to help identify multiple diseases, and this review discusses these systems. Al-Aidaroo et al. [16] classified and detected multiple diseases, involving hepatitis, cancer, and liver disorders, with an accuracy of 97.43%. With 86% and 84% sensitivity, [24] based on a rule- based algorithm, definite and probable Systemic lupus erythematosus (SLE) were detected. Shi et al [26] is another researcher focused on multiple diseases Cerebral infarction (CI), Pulmonary Infarction (PI), and Coronary Heart (CH) detected in this system, accuracy reached for each disease was CI 96.5%, PI 95.6%, CH 93.6%. Another system that is used to detect multiple diseases [21] is used to derive 78 diseases for this dataset taken from the Mount Sinai data warehouse of 7000 patients, this system received a 92.9% accuracy.

Data from the literature study indicates that certain approaches are more effective than others. While certain techniques may be more accurate for some illnesses but less accurate for others.

V. CONCLUSION

According to this review, several EMR system studies have been conducted recently to learn new facts about healthcare using technology. Using various procedures, EMRs provide a lasting record of patient care, reducing vulnerabilities and solving problems in modernized healthcare records.

Physicians can provide better care to patients when they have access to accurate and timely information. EMRs assist physicians in providing safer care, reducing medical errors, and improving the diagnosis of diseases. A competent EHR not only keeps track of patient allergies and medications but also checks for concerns when new medications are administered. An EMR can identify patterns of potentially related adverse outcomes and alert at-risk patients quickly. With the advancement of IT, EMR systems are now widely used to manage medical data and prescribe medication.

Different EMR systems using different techniques are installed and used in various healthcare facilities and these EMR systems have proven essential to delivering better patient care. In this review, it is classified into three primary categories machine learning, rule-based approach, and deep learning method which are then further subdivided depending on the suggested algorithm and have attempted to cover the most recent and current studies on autonomous diagnosis from electronic data. As discussed throughout the review, some methods can give accurate results in one type of disease, but not in another, and most systems are designed to predict and diagnose one specific disease, but very few systems have been able to detect multiple diseases simultaneously. According to the literature study, certain approaches were more effective than others.

Although EMR systems have their benefits, there are still some drawbacks, such as the need to update patient records after every appointment or consultation. Otherwise, physicians or clinical supervisors may later check the system and find incorrect information resulting in an inappropriate treatment plan. It is also possible that records may not be updated or inaccessible for an extended period if there is a power outage, location problems, or another issue. Another disadvantage is that they are still quite expensive.

EHRs will be capable of handling massive amounts of data and complicated clinical test results in the future and eliminate current limitations and develop by using advanced existing methods and techniques to predict diseases more accurately. Related issues such as uncertainty in drawing conclusions and privacy issues will be addressed, and EHRs will come up with the genetic and behavioral data required for accurate prescribing and patient care improvement.

REFERENCES

[1] J. Wu, J. Roy, and W. F. Stewart, “Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches,” Med. Care, vol. 48, no. 6 Suppl, pp. S106-13, 2010.

[2] S. Ford, “Patient-centered Medicine, Transforming the Clinical Method (2nd edition),” Health Expect., vol. 7, no. 2, pp. 181–182, 2004.

[3] M. A. Alkureishi, W. W. Lee, S. Webb, and V. Arora, “Integrating patient-centered electronic health record communication training into resident onboarding: Curriculum development and post- implementation survey among house staff,” JMIR Med. Educ., vol. 4, no. 1, p. e1, 2018.

[4] J. Stausberg, D. Koch, J. Ingenerf, and M. Betzler, “‘Comparing paperbased with electronic patient records: Lessons learned during a study on diagnosis and procedure codes,’’,” J. Amer. Med. Inform. Assoc, vol. 10, no. 5, pp. 470–477, 2003.

[5] G. Makoul, R. H. Curry, and P. C. Tang, “‘The use of electronic medical records: Communication patterns in outpatient encounters,” J. Amer. Med. Inform. Assoc, vol. 8, no. 6, pp. 610–615, 2001.

[6] W. R. Hersh, “‘The electronic medical record: Promises and problems,” J. Amer. Soc. for Inf. Sci, vol. 46, no. 10, pp. 772–776, 1995.

[7] C.-S. Yu, Y.-J. Lin, C.-H. Lin, S.-Y. Lin, J. L. Wu, and S.-S. Chang, “‘Development of an online health care assessment for preventive medicine: A machine learning approach,’’,” J. Med. Internet Res, vol. 22, no. 6, 2020.

[8] Y. Si et al., “Deep representation learning of patient data from Electronic Health Records (EHR): A systematic review,” J. Biomed. Inform., vol. 115, no. 103671, p. 103671, 2021.

[9] X. Zhang, J. Xiao, and F. Gu, “‘Applying support vector machine to electronic health records for cancer classification,” in Proc. Spring Simul. Conf. (SpringSim), 2019, pp. 1–9.

[10] Z. Zeng et al., “‘Contralateral breast cancer event detection using nature language processing,” in Proc. AMIA Annu. Symp, 2017.

[11] M. Jamaluddin and A. D. Wibawa, “Patient diagnosis classification based on electronic medical record using text mining and support vector machine,” in 2021 International Seminar on Application for Technology of Information and Communication (iSemantic), 2021, pp. 243–248.

[12] R. J. Carroll, A. E. Eyler, and J. C. Denny, “‘Naïve electronic health record phenotype identification for rheumatoid arthritis,” Proc. AMIA Annu. Symp, 2011.

[13] Y. Shen et al., “CBN: Constructing a clinical Bayesian network based on data from the electronic medical record,” J. Biomed. Inform., vol. 88, pp. 1–10, 2018.

[14] Q. T. Zeng, S. Goryachev, S. Weiss, M. Sordo, S. N. Murphy, and R. Lazarus, “‘Extracting principal diagnosis, co-morbidity and smoking status for asthma research: Evaluation of a natural language processing system, ’’ BMC Med,” BMC Med. Informat. Decis. Making, vol. 6, no. 1, 2006.

[15] S. Sakai, K. Kobayashi, J. Nakamura, S. Toyabe, and K. Akazawa, “‘Accuracy in the diagnostic prediction of acute appendicitis based on the Bayesian network model, ’’ Methods Inf,” Methods Inf. Med, vol. 46, no. 06, pp. 723–726, 2007.

[16] K. M. Al-Aidaroo, A. A. Bakar, and Z. Othman, “‘Medical data classification with naive bayes approach,’’ Inf,” Technol. J, vol. 11, no. 9, pp. 1166–1174, 2012.

[17] J. Kazmierska and J. Malicki, “‘Application of the Naïve Bayesian Classifier to optimize treatment decisions,” Radiotherapy Oncol, vol. 86, no. 2, pp. 211–216, 2008.

[18] Y. Sun and D. Zhang, “‘Diagnosis and analysis of diabetic retinopathy based on electronic health records,” IEEE Access, vol. 7, pp. 86115– 86120, 2019.

[19] A. Lungu, A. J. Swift, D. Capener, D. Kiely, R. Hose, and J. M. Wild, “‘Diagnosis of pulmonary hypertension from magnetic resonance imaging-based computational models and decision tree analysis,” Pulmonary Circulat, vol. 6, no. 2, pp. 181–190, 2016.

[20] A. Tyagi and P. Singh, “‘Asthma diagnosis and level of control using decision tree and fuzzy system,’,” Int. J. Biomed. Eng. Technol, vol. 16, no. 2, pp. 169–181, 2014.

[21] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, “‘Deep patient: An unsupervised representation to predict the future of patients from the electronic health records,’’,” Sci. Rep, vol. 6, no. 1, 2016.

[22] H. Xu et al., “Extracting and integrating data from entire electronic health records for detecting colorectal cancer cases,” AMIA Annu. Symp. Proc., vol. 2011, pp. 1564–1572, 2011.

[23] C. Breischneider, S. Zillner, M. Hammon, P. Gass, and D. Sonntag, “Automatic extraction of breast cancer information from clinical reports,’’ in Proc,” IEEE 30th Int. Symp. Comput.-Based Med. Syst. (CBMS), pp. 213–218, 2017.

[24] A. Jorge et al., “‘Identifying lupus patients in electronic health records: Development and validation of machine learning algorithms and application of rule-based algorithms,” Seminars in Arthritis and Rheumatism, 2019.

[25] S. Mehrabi et al., “‘Temporal pattern and association discovery of diagnosis codes using deep learning,” in Proc. Int. Conf. Healthcare Informat, 2015, pp. 408–416.

[26] X. Shi et al., “‘Multiple disease risk assessment with uniform model based on medical clinical notes,” IEEE Access, vol. 4, pp. 7074–7083, 2016.

[27] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, “Doctor AI: Predicting clinical events via recurrent neural networks,” arXiv [cs.LG], 2015.

[28] S. Wu et al., “‘Modeling asynchronous event sequences with RNNs,’’,” J. Biomed. Informat, vol. 83, pp. 167–177, 2018.

[29] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, “Deep patient: an unsupervised representation to predict the future of patients from the electronic health records,” Sci. Rep, vol. 6, 2016.

[30] A. H. Al-Fatlawi, M. H. Jabardi, and S. H. Ling, "'Efficient diagnosis system for Parkinson's disease using deep belief network,'' in Proc," IEEE Congr. Evol. Comput. (CEC), pp. 1324–1330, 2016.

[31] S. Farzi, S. Kianian, and I. Rastkhadive, "'Diagnosis of attention deficit hyperactivity disorder using deep belief network based on greedy approach," in Proc. 5th Int. Symp. Comput. Bus. Intell. (ISCBI), 2017, pp. 96–99.

[32] K. Adem, S. Kilicarslan, and O. Cömert, "'Classification and diagnosis of cervical cancer with softmax classification with stacked autoencoder," Expert Syst. Appl, vol. 115, pp. 557–564, 2019.

[33] U. Hwang, S. Choi, H.-B. Lee, and S. Yoon, "Adversarial training for disease prediction from electronic health records with missing data," arXiv [cs.LG], 2017.

[34] J. Latif, C. Xiao, S. Tu, S. U. Rehman, A. Imran, and A. Bilal, "Implementation and use of disease diagnosis systems for electronic medical records based on machine learning: A complete review," IEEE Access, vol. 8, pp. 150489–150513, 2020