



RESEARCH ARTICLE

OPEN ACCESS

PREDICTIVE ANALYTICS IN HEALTHCARE: A REVIEW OF ARTIFICIAL INTELLIGENCE TECHNIQUES USING EHR DATA

Rammohanreddy Dondeti^{*1}, Gatram Rama Mohan Babu²

¹Department of Computer Science & Engineering, Acharya Nagarjuna University, University college of Engineering & Technology, Acharya Nagarjuna University, Nagarjuna Sagar, GUNTUR, Andhra Pradesh, India.

²Department of CSE, R.V.R & J.C College of Engineering, Chowdavaram, GUNTUR, India.

¹<https://orcid.org/0009-0005-3891-9463>, ²<https://orcid.org/0000-0002-8930-606X>

Email: *rammohanreddy.51@gmail.com

ARTICLE INFO

Article History

Received: January 30, 2026

Reviewed: July 28, 2026

Accepted: August 28, 2026

Published: September 30, 2026

Keywords:

Electronic Health record,
Machine Learning,
Deep learning,
Natural language Processing,
Risk Assessment,
Disease prediction.

ABSTRACT

The advent of electronic health records (EHRs) has revolutionized the healthcare industry by enabling the digitization and systematic organization of patient data. However, the massive volume and complexity of health data present significant challenges for manual analysis and summarization. This review article focuses on three main areas to explore the applications of deep learning (DL) and machine learning (ML) methods in the medical sector. First, it examines the use of EHR data for disease prediction, highlighting significant advancements in supervised learning, unsupervised learning, and various DL architectures. Second, it delves into risk prediction using EHR data, illustrating how ML and DL models are employed to assess and predict patient risks. This section discusses various algorithms and models, such as gradient boosting and recurrent neural networks, used to predict conditions like stroke, atrial fibrillation, and cardiovascular diseases. Lastly, the paper addresses the crucial topic of confidentiality and safety precautions for EHR data. Additionally, it discusses the evaluation metrics used to benchmark these models and examines the challenges and limitations in this domain. Future research directions and emerging trends are also outlined to further enhance the efficacy and adoption of ML and DL in health record management. The purpose of this study is to provide scholars and medical practitioners with a comprehensive overview of the current state and potential future applications of ML and DL for improving health record management.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

In today's increasingly digitized healthcare landscape, health records have become indispensable assets for enhancing patient care and driving medical research forward. These records encompass a comprehensive array of patient information, including detailed diagnoses, treatment histories, demographic profiles, and intricate medical narratives. They serve as invaluable repositories of knowledge, offering healthcare providers critical insights into patient health trajectories, treatment outcomes, and disease patterns. However, the sheer volume and complexity of health data present formidable obstacles to their efficient analysis and utilization. Healthcare facilities and research institutions are inundated with vast quantities of data generated from EHRs, medical imaging, genetic sequencing, wearable devices, and more [1]. Managing this wealth of information manually is not only labor-intensive but also prone to errors and inconsistencies, potentially compromising the quality of patient care and hindering scientific advancements. Moreover, the variability in data formats, standards, and terminology across different healthcare systems further complicates the integration and interoperability of health records. This fragmentation not only impedes seamless information exchange but also undermines efforts to derive comprehensive insights from aggregated data sources [2]. Manual analysis and summarization of health records have traditionally been labor-intensive and prone to human error, presenting significant hurdles in the efficient utilization of healthcare data [3]. Healthcare professionals, including physicians, nurses, and researchers, often face daunting tasks when attempting to extract relevant information from extensive and complex health records [4]. The process involves meticulously combing through vast amounts of textual and structured data, identifying pertinent patterns, and synthesizing findings from various disparate sources.

This manual approach is inherently timeconsuming, as it requires considerable effort and attention to detail. Professionals must dedicate substantial portions of their time to review patient histories, treatment records, laboratory results, and other pertinent documents. This extensive effort is not only a drain on valuable human resources but also detracts from the time available for direct patient care and other critical responsibilities [5]. Additionally, the manual nature of this process introduces the potential for human error. Even the most diligent healthcare professionals can overlook crucial details, misinterpret data, or inconsistently apply clinical guidelines. These errors can have serious implications, leading to incorrect diagnoses, inappropriate treatments, and ultimately compromising patient safety and outcomes [6]. The subjective nature of manual analysis further exacerbates these issues, as different individuals may interpret the same data differently, leading to variability and inconsistency in care. The limitations of manual analysis extend beyond the clinical setting to research and public health [7]. Researchers attempting to analyze large datasets manually are constrained by the sheer volume of data, limiting their ability to conduct comprehensive and large-scale studies. The depth of insights that can be gleaned from the data is restricted by the human capacity to process information, making it challenging to uncover subtle patterns and correlations that could inform medical research and public health strategies. In light of these challenges, there is a pressing need to leverage advanced technologies, such as AI and ML, to revolutionize health record management [8]. These technologies offer promising avenues for automating data extraction, pattern recognition, and predictive analytics, thereby empowering healthcare providers with timely and actionable information [9]. Healthcare companies may expedite processes, enhance diagnostic precision, customize treatment regimens, and accelerate medical advancements by utilizing AI-driven technologies. In summary, while health records constitute invaluable reservoirs of patient information and research data, their effective utilization hinges on overcoming the complexities posed by their sheer volume and diversity.

I.1 PROBLEM STATEMENT

Despite significant advancements in the digitization of health records, the manual analysis and summarization of these records remain a major bottleneck in healthcare operations [10]. The current human-driven processes are often slow, error-prone, and inconsistent, which hinders timely access to the critical information necessary for patient care and clinical research. The challenges are compounded by the variability in data formats and the sheer volume of information, making the task of deriving meaningful conclusions from health records increasingly complex. Manual analysis of health records is time-consuming and labor-intensive for healthcare professionals [11]. It involves reviewing patient histories, treatment records, laboratory results, and imaging reports. This process detracts from direct patient care and other critical responsibilities. Delays in accessing and analyzing data can slow decision-making processes, potentially leading to delays in diagnosis and treatment. Therefore, healthcare professionals should prioritize patient care and avoid manual analysis [12].

I.2 OBJECTIVE

The most recent advances in ML and DL methods for automated health record analysis are reviewed in this study. By examining recent advancements in AI-driven approaches, the objective is to critically evaluate their efficacy in transforming the analysis of health records, thereby addressing the challenges posed by manual methods. Through this review, insights will be gained into how AI can enhance the efficiency, accuracy, and scalability of health record analysis, ultimately improving healthcare delivery and patient outcomes.

I.3 CONTRIBUTION AND ORGANIZATION OF THIS PAPER

This review paper goes into the direction that involves the use of both ML and DL techniques to automatically process, analyse and interpret health records without the need for manual intervention.

- The paper reviews advancements in ML and DL techniques for predicting diseases from EHR data. It covers various approaches, including Natural Language Processing (NLP), supervised learning, and unsupervised learning approaches, along with specific DL architectures.
- The review highlights the use of ML and DL models in assessing and predicting patient risks based on EHR data. It discusses a range of algorithms, such as gradient boosting and recurrent neural networks, employed in predicting conditions like stroke, atrial fibrillation, and cardiovascular diseases.
- The paper addresses the critical aspect of safeguarding sensitive patient information. It reviews state-of-the-art security techniques which are designed to secure EHR data from breaches, illegal access, and other security risks.

This document is structured as follows: The background of the review article is explained in Section 2, and the systematic review of the proposed research is covered in Section 3, section 4 explains about the survey paper on ML and DL techniques based on automated health records and section 5 describes about the summary and discussion of the proposed work and section 6 concludes the paper with references.

II. BACKGROUND

The theoretical terms used in the literature review—specifically, EHRs and EHR privacy protection—is introduced in this section. Review of the literature on various technologies and protecting EHR privacy is provided at the conclusion, along with a thorough explanation of the research's objective. The glossary and main terms used in this article are listed in Table 1 [13].

II.1 ELECTRONIC HEALTH RECORDS

EHRs are digital patient records that can be safely shared and kept, allowing integrated healthcare to be managed continuously and successfully [14-16]. They include information about the health history of the patients, include diagnosis, test results, hospital admissions, surgeries, and medications. Further patient diagnosis and care are made possible by this data [17]. When necessary, EHRs can be accessed with other medical professionals; but, during transmission, they are vulnerable to breaches of confidentiality and security [18]. The healthcare industry is becoming more interested in creating a secure EHR sharing environment because of its broad use [19], [20].

Cost reductions, improved healthcare quality, evidence-based medical innovation, more thorough data collecting, and flexibility are all advantages of EHR software. EHR refers to a system that ensures data integrity, resistance to failure, high availability, and consistency of security standards in addition to an electronic database used for keeping and accessing medical records.

Table 1: PCS with details.

Glossary	Description
Privacy (P)	In an EHR, privacy means a person's right to preserve confidentiality of their medical data as well as to manage who may access and use it. This right is vital to patient care and is protected by a number of rules and regulations.
Confidentiality (C)	In order to maintain patient-provider confidence, healthcare workers are required by law and ethics to protect personal medical data from unauthorized disclosure.
Security (S)	The technological and administrative safeguards included in the security measures make sure that EHR data is available and secure by preventing unauthorized entry, disclosure, interruption, alteration, or destruction.
EHR Management	Effective and safe handling of information, storage, and transmission are all part of the comprehensive method for managing EHR systems, which guarantees ethical and legal compliance.
EHR systems and technologies	Cloud computing, blockchain, and AI-driven analytics are just a few examples of the technologies that fall under this general phrase when referring to EHR systems' hardware, software, and methods.
PCS framework	The interdependence of Privacy, Confidentiality, and Security is emphasized by the combined approach of these three factors in EHR systems, which guarantees comprehensive protection and administration of medical data.

Source: Authors, (2026).

II.1.1 EHR Privacy

The preservation of patients' rights over their medical information, particularly data security and physical privacy, is known as EHR privacy. In order to ensure patient ownership over their data, strict confidentiality and safety regulations must be maintained, and data access and transmission must be tracked [21]. Patients and healthcare professionals must trust one another, and maintaining privacy is a major cybersecurity concern. Strict regulations must be in place both nationally and internationally, and patients utilizing EHR systems ought to have ownership over their medical information. This covers recompense for data breaches as well as safeguarding against societal or financial prejudice. Appropriate privacy-preserving technologies are necessary for patients to have control over their data, and they have a right to know how their own medical data is used.

II.1.2 EHR Confidentiality

EHRs maintain the privacy of identifying patient data and can only be shared with explicit, informed agreement. It guarantees that private information is protected against unwanted dissemination [22]. Data encryption and respect for privacy regulations are important precautions. Data secrecy cannot be avoided, and providing access to data might jeopardize one's privacy. To prevent confidential details from unwanted exposure and preserve confidentiality, which is a fundamental component of cybersecurity, electronic data must be encrypted [23].

II.1.3 EHR Security

To prevent illegal entry, abuse, or breaches of medical data, EHR safety is essential. It entails access control, authorization, and verification processes, and adherence to laws like HITECH and HIPAA is crucial [24]. Since EHRs are accessible by multiple systems, there is a risk of abuse or illegal access, which might compromise patient privacy. Policies and procedures for access control are crucial for protecting medical records. Data safety protects against illegal access, disclosure, modification, loss, and destruction of personal data. Creating, executing, and overseeing shared care models all depend on maintaining EHR privacy. Healthcare organizations emphasize the importance of standards for shared and interoperable EHR services [25]. Figure 1 shows a few of the standard safety measures used to protect the confidentiality and safety of EHRs.

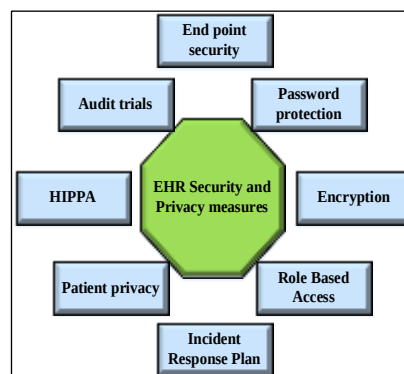


Figure 1: Security and Privacy measures.

Source: Authors, (2026).

II.1.4 The Distinction between Security, Confidentiality, and Privacy in EHRs

It is essential to understand the differences between EHR security, confidentiality, and privacy in order to safeguard all patient data. The Cyber Kill Chain (CKC) model describes these concepts that are very similar but distinctive in how they are used and where a compromise would occur in the context of a cyberattack.

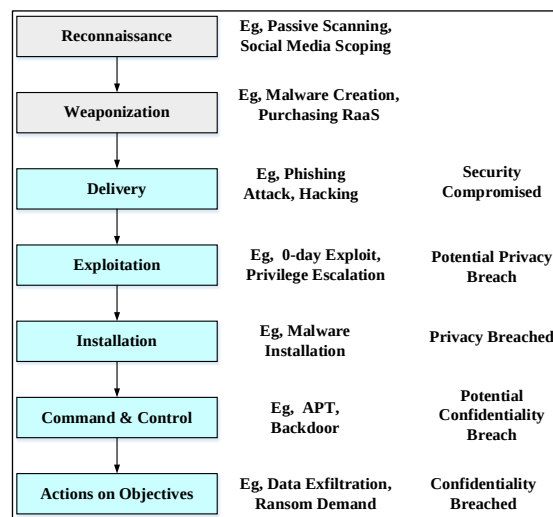


Figure 2: Cyber Kill Chain, highlights the vulnerabilities in EHR systems' security, privacy, and confidentiality.
Source: Authors, (2026).

This alignment with CKC stages (Figure 2) shows that, although a delivery mechanism alone may compromise safety, unauthorized viewing of patient data specifically compromises privacy, and extraction of information and potential use for malicious purposes, includes blackmail or public disclosure, compromises confidentiality. Understanding the distinctions and interrelationships between these concepts is essential for developing comprehensive strategies to protect patient data. By addressing the unique challenges associated with each aspect, healthcare organizations can enhance their EHR systems' resilience to cyber threats, thereby maintaining trust between patients and healthcare providers. Implementing robust privacy policies, ensuring confidentiality through stringent access controls, and maintaining strong security measures are all crucial steps toward achieving a secure and effective EHR environment.

III. REVIEW BASED ON ELECTRONIC HEALTH RECORDS

This section delves into a comprehensive survey of EHR and their summarization through the application of various ML and DL methods. It examines methodologies used in the processing and analysis of EHR data, highlighting the specific learning techniques employed.

III.1 DISEASE PREDICTIONS THROUGH EHR

This part offers a review of multiple studies publications that support the research on illness prediction-based EHR analysis. The effectiveness and therapeutic usefulness of ML techniques for identifying familial hypercholesterolemia (FH) in a sizable primary care population in UK were examined in research by [26]. Utilizing AUC and calibration slope, the investigation evaluated the predictive accuracy of 4,027,775 routine clinical records. The following five popular ML methods were evaluated: neural networks, LR, RF, gradient boosting machines, and ensemble learning. The methods detected increased triglycerides, which decreased the risk of FH, and elevated overall cholesterol levels that is recognized to be a clinical characteristic of FH, together with a family history of early coronary heart disease. With a considerable variation in the percentage of high likelihood instances for clinical examination between designs, all other ML techniques demonstrated strong prediction accuracy. According to the research findings, ML designs are quite accurate in identifying FH, however the amount of clinical labor required to find patients varies greatly throughout models. By [27] developed a DL procedure for early recognition of acute kidney injury (AKI) following cardiothoracic surgery. They used a RNN with 96 parameters to train on a balanced dataset of 2,224 admissions. The RNN outperformed experienced clinicians and showed superior calibration. The study suggests the RNN could be used in hospitals' EHR for real-time patient monitoring, potentially improving early AKI detection and perioperative care adjustments. However, unstructured text information was not available. A study by [28] aimed to diagnose Alzheimer's disease (AD) using electronic medical record (EMR) data without the need for diagnostic imaging. The ambulatory EMR data from US primary care practices was one of the datasets from which investigators took data. They trained a DL model with this information in order to forecast the chance of patients acquiring AD. When 90% of the SCRP dataset was used for training, the algorithm produced a high out-of-sample Area Under the Precision-Recall Curve score. The study concluded that integrating qualitative medical knowledge with advanced DL techniques significantly improved AD prediction accuracy. To have a better understanding of AD forecasting, further research on various medication groups, measures, and situations is required. In research by [29], 48,151 individuals with hepatitis C virus-related cirrhosis had their probability of acquiring hepatocellular carcinoma (HCC) predicted using DL RNN algorithms. RNN algorithms functioned better than traditional logistic regression designs, according to the research, particularly for patients who had a prolonged virologic response. This shows that RNN models can improve risk-based HCC outreach and surveillance programs by efficiently identifying people who are at high risk for acquiring HCC. But the research also emphasized its shortcomings, including its high processing cost and lack of external confirmation.

In study [30], 2.27 million patients from the UCHealth system were included in a large-scale case-control research to forecast the results of myocardial infarction (MI) using DL and ML algorithms. To solve the dataset imbalance, the research employed a variety of over- and under-sampling strategies. In comparison to a logistic regression model that used only known risk variables, the investigators discovered that a deep neural network with random undersampling produced higher classification performance, but the benefit was only marginal. Despite having sufficient discrimination, the whole group had poor calibration, mostly as a result of overfitting brought on by the infrequent nature of MI occurrences. Utilizing information from Healthcare Cost and Utilization Project State Inpatient Databases of California, Branimir and his colleagues [31] sought to forecast the beginning of 10 problems in patients with type 2 diabetes. The study examined how well RNN models—more especially, the gated recurrent unit (GRU) and LSTM—performed in comparison to more conventional approaches like the RF and multilayer perceptron. The RNN GRU design outperformed traditional models, achieving better prediction accuracy. However, prediction accuracy diminished over time. The highest accuracy was observed for depressive disorder and chronic ischemic heart disease, highlighting the potential of advanced DL models in clinical prognosis for DM2 patients. Deep neural networks (DNNs) were utilized in research by [32] to recognize patients treated for deliberate self-harm and forecast future self-harm incidents. Intentional self-harm ICD codes were available for 835 patients in this investigation; 1670 matched controls lacked these data. A variety of algorithms, involving conventional bag-of-words systems and two DNN systems—a CNN and a long short-term memory model—were evaluated using clinical notes associated with ICD codes. An analysis of a group of patients having clinical notes from one to six months prior to their first self-harm incident was conducted to assess the prediction ability of DNNs. The CNN achieved the highest performance in the phenotyping task, with highest AUC and an F1 score. Pretraining with Word2vec (W2V) reduced DNN training time but did not enhance performance. A multi-output Gaussian procedure and recurrent neural network (MGP-RNN) system was created by [33] to identify sepsis in adult hospital patients. Information from 2014 to 2015 were used to train the framework, it was verified. A blood culture order, at least one symptom of end-organ failure, and two or more systemic inflammatory response syndrome symptoms were used to identify sepsis. Compared to RF, Cox regression, penalized logistic regression, and clinical ratings, the MGP-RNN performed better. Sepsis was identified on average five hours in advance. Nevertheless, the design's shortcomings including its incapacity to handle aberrant laboratory data or higher vital signs brought on by chronic organ malfunction.

In order to construct and assess ANN classifiers for timely identification of cardiac arrest risks in emergency department patients, [34] carried out a single-center HER based research. Their study focused on three types of ANN designs: MLP, LSTM, and a hybrid model. Evaluating the models' performance using AUROC, they found that the MLP, LSTM, and hybrid models achieved AUROC values indicating superior predictive capability compared to non-ANN models. The hybrid ANN techniques, which integrated both baseline and sequence data, demonstrated the highest performance across various test characteristics, surpassing individual MLP and LSTM models. These findings underscore the potential of ANN classifiers, particularly hybrid models, for enhancing early detection systems in clinical settings. By [35] conducted a study on leveraging big data and DL to enhance predictive modeling in healthcare, specifically focusing on heart failure diagnosis using EHRs. Their research utilized recurrent neural networks (RNNs), including LSTM, to model temporal relations among patient events recorded over time. By integrating substantial historical EHR data, their proposed LSTM model, employing SiLU and tanh activations with Softmax output and Bridgeout regularization, demonstrated enhanced accuracy in predicting heart failure compared to traditional methods. Their findings underscored the potential of DL in improving initial recognition and management of heart failure, thereby presenting opportunities for cost reduction and better patient outcomes. Utilizing acute kidney failure as a case study, [36] established a DL technique for continuous risk estimation of patient deterioration. Their model, based on a comprehensive dataset of EHR from diverse clinical settings, accurately predicts a significant portion of inpatient episodes and identifies patients at risk of requiring dialysis up to 48 hours in advance, albeit with a 2:1 ratio of false to true alerts. Beyond predictions, the model provides confidence assessments, highlights relevant clinical features, and forecasts future trajectories of pertinent blood tests, offering potential for early intervention in acute kidney injury cases. In turn [37] used information from the Taiwan National Health Insurance database to create a DL algorithm that estimates the probability of developed colorectal cancer in adult Taiwanese citizens. CNN included patients' comorbidities and medication histories, and three years of medical data before the diagnosis of CRC. The study demonstrates the effectiveness of CNNs in predicting CRC depends on non-image and multi-dimensional medical records, with the robust performance of the model attributed to comprehensive patient data and extended medical records.

According to [38] developed a DNN for predicting stroke using a large EHR database. The research used Taiwan National Health Insurance Research Database in a retrospective population-based design, dividing the dataset into a development dataset and two testing datasets. Any ischemic stroke noted in medical files within three years of research enrollment was the main endpoint. The DNN model maintained high predictability up to five years after development, achieving performance comparable to other clinical risk assessment scores. This research highlights the potential of DNN-based intelligent decision support systems in clinical practice however, more investigation is required to verify its efficacy in enhancing therapeutic results. By [39] main goal was to utilize the commonly used CNN model to forecast PCA patients one year ahead of time based just on the most basic information from electronic health data. Nonetheless, the CNN method has demonstrated excellent specificity and sensitivity in predicting PCA with the fewest EHR features. These recent results make it abundantly clear that deep learning prediction models might provide doctors a great deal of potential to enhance patient outcomes by early PCA diagnosis and subsequent follow-up of high-risk PCA patients. Treating patients in a timely manner may increase survival rates, lower treatment costs, and lower the chance of death. This algorithm may be used to detect individuals at high risk of PCA one year ahead of time, based on patient demographics, comorbidities, and drug histories. Research by [40] used ML and DL methods to forecast hypotension following tracheal intubation. The research aimed to develop predictive models capable of forecasting hypotension one minute before tracheal intubation. Information from EHR and vital records of 282 out of 319 patients who had laparoscopic cholecystectomy during October 2018 and July 2019 were used to train the designs. 151 of the 282 individuals who underwent induction of hypotension suffered from it. The two experimental situations that comprised the study's framework were employing unprocessed vital records and incorporating feature engineering into the vital information. The disease forecast via EHR is displayed in Table 2. By [41] study demonstrates that EHRs and knowledge networks can be used to predict AD onset, prioritize biological hypotheses, and contextualize sex dimorphism in AD risk. They used a dataset consisting of 250,545 controls and 749 AD patients to train RF algorithms for predicting the onset of AD.

The models achieved a mean AUC of 0.72 when predicting AD onset 7 years prior and 0.81 when predicting onset 1-day prior. The study also highlighted the use of knowledge networks to identify shared genes among top predictors and AD, including genes such as APOE, ACTB, IL6, and INS. Genetic colocalization study showed a greater female AD connection with osteoporosis at a gene near MS4A6A and a link between AD and hyperlipidemia at the APOE locus. Research by [42] used five-year EHRs from a Taiwan hospital database to forecast chronic illnesses like diabetes. They found limitations in previous research, such as the use of publicly available datasets and imbalanced data. In order to forecast the risk of chronic disease, they created an original structure called Large Language Multimodal Models (LLMMs), which integrates multimodal data from laboratory test results and clinical notes. Their approach learned laboratory test values by combining a multi-head attention layer and text embedding encoder, and then merged blood characteristics with semantics of chronic diseases utilizing a DNN component. The study underscores the importance of integrating multimodal data and advanced language models in chronic disease prediction, paving the way for more accurate and early diagnosis, and enhancing the potential for personalized healthcare interventions.

Table 2: Review of various disease prediction using EHR.

Authors	Techniques used	Disease prediction	Performance	Limitations
Akyea et al [26]	MLP	Identifying familial hypercholesterolaemia in primary care	AUC > 0.89	clinical workload for detecting familial hypercholesterolaemia (FH) varies significantly
Rank et al [27]	RNN	Actual forecast of acute kidney injury	AUC 0.893	Lack of access to unstructured text information in EHR
Ljubic, Roychoudhury, et al [28]	LSTM	Alzheimer's disease prediction	AUPRC 0.89	insufficient diagnostic imaging data hinders the thorough evaluation of AD prediction models
Ioannou et al., [29]	RNN	Forecast Hepatocellular Carcinoma in Patients with Hepatitis-C Cirrhosis.	AUROC 0.806	Two major obstacles to HCC prediction are computational expense and lack of external validation.
Mandair et al., [30]	MLP	Forecast of incident myocardial infarction	F1 0.092, AUC 0.835	DL and ML models for myocardial infarction outcome prediction show poor calibration despite adequate discrimination, mainly due to overfitting from imbalanced datasets.
Ljubic, et al., [31]	RNN-GRU, LSTM	Forecasting complications of type-2 diabetes	Accuracy 73%	Use of RNN models for forecasting complications in type 2 diabetes patients is limited due to decreasing prediction accuracy over time.
Obeid et al., [32]	CNN, LSTM	Predicting Intentional Self-Harm in EHR	AUC 0.882, F1 0.769	Limited performance enhancement from pre-training with Word2vec
Bedoya et al.,[33]	RNN	Early detection of sepsis	C-statistic prediction 0.88	Inability to address abnormal laboratory values owing to chronic organ dysfunction limits the clinical utility of the multi-output
Jang et al., [34]	LSTM	cardiac arrest in ED	AUROC 0.933	Influencing the early identification of patients' risk of cardiac arrest in emergency rooms.
Maragat et al [35]	LSTM	Heart Failure in Big Data	AUC 0.894	Despite improved heart failure prediction accuracy, challenges persist in interpreting complex temporal relations among patient events using LSTM models trained on EHR data.
Tomašev et al., [36]	RNN	Forecast of future acute kidney injury	55.8% inpatient acute kidney injury	Challenge of integrating real-time unstructured data are significant limitations
Y. H. Wang et al., [37]	CNN	Recognition of Colorectal Cancer	AUC 0.920	Extending the prediction window could enhance practical applicability but would require further validation.
Hung et al [38]	MLP	Prediction of ischemic stroke	AUC 0.922	Clinical procedures might be improved by DNN-based IDSS.
Islam et al [39]	CNN	Forecast Model for Prostate Cancer	AOC 0.94	Inconsistencies or missing data points may reduce the reliability of predictions.

J.Lee et al [40]	CNN	Forecast Post-Induction Hypotension	Accuracy 74.89%	Limited sample size and lack of diversity may affect the generalizability of the model.
Alice et al [41]	DL models	Alzheimer disease prediction	AUC 0.72	Limited performance enhancement
Jun-En et al [42]	LLMMs framework integrating multimodal data	Chronic disease prediction	73% accuracy in multi-class prediction, 76% AUROC	Use of specific hospital dataset, may not generalize to other populations, data imbalance issues.

Source: Authors, (2026).

The studies reviewed highlight the efficiency of ML and DL models in predicting various medical conditions using EHRs. These models demonstrate high accuracy in detecting diseases such as familial hypercholesterolaemia, acute kidney injury, Alzheimer's disease, hepatocellular carcinoma, myocardial infarction, and complications in type 2 diabetes patients. They often outperform traditional methods like logistic regression and random forest, offering potential for early intervention and improved patient outcomes. Improving model calibration, addressing dataset biases, and integrating unstructured data from clinical notes could further enhance predictive accuracy. Additionally, exploring the integration of genetic, lifestyle, and environmental data with EHRs may offer more comprehensive risk assessments. These advancements hold promise for advancing personalized medicine and proactive healthcare management.

III.2 RISK PREDICTIONS THROUGH EHR

This part offers a thorough analysis of the numerous studies publications that support the risk-prediction-based analysis of the EHR study. AI approaches were used by [43] to enhance the forecasting of stroke risk using EHR text. The NLP-based system uses EHR data to produce structured text. Next, it uses to categorize the language into "good" and "bad" indicators, which are then utilized for forecasting. In this study, the ML models that were utilized are SVM and LR. The outcomes demonstrate that both approaches are capable of properly predicting outcomes and classifying the text. However, limited techniques only implemented and also low accuracy obtained. By [44], the purpose of this work was to create an ML method utilizing EHR to evaluate the likelihood of atrial fibrillation that is discovered in the initial stages following a stroke. This investigation made use of linked data from a hospital stroke registry and a de-identified research-based database that contained administrative claims data and EHRs. EHRs were used for obtaining clinical free text, physiological measures, routine laboratory findings, and demographic data. The prediction model was constructed with the help of the XGB technique. The AS5F and CHASE-LESS scores were compared to the forecasting performance measured by the C-index. To evaluate the forecasting model's clinical utility, more research is required. In turn [45], created an understandable risk calculator for COVID-19 personal risk forecasting system using de-identified EHR. This used longitudinal EHR data from 7,262 patients enrolled in the UAB Health System who were assessed and/or tested for COVID-19 during January and June 2020 to estimate risk using a Credit Scorecard modeling technique. 912 people in this group tested positive for COVID-19. This process used several variable selection approaches, including LASSO and Elastic-Net, to assess the effects while taking into account the timing of illnesses and underlying health problems. The findings led to the creation of questionnaires aimed at estimating the risk of COVID-19 infection for screening at building entrances or optimizing hospital resources. Due to unequal class results caused by high frequency of stroke events, trained neural network algorithms were skewed toward making negative forecasts. According to [46] created and added regularization terms to the conventional cross-entropy loss function in order to solve this problem. False positive and false negative forecasts were punished by these phrases. In this study, they used ROC to assess the trained systems' performance. Using a dual-input architecture, the optimal neural network integrated and merged the various temporal data sources.

A ML based phenotyping system for incident stroke ascertainment was created by [47] employing NLP to retrieve diagnostic codes, procedure codes, and medical terms from clinical notes. The development and validation of an EHR-based system that reliably distinguishes incident stroke occurrences beyond standard case-versus-noncase stroke detection was detailed in this work. The technique's strong performance in an example of overall population indicates its generalizability and potential for other institutions to adopt it. An existent epidemiological dataset of 4914 atrial fibrillation (AF) patients with carefully selected incident stroke occurrences was used to train and verify the method. This work compared several feature set and ML classifier combinations. The goal of study [48] is to apply DL approach and EHRs using risk variables to forecast the LSTM-RNN method for cerebrovascular illness. The EHRs offer comprehensive information on a patient's physical and mental well-being, including their diagnosis, test findings, treatment plan, and course of care. The accuracy rate, recall, and F1 measure score obtained by the LSTM-RNN approach differ from those obtained via backpropagation and the RNN method. This demonstrates that the LSTM technique works well for early detection of any stroke or cerebrovascular illness. By utilizing ML techniques, [49] hope to enhance the risk forecasting of atherosclerotic cardiovascular disease (ASCVD) and enable automated population screening. The 5-year ASCVD risk was predicted using XGBoost technique and Least Absolute Shrinkage and Selection Operator (LASSO) regression methods; the XGBoost model had a considerably better C-statistic than other approaches. XGBoost and LASSO had net reclassification enhancement rates of 3.9% and 2.8%, correspondingly. By [50], studied ML-methods for diagnosing AF risk using EHR data that has been aligned to Observational Medical Outcomes Partnership Common Data Model. Data from 2,252 219 patients treated in three large Colorado hospitals that make up the UCHHealth hospital system between January 1, 2011, and October 1, 2018, were utilized in this diagnostic analysis. Adjudicated retrospectively, the event AF was classified in specified 6-month intervals using the F1 statistic and the AUROC curve. In turn [51], over the next year, a model to forecast risk for incident essential hypertension was created and prospectively verified. Utilizing EHR data from over 1.5 million Maine residents, this study created and prospectively validated a risk forecast framework for future 1-year incident essential hypertension. Maine is one of the states where the current predictive analytics approach is currently being implemented. The EHR-based algorithm correctly classified patients into five different risk groups, ranging from very low to very high, and it reached predictive accuracy in both retrospective and prospective (validation) cohorts. According to [52] suggested a multimodal RNN framework for cardiovascular risk prediction utilizing EHR data, which includes structured clinical data and medicinal texts.

Word embedding's are concatenated to conventional clinical predictors in the suggested multimodal bidirectional long short-term memory (BiLSTM) algorithm before being applied to a final fully connected neural network. CNN and LSTM are two of the DNN designs whose performance is evaluated in these experiments utilizing data from chest X-ray imaging and clinical features. By [53], created and verified a prospective risk forecasting system for identifying individuals in the general community who are at risk of developing lung cancer again during the following year. The Maine Health Information Exchange network was used to retrieve information from each patient's EHR. An XGBoost technique was used in the design of the model. Every person was given a score to indicate the likelihood of receiving a new incident lung cancer diagnosis between October 1, 2016, and September 31, 2017. Compared to the entire cohort, the high-risk group had a 7.7-fold greater incidence of lung cancer. According to [54], patients with congestive heart failure (CHF) frequently experience unplanned 30-day readmissions, which increase risk factors and medical expenses. Targeted intervention strategies for individuals at risk of readmission should be started with a goal to lower readmission rates and control healthcare costs. To do this, high-risk patients must be identified while they are being released from the hospital. In this study, expert features and contextual embedding of clinical ideas are used to formulate an LSTM neural network in a cost-sensitive manner. This investigation addresses the issue of class imbalance and incorporates sequential patterns into an overall structure that targets essential components of an EHR driven forecasting model. Building on the EHR of more than 23,000 cardiac patients, [55], used DL algorithms to forecast improved risk of mortality in CVD patients. This work trained and analyzed BERT and XLNet, two popular transformer methods that rely on discovering bidirectional relationships in sequential data. This effort is initial to use XLNet to forecast mortality using EHR data. The framework was capable to build ever-more complex temporal relationships since patient histories were created as time series composed of various clinical events. BERT and XLNet were able to achieve an average AUROC. Table 3 shows the various methods and their features.

Table 3: Review of various risk prediction using EHR.

Authors	Techniques used	Risk prediction	Performance	Limitations/ Future works
Song et al [43]	SVM, LR	Stroke risk prediction	Accuracy SVM- 73% LR-71.5%	Low accuracy
Sheng et al [44]	Extreme gradient boost	Stroke prediction	C-index 0.840	Research are needed to verify the generalizability of the technique.
Tarun et al [45]	LASSO regression, Elastic Net	COVID -19 prediction	-	this workflow cannot establish causal Relationships.
Douglas et al [46]	Neural networks	Stroke prediction	ROC 0.669	Datasets does not include all risk factors.
Zhao et al [47]	Random forest, Logistic classification	Stroke prediction	Accuracy 83%	Oversampling techniques and imbalance cohort data may have caused overfitting.
Pattanapong et al [48]	LSTM-RNN	Stroke prediction	Accuracy 98% Precision 99%	Apply DL method to an e-stroke application and incorporate other risk indicators for prediction.
Li et al [49]	XG Boost algorithm and LASSO regression	Cardiovascular risk prediction	Higher C-statistic of 0.792.	Apply DL method to an e-stroke application and incorporate other risk indicators for prediction.
Tiwari et al [50]	Logistic regression	Atrial fibrillation risk prediction	ROC-0.794 F1-score-0.079	Used a very basic technique that neglected to take censoring or time-varying effects into consideration when handling the temporal connections among characteristics.
Chengyin et al [51]	XG Boost classifier	Hypertension prediction	AUC-0.917	EHR records will inevitably have missing data, thus organizing and extracting this information is essential.
Ayoub et al [52]	Multimodal BiLSTM	Cardiovascular risk prediction	AUC-83.8%	Laboratory results of patients are missing.
Xiaofang et al [53]	XGBoost algorithm	Lung cancer prediction	AUC-0.881	The EHR data did not adequately reflect tobacco usage; some variables were absent from the dataset.
Awais et al [54]	LSTM technique	Heart failure risk prediction	AUC-0.77, F1-0.51	In order to integrate the framework into clinical workflow, do a feasibility study.
Emmi et al [55]	BERT and XLNet	Cardiovascular risk prediction	Recall 9.8%	More complex representations might yield better outcomes.

Source: Authors, (2026).

The studies discuss the use of ML and AI techniques in predicting health risks using EHR. Some use NLP and ML techniques for stroke risk prediction, while others use extreme gradient boosting for atrial fibrillation risk assessment.

A third creates an interpretable COVID-19 risk calculator using a Credit Scorecard approach, showing potential for screening and resource optimization. An ML-based phenotyping algorithm is developed for stroke ascertainment, while LSTM-RNN algorithms are used to predict cerebrovascular disease. XGBoost and LASSO are applied to predict 5-year ASCVD risk, while ML on harmonized EHR data is used to identify AF risk. A prospective model predicts one-year incident essential hypertension using Maine state EHR data, while a multimodal BiLSTM model incorporating structured clinical data and medical texts forecasts cardiovascular risk. A model predicts new incident lung cancer risk within a year using XGBoost algorithm on Maine EHR data, showing higher incidence in the high-risk category. DL models like BERT and XLNet are trained to predict mortality risk in CVD patients, achieving high performance metrics. These studies highlight the potential of advanced AI and ML techniques in enhancing predictive analytics and healthcare outcomes using EHR data. The next section provides about the security applications in EHR data's.

III.3 SECURITY AND PRIVACY MEASURES IN EHR

This part offers a thorough analysis of several studies publications that support the analysis of EHR work from a privacy and security perspective. By [56] coupled the Data Encryption Standard (DES) method in MATLAB with Arithmetic Coding (AC) technique to improve security, confidentiality, privacy of data, imperceptibility, secrecy, and assured retrieval of personal data for DICOM pictures. In the case of high quality, high embedding capacity, confidentiality, robustness, and security but higher computational cost, performance evaluation with MSE, RMSE, PSNR, SSIM, Weighted PSNR, Universal Quality (UQ), KLD, Manhattan Distance (MD), and Euclidean Distance (ED) demonstrated simple and faster implementation, while low performance and coarse results on time-frequency analysis was observed. In turn [57] suggested using the Interplanetary File System (IPFS) to get Covid-19 history in a genuine manner. This technique further performed safety evaluations against a variety of threats by combining SHA-256 with a chaotic map. Displayed a diagonal, vertical, and horizontal correlation of zero, with an average entropy of 7.998, a UACI of 33.37%, and an NPCR of 99.6437. Lastly, surpassing the current design in terms of confidentiality and security. According to [58] discuss the risks associated with EMR exposure to issues including quantum assaults, data searchable safety, and ledger compulsion. proposing an effective cross-institution EMR sharing system that combines a successful On-chain Ledger and Off-chain Storage (OLOS) with Keyword-Searchable Attribute-Based Encryption (KS-ABE) to enhance EMR's quantum resistance. The obtained times were 217.72 ms and 899.59 ms, respectively, and 29.09 s for the 1000 keyword index. By [59] integrated RSA- and AES-based steganography on RGB pixels of 322 PNG medical care photos that were retrieved from MIAS and roughly measured 1024 by 1024. The goal is to safeguard and maintain the privacy of patients with breast cancer by preventing data breaches, unauthorized disclosure, inaccessibility, manipulation assaults, and ensuring safe transfer of information among sender and recipient. The outcomes indicated that there was no loss for medical information, a boost in execution time, and 100% integrity. The MSE was 0.00257 and PSNR was 78.79. Hence, with an average concealing capacity of 266227, enhanced secrecy, confidentiality, and authenticity could be guaranteed.

In turn [60] carried out three-phase DNN-based medical picture steganography: 1) Preparation Network: in this case, RGB cover is used to resize the secret picture; 2) Hiding Network: an RGB image container is made just for the reason of transmission; 3) Reveal Network: secret message is then extracted from container image. In order to achieve an average PSNR loss of 17.54% for the secret to decoded picture, comparison tests were conducted using LSB, Discrete Coefficient Transform (DCT), Discrete Wavelet Transform (DWT), and binary pattern complexity. Additionally, analysis was conducted. SSIM was 0.94, KL-Di was 957.82, Intersection was 0.95, and R was 0.97. According to [61] used MATLAB and Python to create neural Stacked Auto-Encoder (SAE) based medical image encryption using the chaos method. Medical data sets were prepared, data bits were shuffled, a chaotic sequence was generated, medical picture pixels were shuffled, and bewilderment resulted. The decoded picture had a PSNR of 23.1090, SSIM of 0.7428, FSIM of 0.8727, and an encryption and decryption time of 2.216. For gaussian noise, the PSNR was 27.5669, SSIM was 0.9381, FSIM was 0.987, and for speckle noise, it was 29.4350. With SSIM of 0.9425, FSIM of 0.9618, and Edge Differential Ratio (EDR) of 0.8660, the system is computationally demanding but also cost-effective, efficient, secret, and has higher performance and security against threats. By [62] created a safe searchable encryption system on Hyperledger blockchain using DL. Establishing a 94.34% accuracy with 8.89 loss and an IoT model accuracy of 88.38% with 8.92 loss, the suggested technique was graphically compared to benchmark algorithms such as Medblock, Medchain, and Cash et al. This demonstrated that the technique has enhanced performance, less search and communication time, less time for execution, less index-search time, more efficiency, throughput, and greater resistance against collision attacks, DoS, and DDoS, making it more effective. In turn [63] applied the suggested cryptographic technique on genuine Holter cardiac data, photoplethysmography signals, and electrocardiograms (ECGs). Employing Microsoft Visual C++ to secure medical data during transmission against illegal access and maintain confidentiality. SNR varied from 31.83 to 46.37, and compression ranged from 3.84 (PPG) to 4.98 (Holter). A excellent quality signal was produced for length 128 with a PSNR of 43.62 dB, MSE of 0.051 dB, PRD of 6.31%, RMSE of 0.2258, MAPE of 0.013±0.002%, BER of 0.11, and KLD of 0.014. Table 4 presents a comparative examination of several EHR security strategies.

According to [64] applied IWT-LSB-based steganography to a medical image. Using a piecewise chaotic map to apply pixels, she generated a key stream, which was then generated by decomposing bits, which led to diffusion. Next, utilizing the IWT lifting approach, she embedded a secret picture using cyclic right shift method with LSB. Then, with the MSE of color images ranging from 0.1 to 0.15 and that of grey images from 0.15 to 0.2, PSNR of color images ranging from 54 to 56 and that of grey images from 52 to 54dB, and both demonstrating a rise in privacy and reliability with a Normalized Cross-Correlation (NCC) of 1, a Hybrid Neural Network (HFNN) improved image quality also led to an improve in data hiding capacity. By [65] showed a research on the development of cloud-based EHRs over the past decade, focusing on their role in remote patient monitoring. The study highlighted the importance of Healthcare 4.0, which integrates IoT components and cloud computing for remote medical operations. The study highlighted the challenges in safeguarding patient data from unauthorized access, requiring robust security measures. Traditional approaches sometimes fall short in balancing effective safety, service-side verification, and computing performance in the absence of a trusted third party, even when cryptographic techniques are used. The capacity of blockchain-based safety measures to offer robust safety for data exchange and storage with little computational overhead has drawn attention. In turn [66] studied the integration of blockchain platforms in healthcare management, focusing on privacy in medical records for patient diagnosis and treatment.

They identified challenges such as authorized record transfers, security across platforms, content diversity, and platform interoperability. They proposed a permissioned block chain for healthcare record management, using a hyper-fabric ledger on a peer-to-peer chain. The system involved patients, practitioners, and health-related officers who created, retrieve, and stored patient records. The system demonstrated a transaction response time of 0.56 seconds for queries and 0.42 seconds for HTTPS requests for 500 users. They also presented an electronic health block chain-based support system, using a hyper-fabric ledger to record world-state key values, transform records, and aid interoperability.

Table 4: Security measures in EHR.

Authors & Ref no:	Security technique used	Data format used	Data used	EHR data used or not
Ahmed et al [56]	Enhanced wavelet steganography	Cover: DICOM images; secret message: characters	10 images of different wavelengths hidden with 1000, 5000, 10000, 17400, and 20000 characters	Yes
Falmari et al [57]	Cryptography	MRI, X-Ray, and CT-Scan all as digital images	Database for British Society of Thoracic Imaging COVID-19	Yes
Li et al [58]	Lattice cryptography-based Keyword-searchable attribute-based encryption method (KS-ABE)	OLOS model for a lightweight block-chain ledger	-	Yes
Khalifeh et al [59]	RSA and AES steganography combined with LSB	Image of cover signal; the secret signal is a patient data	A total of 322 identically sized mammography pictures 1024 x 1024	Yes
Desari et al [60]	Deep neural network-based steganography	Cover picture: A medical picture used as a covert signal	10 cover images	Yes
El-Shafari et al [61]	Stacked-Auto Encoder Cryptography	Image	5 various grey medical images of size 256×256	Yes
Ali et al [62]	cryptographic technique utilizing safe searchable block chain encryption with DL homomorphism	Hyperledger	Log file	Yes
Georgieva et al [63]	RSA cryptography for steganography and public key encryption	ECG, PPG, Holter records	Two hours of cardiac recording using 24 distinct triple parallel ECG, PPG, and Holter signal recordings	Yes
Aleisa et al [64]	IWT-LSB Steganography	Cover photo containing confidential information	Secret text up to 8 and 192 digits	Yes
Hemant et al [65]	Blockchain-based medical care cloud record security	Patient sensitive dataset	There are twenty medical users, and the data sizes range from 128 to 64 MB (512 KB, 2, 8, 32, and 64 MB).	Yes
James et al [66]	Permissioned Block-chain with hyper fabric ledger	Key value pair with health information	Patient medical records with transactions	Yes

Source: Authors, (2026).

The studies collectively highlight the critical importance of enhancing security and privacy for EHR and medical images. They emphasize the need for advanced methods to ensure the protection of sensitive patient information. High performance metrics were achieved, demonstrating improvements in data confidentiality, integrity, and robustness against various attacks. These methods contribute to reducing unauthorized access, data breaches, and ensuring secure data transmission. These studies collectively emphasize the critical importance of robust security and privacy measures in handling EHR data. High-performance metrics achieved through advanced encryption, steganography, and block-chain techniques demonstrate their potential to protect sensitive medical information effectively. The integration of these methods ensures secure, efficient, and reliable handling of EHRs, addressing the growing need for confidentiality and data integrity in healthcare.

IV. SUMMARY AND DISCUSSION

ML and DL models have exposed important potential in predicting various medical conditions using EHR. These models have shown high accuracy in identifying diseases and these models often surpass traditional methods, facilitating early intervention and improved patient outcomes. Enhancements in model calibration, mitigation of dataset biases, and integration of unstructured clinical notes can further enhance predictive accuracy. Integrating genetic, lifestyle, and environmental data with EHRs can provide more comprehensive risk assessments, paving the way for personalized medicine and proactive healthcare management. ML and AI techniques have also been applied to predict health risks using EHRs. Techniques like NLP and extreme gradient boosting have shown promise in predicting stroke and atrial fibrillation risks. Models like the COVID-19 risk calculator demonstrate potential applications in screening and resource optimization. Deep learning models like BERT and XLNet achieve high performance metrics in predicting mortality risk in cardiovascular disease patients. Security and privacy are crucial in handling EHRs and medical images.

Advanced encryption, steganography, and block chain techniques have been shown to improve data confidentiality, integrity, and robustness against various attacks. Future research could focus on developing more sophisticated encryption algorithms and exploring block chain technology for secure data management solutions.

V. CONCLUSION

The integration of ML and DL techniques with EHR marks a transformative shift in healthcare. This review has explored three primary applications: disease prediction, risk assessment, and security and privacy measures. By leveraging advancements in NLP, supervised and unsupervised learning, and various DL architectures, significant strides have been made in predicting diseases and assessing patient risks. These technologies enable early interventions, personalized treatments, and overall improved patient outcomes. Concurrently, addressing the critical need for data security and privacy, advanced encryption, steganography, and blockchain methods ensure the protection and integrity of sensitive patient information. As the adoption of these technologies continues to grow, the healthcare industry stands to benefit from more efficient, secure, and accurate health record management. Expanding the utilize of blockchain technology in EHR management can provide immutable and decentralized data storage, ensuring data integrity and enhancing security measures against unauthorized access and tampering. The healthcare sector may handle EHRs in a more safe, reliable, and effective manner by going in new directions, which will eventually improve patient care and medical results.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Rammohanreddy Dondeti, Gatram Rama Mohan Babu.

Methodology: Rammohanreddy Dondeti, Gatram Rama Mohan Babu.

Investigation: Rammohanreddy Dondeti, Gatram Rama Mohan Babu.

Discussion of results: Rammohanreddy Dondeti, Gatram Rama Mohan Babu.

Writing – Original Draft: Rammohanreddy Dondeti, Gatram Rama Mohan Babu.

Writing – Review and Editing: Rammohanreddy Dondeti, Gatram Rama Mohan Babu.

Resources: Rammohanreddy Dondeti, Gatram Rama Mohan Babu.

Supervision: Rammohanreddy Dondeti, Gatram Rama Mohan Babu.

Approval of the final text: Rammohanreddy Dondeti, Gatram Rama Mohan Babu.

VII. ACKNOWLEDGMENTS

The authors would like to thank the Deanship of Acharya Nagarjuna University for supporting this work.

VIII. REFERENCES

- [1] P. Bhambri and S. Rani, "Bioengineering and healthcare data analysis: Introduction, advances, and challenges," in *Computational Intelligence and Blockchain in Biomedical and Health Informatics*, pp. 1–25, 2024.
- [2] A. Y. Lin, S. Arabandi, T. Beale, W. D. Duncan, A. Hicks, W. R. Hogan, and S. Schulz, "Improving the quality and utility of electronic health record data through ontologies," *Standards*, vol. 3, no. 3, pp. 316–340, 2023.
- [3] T. I. Amosa, L. I. B. Izhar, P. Sebastian, I. B. Ismail, O. Ibrahim, and S. L. Ayinla, "Clinical errors from acronym use in electronic health record: A review of NLP-based disambiguation techniques," *IEEE Access*, vol. 11, pp. 59297–59316, 2023.
- [4] D. Nakikj, D. Kreda, and N. Gehlenborg, "Alerts and collections for automating patients' sensemaking and organizing of their electronic health record data for reflection, planning, and clinical visits: Qualitative research-through-design study," *JMIR Human Factors*, vol. 10, p. e41552, 2023.
- [5] N. W. Rasheed, M. Y. Abdullah, N. H. Sherif, and O. Dudnikov, "The risk of data loss in computerized medical record systems," in *Proc. 35th Conf. Open Innovations Association (FRUCT)*, Apr. 2024, pp. 560–569.
- [6] J. Benbassat, "Medical errors and quality assurance of healthcare," in *Teaching Professional Attitudes and Basic Clinical Skills to Medical Students: A Practical Guide*. Cham, Switzerland: Springer, 2023, pp. 125–139.
- [7] N. Faizi and Y. Alvi, *Biostatistics Manual for Health Research: A Practical Guide to Data Analysis*. Amsterdam, The Netherlands: Elsevier, 2023.
- [8] S. A. Alowais *et al.*, "Revolutionizing healthcare: The role of artificial intelligence in clinical practice," *BMC Medical Education*, vol. 23, no. 1, p. 689, 2023.
- [9] K. Olson, "A comprehensive review on healthcare data analytics," *Journal of Biomedical and Sustainable Healthcare Applications*, vol. 3, no. 2, pp. 95–105, 2023.
- [10] J. A. Balch *et al.*, "Machine learning-enabled clinical information systems using fast healthcare interoperability resources data standards: Scoping review," *JMIR Medical Informatics*, vol. 11, p. e48297, 2023.
- [11] S. Patel, B. Gayakvad, R. Solanki, R. Patel, and D. Khunt, "Towards the digitization of healthcare record management," in *Human-Machine Interface: Making Healthcare Digital*, pp. 411–447, 2023.
- [12] L. A. Baumann and A. L. Brütt, "Comparing an in-person workshop and a postal Delphi survey for involving health service users in health care and health research prioritization," *Health Expectations*, vol. 26, no. 1, pp. 199–212, 2023.
- [13] R. Nowrozy, K. Ahmed, A. S. M. Kayes, H. Wang, and T. R. McIntosh, "Privacy preservation of electronic health records in the modern era: A systematic survey," *ACM Computing Surveys*, vol. 56, no. 8, pp. 1–37, 2024.
- [14] M. Anshari, "Redefining electronic health records (EHR) and electronic medical records (EMR) to promote patient empowerment," *International Journal on Informatics for Development*, vol. 8, no. 1, pp. 35–39, 2019.

- [15] H. Y. Chen, Z. Y. Wu, T. L. Chen, Y. M. Huang, and C. H. Liu, "Security privacy and policy for cryptographic based electronic medical information system," *Sensors*, vol. 21, no. 3, p. 713, 2021.
- [16] M. M. Tripathi, M. Haroon, Z. Khan, and M. S. Husain, "Security in digital healthcare system," in *Pervasive Healthcare: A Compendium of Critical Factors for Success*, pp. 217–231, 2022.
- [17] R. Nowrozy, K. Ahmed, A. S. M. Kayes, H. Wang, and T. R. McIntosh, "Privacy preservation of electronic health records in the modern era: A systematic survey," *ACM Computing Surveys*, vol. 56, no. 8, pp. 1–37, 2024.
- [18] R. Nowrozy, K. Ahmed, H. Wang, and T. McIntosh, "Towards a universal privacy model for electronic health record systems: An ontology and machine learning approach," in *Informatics*, vol. 10, no. 3, p. 60, Jul. 2023.
- [19] B. Al Bahrani and I. Medhi, "Copy-pasting in patients' electronic medical records (EMRs): Use judiciously and with caution," *Cureus*, vol. 15, no. 6, 2023.
- [20] R. Cerchione, P. Centobelli, E. Riccio, S. Abbate, and E. Oropallo, "Blockchain's coming to hospital to digitalize healthcare services: Designing a distributed electronic health record ecosystem," *Technovation*, vol. 120, p. 102480, 2023.
- [21] M. Farhadi, H. Haddad, and H. Shahriar, "Static analysis of HIPAA security requirements in electronic health record applications," in *Proc. IEEE 42nd Annual Computer Software and Applications Conf. (COMPSAC)*, Jul. 2018, vol. 2, pp. 474–479.
- [22] A. R. Caffrey and A. R. Horn, "Considerations for protecting research participants," in *Pragmatic Randomized Clinical Trials*. Cambridge, MA, USA: Academic Press, 2021, pp. 273–292.
- [23] G. K. Lettieri *et al.*, "Medical confidentiality in the digital era: An analysis of physician-patient relations," *Revista Bioetica*, vol. 29, pp. 814–824, 2022.
- [24] B. D. Deebak and F. Al-Turjman, "Secure-user sign-in authentication for IoT-based eHealth systems," *Complex & Intelligent Systems*, vol. 9, no. 3, pp. 2629–2649, 2023.
- [25] J. Yin, M. Tang, J. Cao, M. You, H. Wang, and M. Alazab, "Knowledge-driven cybersecurity intelligence: Software vulnerability coexploitation behavior discovery," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 4, pp. 5593–5601, 2022.
- [26] R. K. Akyea, N. Qureshi, J. Kai, and S. F. Weng, "Performance and clinical utility of supervised machine-learning approaches in detecting familial hypercholesterolaemia in primary care," *NPJ Digital Medicine*, vol. 3, no. 1, p. 142, 2020.
- [27] N. Rank *et al.*, "Deep-learning-based real-time prediction of acute kidney injury outperforms human predictive performance," *NPJ Digital Medicine*, vol. 3, no. 1, p. 139, 2020.
- [28] B. Ljubic *et al.*, "Influence of medical domain knowledge on deep learning for Alzheimer's disease prediction," *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105765, 2020.
- [29] G. N. Ioannou *et al.*, "Assessment of a deep learning model to predict hepatocellular carcinoma in patients with hepatitis C cirrhosis," *JAMA Network Open*, vol. 3, no. 9, p. e2015626, 2020.
- [30] D. Mandair, P. Tiwari, S. Simon, K. L. Colborn, and M. A. Rosenberg, "Prediction of incident myocardial infarction using machine learning applied to harmonized electronic health record data," *BMC Medical Informatics and Decision Making*, vol. 20, pp. 1–10, 2020.
- [31] B. Ljubic *et al.*, "Predicting complications of diabetes mellitus using advanced machine learning algorithms," *Journal of the American Medical Informatics Association*, vol. 27, no. 9, pp. 1343–1351, 2020.
- [32] J. S. Obeid *et al.*, "Identifying and predicting intentional self-harm in electronic health record clinical notes: Deep learning approach," *JMIR Medical Informatics*, vol. 8, no. 7, p. e17784, 2020.
- [33] A. D. Bedoya *et al.*, "Machine learning for early detection of sepsis: An internal and temporal validation study," *JAMIA Open*, vol. 3, no. 2, pp. 252–260, 2020.
- [34] D. H. Jang *et al.*, "Developing neural network models for early detection of cardiac arrest in emergency department," *The American Journal of Emergency Medicine*, vol. 38, no. 1, pp. 43–49, 2020.
- [35] G. Maragatham and S. Devi, "Retracted article: LSTM model for prediction of heart failure in big data," *Journal of Medical Systems*, vol. 43, no. 5, p. 111, 2019.
- [36] N. Tomašev *et al.*, "A clinically applicable approach to continuous prediction of future acute kidney injury," *Nature*, vol. 572, no. 7767, pp. 116–119, 2019.
- [37] Y. H. Wang, P. A. Nguyen, M. M. Islam, Y. C. Li, and H. C. Yang, "Development of deep learning algorithm for detection of colorectal cancer in EHR data," in *MEDINFO 2019: Health and Wellbeing e-Networks for All*. Amsterdam, The Netherlands: IOS Press, 2019, pp. 438–441.
- [38] C. Y. Hung, C. H. Lin, T. H. Lan, G. S. Peng, and C. C. Lee, "Development of an intelligent decision support system for ischemic stroke risk assessment in a population-based electronic health record database," *PLOS ONE*, vol. 14, no. 3, p. e0213007, 2019.
- [39] M. Islam, H. C. Yang, P. A. Nguyen, Y. H. Wang, T. N. Poly, and Y. J. Li, "Deep learning approach for the development of a novel predictive model for prostate cancer," *Studies in Health Technology and Informatics*, vol. 270, pp. 1241–1242, 2020.
- [40] J. Lee *et al.*, "Comparative analysis on machine learning and deep learning to predict post-induction hypotension," *Sensors*, vol. 20, pp. 1–21, 2020.
- [41] A. S. Tang *et al.*, "Leveraging electronic health records and knowledge networks for Alzheimer's disease prediction and sex-specific biological insights," *Nature Aging*, vol. 4, no. 3, pp. 379–395, 2024.
- [42] J. E. Ding *et al.*, "Large language multimodal models for 5-year chronic disease cohort prediction using EHR data," *arXiv preprint arXiv:2403.04785*, 2024.
- [43] S. Jiang, Y. Gu, and E. Kumar, "Stroke risk prediction using artificial intelligence techniques through electronic health records," *Artificial Intelligence Evolution*, pp. 88–98, 2023.

- [44] S. F. Sung, K. L. Sung, R. C. Pan, P. J. Lee, and Y. H. Hu, "Automated risk assessment of newly detected atrial fibrillation poststroke from electronic health record data using machine learning and natural language processing," *Frontiers in Cardiovascular Medicine*, vol. 9, p. 941237, 2022.
- [45] T. K. K. Mamidi, T. K. Tran-Nguyen, R. L. Melvin, and E. A. Worthey, "Development of an individualized risk prediction model for COVID-19 using electronic health record data," *Frontiers in Big Data*, vol. 4, p. 675882, 2021.
- [46] D. Teoh, "Towards stroke prediction using electronic health records," *BMC Medical Informatics and Decision Making*, vol. 18, pp. 1–11, 2018.
- [47] Y. Zhao *et al.*, "Natural language processing and machine learning for identifying incident stroke from electronic health records: Algorithm development and validation," *Journal of Medical Internet Research*, vol. 23, no. 3, p. e22951, 2021.
- [48] P. Chantamit-O-Pas and M. Goyal, "Long short-term memory recurrent neural network for stroke prediction," in *Machine Learning and Data Mining in Pattern Recognition*, Cham, Switzerland: Springer, 2018, pp. 312–323.
- [49] C. Li *et al.*, "Improving cardiovascular risk prediction through machine learning modelling of irregularly repeated electronic health records," *European Heart Journal-Digital Health*, vol. 5, no. 1, pp. 30–40, 2024.
- [50] P. Tiwari *et al.*, "Assessment of a machine learning model applied to harmonized electronic health record data for the prediction of incident atrial fibrillation," *JAMA Network Open*, vol. 3, no. 1, p. e1919396, 2020.
- [51] C. Ye *et al.*, "Prediction of incident hypertension within the next year: Prospective study using statewide electronic health records and machine learning," *Journal of Medical Internet Research*, vol. 20, no. 1, p. e22, 2018.
- [52] A. Bagheri, T. K. J. Groenhof, W. B. Veldhuis, P. A. de Jong, F. W. Asselbergs, and D. L. Oberski, "Multimodal learning for cardiovascular risk prediction using EHR data," *arXiv preprint arXiv:2008.11979*, 2020.
- [53] X. Wang *et al.*, "Prediction of the 1-year risk of incident lung cancer: Prospective study using electronic health records from the state of Maine," *Journal of Medical Internet Research*, vol. 21, no. 5, p. e13260, 2019.
- [54] A. Ashfaq, A. Sant'Anna, M. Lingman, and S. Nowaczyk, "Readmission prediction using deep learning on electronic health records," *Journal of Biomedical Informatics*, vol. 97, p. 103256, 2019.
- [55] E. Antikainen *et al.*, "Transformers for cardiac patient mortality risk prediction from heterogeneous electronic health records," *Scientific Reports*, vol. 13, no. 1, p. 3517, 2023.
- [56] M. A. Ahmad *et al.*, "Hiding patients' medical reports using an enhanced wavelet steganography algorithm in DICOM images," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 10577–10592, 2022.
- [57] V. R. Falmari, M. Brindha, and S. Ko, "Secure COVID-19 electronic health records management for telediagnosis and travel ticket assistant system using cryptographic approaches," *SN Computer Science*, vol. 3, no. 5, p. 403, 2022.
- [58] C. Li *et al.*, "Efficient medical big data management with keyword-searchable encryption in healthchain," *IEEE Systems Journal*, vol. 16, no. 4, pp. 5521–5532, 2022.
- [59] S. Khalifeh, J. Georgi, and S. Shakhathreh, "Design and implementation of a steganography-based system that provides protection for breast cancer patient's data," in *Proc. 56th Annual Conf. Information Sciences and Systems (CISS)*, Mar. 2022, pp. 19–24.
- [60] S. D. Desai, N. Patil, S. R. Nirmala, S. Kulkarni, P. D. Desai, and D. Shinde, "Deep neural network based medical image steganography," in *Proc. Int. Conf. Smart Technologies and Systems for Next Generation Computing (ICSTSN)*, Mar. 2022, pp. 1–5.
- [61] W. El-Shafai, F. Khallaf, E. S. M. El-Rabaie, and F. E. Abd El-Samie, "Proposed neural SAE-based medical image cryptography framework using deep extracted features for smart IoT healthcare applications," *Neural Computing and Applications*, vol. 34, no. 13, pp. 10629–10653, 2022.
- [62] A. Ali *et al.*, "Deep learning based homomorphic searchable encryption for keyword search in blockchain healthcare system: A novel approach to cryptography," *Sensors*, vol. 22, no. 2, p. 528, 2022.
- [63] G. Georgieva-Tsaneva, G. Bogdanova, and E. Gospodinova, "Mathematically based assessment of the accuracy of protection of cardiac data realized with the help of cryptography and steganography," *Mathematics*, vol. 10, no. 3, p. 390, 2022.
- [64] H. N. AlEisa, "Data confidentiality in healthcare monitoring systems based on image steganography to improve the exchange of patient information using the internet of things," *Journal of Healthcare Engineering*, vol. 2022, no. 1, p. 7528583, 2022.
- [65] H. B. Mahajan *et al.*, "RETRACTED ARTICLE: Integration of Healthcare 4.0 and blockchain into secure cloud-based electronic health records systems," *Applied Nanoscience*, vol. 13, no. 3, pp. 2329–2342, 2023.
- [66] J. K. Oladele *et al.*, "BEHeDaS: A blockchain electronic health data system for secure medical records exchange," *Journal of Computing Theories and Applications*, vol. 3024, pp. 9104, 2024.