Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 7, No.2, April 2025, pp. 02502011-01 ~ 02502011-014
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v7i2.1312

Adaptive Learning Systems for Data Conversion in EHRs through Machine Learning

Janardhan Deepa[1], Jayashree Jayaraman[2*]

[1]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamilnadu, India

[2]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamilnadu, India

[*] jayashree.j@vit.ac.in

Abstract. Healthcare data management has advanced with Electronic Health Records (EHRs), enhancing the efficiency of medical procedures. Machine learning applied to EHRs transitions healthcare from reactive to proactive, supporting the cost-efficiency and sustainability goals of smart cities. However, digitizing medical records introduces security risks, especially from internal threats, necessitating strong detection systems. Research into machine learning techniques, such as decision trees, random forests, and support vector machines (SVMs), shows their effectiveness in detecting EHR breaches. Balancing system usability with patient privacy remains a key challenge amid widespread data sharing. This study highlights SVMs and deep learning models as promising for improving EHR data accuracy, enhancing detection efficiency, and supporting clinical decisions. Despite advancements in AI, deep learning continues to play a crucial role in refining clinical decision systems, including translating EHR data using technologies like natural language processing (NLP). The study provides a qualitative analysis of how deep learning can optimize EHR processes while addressing security and functional challenges.

Keywords: machine learning, transfer learning, deep learning, fine-tuning

(Received 2024-11-30, Accepted 2025-03-19, Available Online by 2025-04-30)

1. Introduction

The use of Electronic Health Records has completely changed how patient data is recorded, maintained, and exchanged in the modern healthcare environment. By leveraging Machine Learning methods, the precision of data conversion in EHRs can be enhanced[1-8]. Improved patient care, faster processes, and more accessibility are all benefits of the digital transformation of medical records. Data should be cleaned to eliminate mistakes and inconsistencies by employing preprocessing methods such as selecting, extracting, and standardizing features[1]. By contrasting several algorithms like Random Forest and Artificial Neural Networks, the most effective ML algorithm for the task can be chosen[1]. In healthcare, patient safety is of utmost importance, and preventing avoidable accidents is a fundamental right[2]. By integrating automated Machine Learning, particularly Artificial Intelligence

(AI), with EHRs, healthcare providers can more effectively detect and predict safety incidents, leading to improved healthcare quality by identifying and mitigating potential risks as shown in fig.1.

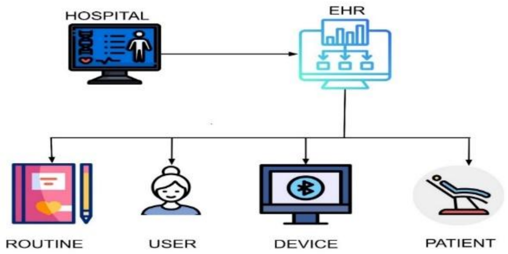

Figure 1. Securing Electronic Health Records and Disaggregation

Utilizing cutting-edge data science methods like deep learning,machine learning and natural language processing, this method was created. Nevertheless, the difficulty is in effectively evaluating the abundance of important data in unstructured EHRs, which calls for automated techniques[9-13].Utilizing feature selection methods alongside classification algorithms can enhance precision of predictive[14 to 22] models by selecting the most relevant features for analysis. Despite the advantages of Electronic Health Records , safeguarding patient privacy remains a challenge. These techniques are particularly relevant in light of potential insider threats to healthcare data. Prior to data analysis, data preprocessing is necessary to ensure accurate and effective results[23 -28].These include: Normalizing numerical variables to ensure they are on a similar scale. [29-37]Changing the analysis-ready format of categorical variables. Filling in missing values with appropriate estimates. The resulting data is divided into multiple filter windows which compute dot-products between input signals and filter kernels preceding the backpropagation part where gradients are backpropagation[37-44].

The article emphasizes the numerous advantages of digital health records , highlighting their growing popularity. These advantages include improved efficiency, accessibility, and communication, ultimately leading to enhanced patient care[45 - 52]. The article underscores the critical role EHRs play in powering machine learning systems that aid in automated decision-making. For instance, predicting hospital readmissions helps minimize associated costs and mortality risks. Non-linear neural networks demonstrate exceptional learning capabilities, enabling them to recognize intricate patterns within data[53 - 61]. This is because non-linear activation functions make it easier for these networks to converge during the training process. Every participant in an EHR system is mandated to interact with the two adjacent EHR systems. Even though the threat paradigm assumes that the Electronic health record systems are nosy but honest, there are cases where the pseudo EHR systems are semihonest (honest-but-curious) and being literrally created.While EHRs have evolved, they remain vulnerable to insider and cyber threats. Therefore, the focus has shifted to safeguarding patient data and protecting it from potential attacks that [62-72]could compromise its integrity and confidentiality. Researchers are currently examining machine learning algorithms that could use EHR data to recognize and predict patient safety issues, with all hospitals participating in the survey looking forward to implementing an EMR system or growing their present ones into EHRs according to researcher’s answers. This is crucial because it simplifies recording and sharing data, then utilizing it to enhance health outcomes.

Throughout the nation, electronic health records are steadily gaining acceptance, and they have the advantage to both patients and nurses. It navigates the intricate balance between harnessing the benefits of Electronic Health Records and proactively addressing the potential risks and challenges associated with their integration into modern healthcare environments.

1.1. Problem statement

Efficient and reliable data conversion is essential for organisations across multiple sectors in today's data-centric world, yet the conventional data conversion approaches often sacrifice precision leading to errors, discrepancies and unproductive data processing procedures.

The aim of this study is to enhance the exactness along with precision in processes utilized during translation. The purpose of the study is to improve upon precision as well as accuracy of conversion processes. These aforementioned constraints may be surmounted through their utilization within ML techniques. Machine Learning techniques are to be used in order to overcome these limitations so as to improve data conversion methods. We aim to design Machine Learning models capable of deciphering the intrinsic sense and organization of multiple data structures to enable them to make sensible decisions during data conversion. This requires training models on labeled data sets which include diverse types as well as ensure stable performance across various situations including those that are extraordinary.

2. Literature Review

We were discussing about different Enhancing Precision of Data Conversion in Electronic Health Records through Machine Learning Methods Rodrigo Tertulino et.al., R[1] suggested the use of ML algorithms for the precision of data conversion. And this paper discuss about the This study is essential as we investigate different machine learning techniques with one of them being compared to in the study. The study ranges from discussing several conditional proxy re-encryption algorithms to how privacy and security are risks with e-health systems Praveen kumar kannoju et.al., R[5] suggest about the radical change in technical approach is needed for efficient healthcare delivery. And the implementation of EHR. The researchers have compared in this article three different machine learning approaches; the Eigen face approach to object recognition, communication encryption using the transmission technique, and a method for reception with less noise.

Jiabao Xu et al., R[8] addressed the issue of low supply of available electronic health records data, absence of substantial datasets, as well as model generalization’s poor performance. Auto-Encoder (AE), Denoising AE, Stacked Denoising AE, Variational AE, Compressed AE, Deep Learning Models (overview), NLP techniques (NER and Relation Extraction) for EHR data extraction are the seven machine learning algorithms addressed in this study.

In this work Zeeshan Ahmed et al., R [31] look at an example where poor management of resources leads to over evaluation of organizational resources there is a need for problems concerning privacy and ethics as far as data is concerned. Aspects pertaining to data ownership, permission processes, and the security associated with information become prevalent within this case study that will also include comparative analysis on three different machine learning approaches; NLP Algorithm (Natural Language Processing), K-Nearest Neighbor-KNN) algorithm, and the Deep Learning Algorithm.

Nikhil Mukund Jagirdar R[40] discuss about the Batch learning algorithms cannot adapt to non-stationary datasets and Batch learning algorithms cannot dynamically adapt to new data arrivals. This paper discuss about the three Machine Learning algorithms they are Incremental random forest algorithm, Adaptive Boosting (AdaBoost) and Online or Incremental Learning Algorithm. The exploration of various algorithms in Electronic Health Records (EHR) management and healthcare analytics has yielded significant advancements, but numerous challenges persist. Rodrigo Tertulino et al. introduced the Conditional Proxy Re-Encryption algorithm to enhance security in electronic health systems, though concerns over privacy remain. Praveen Kumar Kannoju et al. utilized transmission and reception algorithms along with Eigen face recognition, highlighting the necessity for a drastic shift in EHR implementation strategies, addressing static and dynamic data challenges, and improving transmission efficiency. William Hurst et al. employed machine learning models, including Random Forest Decision Trees and Support Vector Machines (SVMs), but faced difficulties in ensuring EHR compatibility with other clinical systems and data extraction processes. Similarly, Wenju Xu et al. proposed enhancements to the PPDARM scheme using homomorphic cryptography to secure EHR item sets, yet existing communication flaws necessitated further refinement. Ghasem Deimazar et al. implemented various machine learning classifiers, such as Support Vector Machines, Recurrent Neural

Networks, and Logistic Regression, but only considered patient safety instances in clinical notes, limiting broader application in healthcare analytics. Meanwhile, Jose Roberto et al. leveraged deep learning models, including Autoencoders and Convolutional Neural Networks (CNNs), successfully predicting heart failure and emergency admissions; however, the generalizability of EHR-based predictions remains a challenge. Thomas Sequist et al. applied multivariable logistic regression but found that primary EHR functionalities were underutilized. Rachana Patil’s work in thematic analysis for data extraction raised concerns over potential privacy breaches and the restrictive nature of inclusion criteria, affecting generalizability. Jiabao Xu et al. explored autoencoders and deep learning for Named Entity Recognition (NER) and relation extraction in EHRs, but their models suffered from insufficient publicly available datasets, hampering generalization. Manohara Pai addressed interoperability challenges by employing security algorithms like SSL and SOAP, yet the lack of semantic data sharing between hospitals and variations in IT infrastructures presented obstacles. Mario Ciampi et al. focused on de-identification and data transformation but limited their proposal to secondary data use, restricting broader applications. Yingtao Luo et al. developed a pseudo-code training algorithm for the CHE method but faced instability in representation learning for diagnosis predictions. Zhichao Zhu et al. introduced a hybrid CRF model using the Viterbi algorithm, showing superior performance through stacked BiLSTM, yet opportunities exist for external knowledge integration to further enhance outcomes. Yuan Su et al. proposed the KeyGen and Enc algorithm for encrypting EHRs, yet issues related to controlling public errors in encryption persist. Joffrey Leevy et al. combined RNN models, LSTM algorithms, CRFs, and rule-based systems for de-identification, but challenges remain in ensuring stepwise de-identification efficiency. Zeeshan Ahmed et al. incorporated NLP, K-Nearest Neighbors (KNN), and deep learning, yet resource mismanagement, ethical concerns, and data privacy issues hinder implementation. Rosario Catelli et al. employed Stochastic Gradient Descent (SGD), struggling with multilingual data scarcity and suboptimal performance of BERT models. Nikhil Mukund Jagirdar utilized incremental learning through Random Forest and AdaBoost, yet batch learning limitations hinder adaptation to non-stationary datasets. Lastly, Zhichao Zhu et al. demonstrated the effectiveness of stacked BiLSTM with the Viterbi algorithm in enhancing context-dependency, yet further refinement with external knowledge is required. Across these studies, it is evident that while algorithms significantly improve EHR processing, ongoing challenges in security, interoperability, privacy, and model adaptability necessitate continued research and innovation in healthcare technology.

3. Challenges in Data Conversion Precision

Precision health is an individualized approach to healthcare that focuses on prevention and early intervention. Several sources of this data are used none, One’s lifestyle, medical history, and genetics, to develop personalized treatments and interventions[26]. Precision health aims to predict and prevent diseases before they develop, rather than simply treating them after they occur, as is commonly done in traditional healthcare systems[37]. Challenges in effectively transforming data from mobile sensors include handling data that: - Changes frequently and contains numerous variables - Shows strong[1] correlations - Comes from different areas and presents different time intervals.

3.1 Complexity of Healthcare Data

It is challenging to use EHR data in precision medicine research because of these issues. Management of permission and ensuring the confidentiality and security of health data during computation, determining the reliability of precision health data, and adhering to legal and ethical criteria are some of the primary problems In relation to privacy and data [5]security for precision medicine. Acquiring accuracy in data conversion is severely hampered by the complex nature of healthcare data, which comprises a variety of formats, terminologies, and coding systems. For Machine Learning techniques to guarantee precise data mapping and conversion into standardised forms, they must successfully negotiate this complexity[29-35].

3.2 Heterogeneous Data Sources

Electronic health records often receive data from various sources, including imaging systems, clinical notes, and diagnostic equipment. However, the diversity of these data sources poses a challenge for developing Machine Learning models[36-39]. Machine Learning tools can analyze and find meaningful patterns in diverse data sources. By doing so, businesses can make better decisions and[4] solve problems more effectively.

3.3 Limited Labelled Data

Machine Learning models, particularly supervised models, require extensive training data with annotations. PH data is mainly stored in two ways, centralized storage or decentralized storage. Nonetheless, there is a risk because when that single server goes down due to failure or attacks by cyber criminals, everything related to it is compromised and this has ripple effects on all systems containing information captured within the very same server[24].In healthcare, assembling large, well-annotated datasets poses challenges

3.4 Data Quality Assurance

The quality of input data directly impacts Machine Learning models used in healthcare. Data with errors, inconsistencies, or missing pieces can skew models and compromise accuracy[21]. A key hurdle to using such data in examining aspects connected to the risk of disease resides in the requirement to determine patients who suffer from illnesses of concern first, a process we will refer to as phenotyping[27].

3.5 Problems with Interoperability in ML Models

We have to properly integrate electronic health records and machine learning models, and it's crucial to address interoperability challenges. This involves ensuring seamless communication between machine learning algorithms and EHR infrastructure[14]. The goal is to enable accurate data transfer and efficient integration without disrupting patient care.

3.6 Explainability and Trustworthiness

Deep learning models in healthcare often face the "black box" problem due to their complex nature. To ensure trust in these models, it is crucial to explain the decision-making process they use[14]. Must be developed with increasing fantastic consideration is directly influenced by how complete and accurate EHR documentation is. This has led to increased studies aimed at enhancing EHR documentation the increase in medical [35]errors related to its application.

3.7 The dynamic nature of healthcare standards

Healthcare standards are constantly evolving due to changes in laws and regulations[11]. To ensure accurate data translation, Machine Learning models must adapt to these evolving requirements.[21]This presents a challenge in maintaining the effectiveness of models over time, as they need to stay up-to-date with the latest standards.

3.8 Recognizing the challenges in analysis of unstructured data

The complexities involved in unstructured data analytics explain why finding efficient frameworks for handling unending pools of suchlike unstructured data is crucial[40-45]. According to the papers, the central theme is how to fuse different forms of data such as imaging, EHR, genomic, and text data to boost predictive modeling and decision-making purposes. According to the papers, in the analysis of complicated healthcare data, Some of the machine learing techniques that are put to use include those of random forests, convolutional neural networks, multilayer perceptrons, and logistic regression. [46-51].

4. Machine Learning Techniques for Improving Precision

Machine Learning methods can be used to identify medical terminology, which is crucial for accurate diagnosis and treatment decisions[3]. Machine Learning plays a crucial role in enhancing patient care, extracting valuable insights from Electronic Health Records , and optimizing healthcare services[29]. Various machine learning techniques are widely employed within the EHR domain. The datasets for Enhancing Precision of Data Conversion in Electronic Health Records through Machine Learning approaches encompass a variety of data formats, including textual and image-based datasets sourced from diverse sources. Textual datasets integrate patient demographics, laboratory test results, clinical records, diagnostic findings, treatment outcomes, symptoms, and medical history. These datasets offer a comprehensive view of data precision, aiding in the development of models capable of identifying the data for conversion precision in below table.

Table 1. Datasets of the EHR

		Datasets		
Ref.no	Year	Size	Types	Features
29	2022	1.In 2017 TAC had two hundred medication labels. 2. Problems with NLP in clinics have become a national craze in Cn2(2) in 2018. 505 discharge summaries	1.structured product 2.unstructured clinical data	1. 4 2. 9
	2024	20001 patients Datasets of the EHR ents details	Unstructured data	9
36	2021	1746 patients	Structured data	25
37	2021	Medicines at the hospital added up to no less than a thousand dollars	Structured data	6
	2021	4413 patient details	structured data	11
13	2020	505 patient discharge summaries	Unstructured clinical data	9

4.1. Detecting anomalies

Detecting anomalies or outliers in a dataset is referred to as anomaly detection. One common technique for this is statistical anomaly detection, which calculates statistical measures such as mean, standard deviation, and z-scores. Instances that differ substantially from the established thresholds are flagged as anomalies[3]. Other information typically found in Electronic Health Record databases (such as patient names, job titles, and departments) is not used in this project due to privacy concerns. This limitation actually improves the effectiveness of the proposed detection methodology by using a smaller dataset. This research focuses on detecting data misuse by insiders, not external unauthorized access[18]. In the EHR dataset, "routine actions" are actions related to accessing patient records, such as viewing assessment forms or medication orders. "Patient user information" is used to identify the patient record being accessed. "Device data" provides information about how the record was accessed, such as the device number used to retrieve the data. In supervised learning, separating (disaggregating) the Electronic Health Record is crucial. These differences need to be identified to establish benchmarks for both typical and abnormal actions[52-59]. The work flow with the electronic health record is shown below in figure 2.

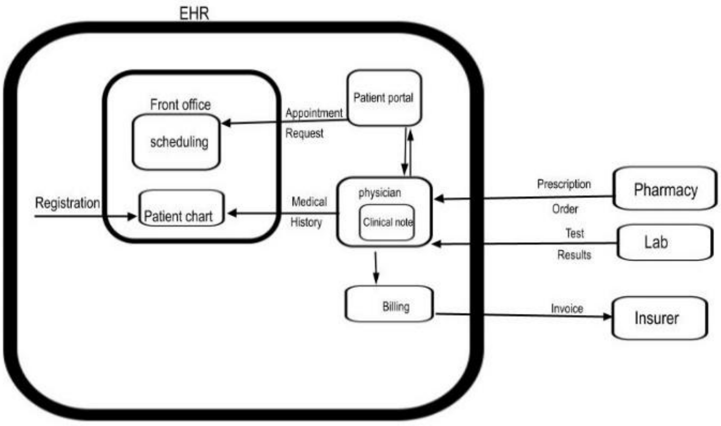

Figure 2. Workflow with the electronic health record system

EHR dataset Two examples of density-based classification techniques include density-based clustering using spatial noise and local outlier factor, help differentiate normal from abnormal behaviors[18]. These algorithms validate anomalous points by referencing the related data record. After labeling, feature extraction involves a detailed analysis of statistical measures, including frequency, mean, mode, standard deviation, and various percentiles[60-64]. Instead of relying on techniques like SMOTE to balance the dataset, the researchers have opted to maintain the imbalanced structure for experimental purposes. The F1score of the algorithms are as follows in table 1

Table 2. F1score of the algorithms .

Ref.no	Entities	F1score
3	Crf	0.84
	LSTM	0.61
	Rnn	0.84
8	Cnn-Rnn	97.2
	Bi-Lstm	96.4
	Cnn	96.0

Classification methodology To identify abnormalities, three methods were chosen:1. Support Vector Machine (SVM): This technique separates data into distinct categories. 2. Random Decision Forest: This involves creating multiple decision trees, each utilizes a haphazard set of attributes; the final projection is derived by amalgamating the results from each tree. 3. decision Tree: This approach involves building a tree-like structure with each node representing a characteristic and each branch representing a decision. SVM is a classifier that finds the best way to separate data points; it's particularly suitable for binary classification (dividing data into two categories) and data with many features[65-72]. In typical scenarios, data is categorized as either "normal" or "abnormal." During testing, these categories are trained using all labeled data, except for extreme values (outliers). By simplifying the task into a binary classification problem (normal vs. abnormal), anomaly detection becomes more manageable. The accuracy of the algorithms are given in table 2.

Table 3. F1 score of the algorithms

Ref.no	Entities	Accuracy
2	Decision tree	0.98
	Random forest	0.98
	SVM	0.96
	Cnn	0.92
18	Federated Learning	0.84
22	Vector Dimension 100	0.87

4.2. Access control

In placing limitations on which patient data to access, the electronic mechanisms that make up Electronic Health Record systems help in safeguarding patient privacy[15]. HIPAA necessitates rules and restrictions as need be to access personal information while permitting a user access health data[68].

4.3. De-Identification

Researchers can use this anonymized data for studies, medical research, and health policy assessments[22]. To perform Named Entity Recognition (NER), which identifies and classifies personal information in PHI, Machine Learning techniques are employed. These techniques follow a typical workflow involving pre-processing, tagging, and post-processing[18]. Our two De-identification systems also adhere to this framework. When working with imperfect data, pre-processing is essential. After pre-processing, entities are tagged using algorithms guided by models or labeled data[54]. Although post-processing is optional, it can significantly enhance system accuracy. Due to time constraints, we have not added the post processing module to our project [22].

Pre-processing In this stage, tokenization or separating sentences from words is focused on. This is because words related to entities (like dates or names) may not be clearly separated from regular words[28]. If the separation is incorrect, the system that identifies entities (tagger) may make mistakes. For instance, in the phrase "09/14/2067CPT Code," "09/14/2067" represents the date, while "CPT Code" doesn't[20]. The point is, some models like Conditional Random Fields (CRFs) and Long Short-Term Memory (LSTMs) recognize the importance of breaking up texts into sentences. [10].

Tagging Unit tags are assigned to entities consisting of a single token. Feature extraction underlies the tagging process, similar to other categorization tasks[18]. While indicator functions are often versatile and efficient, they present challenges as they require manual creation and may need to be modified for different datasets, leading to increased computational costs due to numerous sparse features[22]. Specifically, they have investigated techniques for learning the continuous representation of individual tokens, known as word embeddings or word vectors[18]. Unlike other signals, word embeddings can undergo various operations (translation, rotation, scaling, and superposition), but the exact definitions and significance of these operations remain uncertain[22].

4.4. NLP(Natural Language Processing)

NLP is a field of computer science which focuses on teaching computers how to understand human languages. The medical industry utilizes NLP to gather information from unstructured clinical records such as reports and notes[18]. Named Entity Recognition in Natural Language processing is a task done using various algorithms to process human language and pick out certain information and organic context that supports the information. Currently, the main entity recognition techniques are classified in the order of their development timeline[9].It allows for accurate conversion of unstructured text into structured data. Text categorization methods like Naive Bayes and Support Vector Machines (SVM) can enhance the accuracy. of data conversion by categorizing unstructured text into specific groups or classes. These algorithms enable the efficient identification and extraction of entities from text, enhancing the precision of data conversion[22]. Natural Language processing methods can also assist in improving the quality of data. They can clean up and standardize data, making it more accurate and usable.

4.5.Deep Learning

Deep learning techniques make use of multi-layered neural networks to discover and extract intricate patterns and representations from data[23].Recent times have witnessed an upsurge in research harnessing deep learning to analyze EHRs. Popular deep learning architectures used for[71] this purpose are Feed-Forward Neural Networks (FFNNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs).In a groundbreaking study, Tran and colleagues developed eNRBM (electronic medical records-driven nonnegative restricted Boltzmann machines) to automate feature extraction from medical records[72]. To handle the hierarchical nature of illness coding, eNRBM imposed constraints on structural consistency and positivity[8]. One hand, deep models are able to pick up local patterns and as well as temporal information but that is not the case with logistic regression or random forest which only aggregate time series features meaning they lose some sort of information in process[33].

This extractor has outperformed traditional feature engineering methods in Clinical risk prediction tasks include the prediction of complications in diabetes mellitus, liver and intrahepatic bile duct cancer, anus and rectal cancer, and non-hypertensive congestive heart failure[38].

Neither Self-Developed Architecture (SDA) nor Equi-Norm Representation Boltzmann Machine (eNRBM), including the approach by [40] that divided patients' encounters into distinct time intervals and combined them, effectively incorporated time into modular models. Both eNRBM and SDA overlooked temporal relationships in electronic health records (EHRs)[8]. In SGD, each device updates its model based on a small portion of the training data. It has helped advance deep learning [5]in healthcare applications, including predicting mortality and readmission rates. Deepr shows the fitness journey of a patient as a sequence of medical codes, where each code is transformed into a new space to enable numerical operations and statistical analysis Similar to word embedding in Natural Language processing, Deepr uses "special words" to represent time intervals between events[20]. In situations where sharing raw data is impractical due to resource constraints or privacy issues, another deep learning approach known as SplitNN (Split learning) can be utilized[2]. The analysis model of data has changed from the [41]use of features involved by specialists while engineering to the development driven by the data itself.

5. Conclusion

Machine Learning methods can greatly improve the accuracy of Electronic Health Record data conversion. This can lead to better patient outcomes and personalized healthcare plans. To achieve accurate EHR data conversion, a comprehensive approach is needed that involves technological advancements, careful procedures, and collaboration among stakeholders. This will enhance patient care and the reliability of healthcare data. By leveraging predictive analytics, incorporating algorithms, and promoting collaboration for standardized practices, Machine Learning can significantly improve the precision of EHR data conversion. NLP methods can enhance the accuracy of Electronic Health Record data conversion by converting unstructured text information into organized formats. Healthcare data conversion will become more accurate and reliable with the implementation of advanced Machine Learning techniques, such as adaptive learning systems and semantic interoperability. This comprehensive approach involves both technical and procedural elements. Technical aspects like powerful algorithms, standardization, and error handling are combined with procedural elements like quality assurance and user training. Additionally, continuous monitoring and compliance with regulations are crucial to ensure data conversion accuracy and security in Electronic Health Records.

References

[1] R. Tertulino, N. Antunes, and H. Morais, “Privacy in electronic health records: a systematic mapping study,” Journal of Public Health, Jan. 2023, doi: 10.1007/s10389-022-01795-z. Available: https://doi.org/10.1007/s10389-022-01795-z.

[2] J. Pool, S. Akhlaghpour, F. Fatehi, and A. Burton-Jones, “A systematic analysis of failures in protecting personal health data: A scoping review,” International Journal of Information Management, vol. 74, p. 102719, Feb. 2024, doi: 10.1016/j.ijinfomgt.2023.102719. Available: https://doi.org/10.1016/j.ijinfomgt.2023.102719 .

[3] W. Xu, Q. Zhao, Y. Zhan, B. Wang, and Y. Hu, “Privacy-preserving association rule mining based on electronic medical system,” Wireless Networks, vol. 28, no. 1, pp. 303–317, Jan. 2022, doi: 10.1007/s11276-021-02846-1. Available: https://doi.org/10.1007/s11276-021-02846-1.

[4] J. R. A. Solares et al., “Deep learning for electronic health records: A comparative review of multiple deep neural architectures,” Journal of Biomedical Informatics, vol. 101, p. 103337, Jan. 2020, doi: 10.1016/j.jbi.2019.103337. Available: https://doi.org/10.1016/j.jbi.2019.103337 .

[5] P. K. Kannoju, K. V. Sridhar, and K. S. R. Prasad, “A new paradigm of electronic health record for efficient implementation of health care delivery,” Second International Conference on Intelligent

Systems, Modelling and Simulation., Jan. 2011, doi: 10.1109/isms.2011.28. Available: https://doi.org/10.1109/isms.2011.28.

[6] J. Pool, S. Akhlaghpour, F. Fatehi, and A. Burton-Jones, "A systematic analysis of failures in protecting personal health data: A scoping review," International Journal of Information Management, vol. 74, p. 102719, Feb. 2024, doi: 10.1016/j.ijinfomgt.2023.102719. Available: https://doi.org/10.1016/j.ijinfomgt.2023.102719

[7] W. Xu, Q. Zhao, Y. Zhan, B. Wang, and Y. Hu, "Privacy-preserving association rule mining based on electronic medical system," Wireless Networks, vol. 28, no. 1, pp. 303–317, Jan. 2022, doi: 10.1007/s11276-021-02846-1. Available: https://doi.org/10.1007/s11276-021-02846-1.

[8] J. Xu, X. Xi, J. Chen, V. S. Sheng, J. Ma, and Z. Cui, "A survey of deep learning for electronic health records," *Applied Sciences*, vol. 12, no. 22, p. 11709, Nov. 2022, doi: 10.3390/app122211709. Available: https://doi.org/10.3390/app122211709..

[9] J. Ke, W. Wang, X. Chen, J. Gou, Y. Gao, and S. Jin, "Medical entity recognition and knowledge map relationship analysis of Chinese EMRs based on improved BiLSTM-CRF," *Computers & Electrical Engineering*, vol. 108, p. 108709, May 2023, doi: 10.1016/j.compeleceng.2023.108709. Available: https://doi.org/10.1016/j.compeleceng.2023.108709..

[10] H. Chen, H. Li, G. Xu, Y. Zhang, and X. Luo, "Achieving privacy-preserving federated learning with irrelevant updates over E-Health applications," *IEEE Xplore*, Jun. 2020, doi: 10.1109/icc40277.2020.9149385. Available: https://doi.org/10.1109/icc40277.2020.9149385.

[11] M. M. M. Pai, R. Ganiga, R. M. Pai, and R. K. Sinha, "Standard electronic health record (EHR) framework for Indian healthcare system," Health Services and Outcomes Research Methodology, vol. 21, no. 3, pp. 339–362, Jan. 2021, doi: 10.1007/s10742-020-00238-0. Available: https://doi.org/10.1007/s10742-020-00238-0.

[12] A. Mukherjee, "Implementing Electronic Health Records in India: Status, Issues & Way Forward," Biomedical Journal of Scientific & Technical Research, vol. 33, no. 2, Jan. 2021, doi: 10.26717/bjstr.2021.33.005378. Available: https://doi.org/10.26717/bjstr.2021.33.005378 .

[13] A. O. Ugwu, X. Gao, J. O. Ugwu, and V. Chang, "Ethical Implications of AI in Healthcare Data: A Case Study Using Healthcare Data Breaches from the US Department of Health and Human Services Breach Portal between 2009-2021," IEEE Xplore, Sep. 2022, doi: 10.1109/iiotbdsc57192.2022.00070. Available: https://doi.org/10.1109/iiotbdsc57192.2022.00070.

[14] S. Kohler et al., "Eos and OMOCL: Towards a seamless integration of openEHR records into the OMOP Common Data Model," Journal of Biomedical Informatics, vol. 144, p. 104437, Aug. 2023, doi: 10.1016/j.jbi.2023.104437. Available: https://doi.org/10.1016/j.jbi.2023.104437.

[15] R. Y. Patil, "A secure privacy preserving and access control scheme for medical internet of things (MIoT) using attribute-based signcryption," International Journal of Information Technology, vol. 16, no. 1, pp. 181–191, Oct. 2023, doi: 10.1007/s41870-023-01569-0. Available: https://doi.org/10.1007/s41870-023-01569-0.

[16] T. D. Sequist, T. Cullen, H. Hays, M. M. Taualii, S. R. Simon, and D. W. Bates, "Implementation and Use of an Electronic Health Record within the Indian Health Service," Journal of the American Medical Informatics Association, vol. 14, no. 2, pp. 191–197, Mar. 2007, doi: 10.1197/jamia.m2234. Available: https://doi.org/10.1197/jamia.m2234 .

[17] R. Hoover, "Benefits of using an electronic health record," Nursing Critical Care, vol. 12, no. 1, pp. 9–10, Jan. 2017, doi: 10.1097/01.ccn.0000508631.93151.8d. Available: https://doi.org/10.1097/01.ccn.0000508631.93151.8d.

[18] H. Li *et al.*, "Review on security of federated learning and its application in healthcare," *Future Generation Computer Systems*, vol. 144, pp. 271–290, Jul. 2023, doi: 10.1016/j.future.2023.02.021. Available: https://doi.org/10.1016/j.future.2023.02.021.

[19] K.-L. Tan, C.-H. Chi, and K.-Y. Lam, "Secure and privacy-preserving sharing of personal health records with multi-party pre-authorization verification," *Wireless Networks*, Sep. 2022, doi: 10.1007/s11276-022-03114-6. Available: https://doi.org/10.1007/s11276-022-03114-6.

[20] S. Cao, Q. Zhang, D. Wang, P. Xiangli, and X. Zhang, "Hybrid smart contracts for Privacy-Preserving-Aware Insurance compensation," *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2022, doi: 10.1109/wcnc51071.2022.9771976. Available: https://doi.org/10.1109/wcnc51071.2022.9771976.

[21] M. Ciampi, M. Sicuranza, and S. Silvestri, "A Privacy-Preserving and Standard-Based architecture for secondary use of clinical data," *Information*, vol. 13, no. 2, p. 87, Feb. 2022, doi: 10.3390/info13020087. Available: https://doi.org/10.3390/info13020087.

[22] J. Zhang, J. Zhang, K. Wang, and W. Yan, "Should doctors use or avoid medical terms? The influence of medical terms on service quality of E-health," *Electronic Commerce Research*, vol. 23, no. 3, pp. 1775–1805, Oct. 2021, doi: 10.1007/s10660-021-09516-6. Available: https://doi.org/10.1007/s10660-021-09516-6.

[23] "Y. Luo, Z. Liu, and Q. Liu, "Deep stable representation learning on electronic health records," 2022 IEEE International Conference on Data Mining (ICDM), Nov. 2022, doi: 10.1109/icdm54844.2022.00134. Available: https://doi.org/10.1109/icdm54844.2022.00134.

[24] C. Thapa and S. Camtepe, "Precision health data: Requirements, challenges and existing techniques for data security and privacy," *Computers in Biology and Medicine*, vol. 129, p. 104130, Feb. 2021, doi: 10.1016/j.compbiomed.2020.104130. Available: https://doi.org/10.1016/j.compbiomed.2020.104130.

[25] Z. Zhu, J. Li, Q. Zhao, Y.-C. Wei, and Y. Jia, "Medical named entity recognition of Chinese electronic medical records based on stacked Bidirectional Long Short-Term Memory," *, IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, Jul. 2021, doi: 10.1109/compsac51774.2021.00293. Available: https://doi.org/10.1109/compsac51774.2021.00293.)

[26] "M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, p. n71, Mar. 2021, doi: 10.1136/bmj.n71. Available: https://doi.org/10.1136/bmj.n71.

[27] T. Ferté, S. Cossin, T. Schaeverbeke, T. Barnetche, V. Jouhet, and B. Hejblum, "Automatic phenotyping of electronical health record: PheVis algorithm," *Journal of Biomedical Informatics*, vol. 117, p. 103746, May 2021, doi: 10.1016/j.jbi.2021.103746. Available: https://inria.hal.science/hal-03100435.

[28] "Y. Su, Y. Li, K. Zhang, and B. Yang, "A privacy-preserving public integrity check scheme for outsourced EHRs," Information Sciences, vol. 542, pp. 112–130, Jan. 2021, doi: 10.1016/j.ins.2020.06.043. Available: https://doi.org/10.1016/j.ins.2020.06.043.

[29] "F. Zerka *et al.*, "Systematic Review of Privacy-Preserving Distributed Machine Learning from Federated Databases in Health Care," *JCO Clinical Cancer Informatics*, no. 4, pp. 184–200, Nov. 2020, doi: 10.1200/cci.19.00047. Available: https://doi.org/10.1200/cci.19.00047.

[30] J. L. Leevy, T. M. Khoshgoftaar, and F. Villanustre, "Survey on RNN and CRF models for de-identification of medical free text," *Journal of Big Data*, vol. 7, no. 1, Sep. 2020, doi: 10.1186/s40537-020-00351-4. Available: https://doi.org/10.1186/s40537-020-00351-4..

[31] "Z. Ahmed, K. Mohamed, S. Zeeshan, and X. Dong, "Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine," Database, vol. 2020, Jan. 2020, doi: 10.1093/database/baaa010. Available: https://doi.org/10.1093/database/baaa010 .

[32] "H. Ji, S. Kim, S. Yi, H. Hwang, J.-W. Kim, and S. Yoo, "Converting clinical document architecture documents to the common data model for incorporating health information exchange data in observational health studies: CDA to CDM," *Journal of Biomedical Informatics*, vol. 107, p. 103459, Jul. 2020, doi: 10.1016/j.jbi.2020.103459. Available: https://doi.org/10.1016/j.jbi.2020.103459.

[33] D. Zhang, C. Yin, J. Zeng, X. Yuan, and P. Zhang, "Combining structured and unstructured data for predictive models: a deep learning approach," *medRxiv (Cold Spring Harbor Laboratory)*, Aug.

2020, doi: 10.1101/2020.08.10.20172122. Available: https://doi.org/10.1101/2020.08.10.20172122.
[34] R. Catelli, F. Gargiulo, V. Casola, G. De Pietro, H. Fujita, and M. Esposito, "A novel COVID-19 data set and an effective deep learning approach for the De-Identification of Italian medical records," *IEEE Access*, vol. 9, pp. 19097–19110, Jan. 2021, doi: 10.1109/access.2021.3054479. Available: https://doi.org/10.1109/access.2021.3054479.
[35] L. O. Varela *et al.*, "Evaluation of interventions to improve electronic health record documentation within the inpatient setting: a protocol for a systematic review," *Systematic Reviews*, vol. 8, no. 1, Feb. 2019, doi: 10.1186/s13643-019-0971-2. Available: https://doi.org/10.1186/s13643-019-0971-2..
[36] R. H. Epstein, F. Dexter, and E. S. Schwenk, "Provider access to legacy electronic anesthesia records following implementation of an electronic health record system," *Journal of Medical Systems*, vol. 43, no. 5, Mar. 2019, doi: 10.1007/s10916-019-1232-6. Available: https://doi.org/10.1007/s10916-019-1232-6.
[37] C. A. Nelson, A. J. Butte, and S. E. Baranzini, "Integrating biomedical research and electronic health records to create knowledge based biologically meaningful machine readable embeddings," *bioRxiv (Cold Spring Harbor Laboratory)*, Feb. 2019, doi: 10.1101/540963. Available: https://doi.org/10.1101/540963..
[38] D. A. Kumar and A. Chinnalagu, "Sentiment and Emotion in Social Media COVID-19 Conversations: SAB-LSTM Approach," *IEEE Xplore*, Dec. 2020, doi: 10.1109/smart50582.2020.9337098. Available: https://doi.org/10.1109/smart50582.2020.9337098.
[39] Y. Yang, X. Zheng, W. Guo, X. Liu, and V. Chang, "Privacy-preserving fusion of IoT and big data for e-health," *Future Generation Computer Systems*, vol. 86, pp. 1437–1455, Sep. 2018, doi: 10.1016/j.future.2018.01.003. Available: https://doi.org/10.1016/j.future.2018.01.003.
[40] N. M. Jagirdar, "Online Machine Learning Algorithms Review and Comparison in Healthcare," *University of Tennessee, Knoxville TRACE: Tennessee Research and Creative Exchange.*, Jan. 2018, Available: https://trace.tennessee.edu/cgi/viewcontent.cgi?article=6803&context=utk_gradthes.
[41] C. Xiao, E. Choi, and J. Sun, "Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review," Journal of the American Medical Informatics Association, vol. 25, no. 10, pp. 1419–1428, Jun. 2018, doi: 10.1093/jamia/ocy068. Available: https://doi.org/10.1093/jamia/ocy068.
[42] S. K. Nayak and S. Tripathy, "Privacy Preserving Provable Data Possession for Cloud Based Electronic Health Record System," *2016 IEEE.*, Aug. 2016, doi: 10.1109/trustcom.2016.0149. Available: https://doi.org/10.1109/trustcom.2016.0149.
[43] J. Zhao, P. Papapetrou, L. Asker, and H. Boström, "Learning from heterogeneous temporal data in electronic health records," *Journal of Biomedical Informatics*, vol. 65, pp. 105–119, Jan. 2017, doi: 10.1016/j.jbi.2016.11.006. Available: https://doi.org/10.1016/j.jbi.2016.11.006.
[44] Y. Chen, S. Nyemba, W. Zhang, and B. Malin, "Leveraging social networks to detect anomalous insider actions in collaborative environments," , *IEEE Xplore*, Jul. 2011, doi: 10.1109/isi.2011.5984061. Available: https://doi.org/10.1109/isi.2011.5984061.
[45] Z. Jiang, C. Zhao, B. He, Y. Guan, and J. Jiang, "De-identification of medical records using conditional random fields and long short-term memory networks," Journal of Biomedical Informatics, vol. 75, pp. S43–S53, Nov. 2017, doi: 10.1016/j.jbi.2017.10.003. Available: https://doi.org/10.1016/j.jbi.2017.10.003.
[46] S. El-Gendy, M. S. Elsayed, A. Jurcut, and M. A. Azer, "Privacy preservation using machine learning in the internet of things," *Mathematics*, vol. 11, no. 16, p. 3477, Aug. 2023, doi: 10.3390/math11163477. Available: https://doi.org/10.3390/math11163477.
[47] J. G, S. R, G. H. L, V. Ravi, M. Almeshari, and Y. Alzamil, "Electronic Health Record (EHR) system development for study on EHR data-based early prediction of diabetes using machine

learning algorithms," *The Open Bioinformatics Journal*, vol. 16, no. 1, Oct. 2023, doi: 10.2174/18750362-v16-e230906-2023-15. Available: https://doi.org/10.2174/18750362-v16-e230906-2023-15.

.

[48] A. Kline et al., "Multimodal machine learning in precision health: A scoping review," Npj Digital Medicine, vol. 5, no. 1, Nov. 2022, doi: 10.1038/s41746-022-00712-8. Available: https://doi.org/10.1038/s41746-022-00712-8.

[49] K. V. Kanimozhi and Dr. M. Venkatesan, "Unstructured Data Analysis-A Survey," *IJARCCE*, pp. 223–225, Mar. 2015, doi: 10.17148/ijarcce.2015.4354. Available: https://doi.org/10.17148/ijarcce.2015.4354.

.

[50] M. Zheng, H. Xue, Y. Zhang, T. Wei, and J. C. S. Lui, "Enpublic Apps," *Computer Science Raleigh, North Carolina.*, Apr. 2015, doi: 10.1145/2714576.2714593. Available: https://doi.org/10.1145/2714576.2714593.

.

[51] L. M. Dang, Md. J. Piran, D. Han, K. Min, and H. Moon, "A survey on internet of things and cloud computing for healthcare," *Electronics*, vol. 8, no. 7, p. 768, Jul. 2019, doi: 10.3390/electronics8070768. Available: https://doi.org/10.3390/electronics8070768.

[52] P. Tasatanattakool and C. Techapanupreeda, "User authentication algorithm with role-based access control for electronic health systems to prevent abuse of patient privacy," 3rd IEEE International Conference on Computer and Communications (ICCC)., Dec. 2017, doi: 10.1109/compcomm.2017.8322697. Available: https://doi.org/10.1109/compcomm.2017.8322697.

[53] B. F. Smaradottir, "Security Management in Electronic Health Records: Attitudes and Experiences Among Health Care Professionals," *2018 International Conference on Computational Science and Computational Intelligence (CSCI).*, Dec. 2018, doi: 10.1109/csci46756.2018.00143. Available: https://doi.org/10.1109/csci46756.2018.00143.

[54] Q. Mamun and M. Rana, "A robust authentication model using multi-channel communication for eHealth systems to enhance privacy and security," *Electronics and Mobile Communication Conference (IEMCON).*, Oct. 2017, doi: 10.1109/iemcon.2017.8117210. Available: https://doi.org/10.1109/iemcon.2017.8117210.

[55] A. Odeh, I. Keshta, A. Aboshgifa, and E. Abdelfattah, "Privacy and security in mobile health Technologies: Challenges and concerns," *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, Jan. 2022, doi: 10.1109/ccwc54503.2022.9720863. Available: https://doi.org/10.1109/ccwc54503.2022.9720863.

[56] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, Jan. 2018, doi: 10.1109/access.2017.2788044. Available: https://doi.org/10.1109/access.2017.2788044.

[57] X. Xu, T. Wang, Y. Yang, L. Zuo, F. Shen, and H. T. Shen, "Cross-Modal attention with semantic consistence for Image–Text matching," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 12, pp. 5412–5425, Dec. 2020, doi: 10.1109/tnnls.2020.2967597. Available: https://doi.org/10.1109/tnnls.2020.2967597.

[58] J. Wallcraft et al., "Partnerships for better mental health worldwide: WPA recommendations on best practices in working with service users and family carers," World Psychiatry, vol. 10, no. 3, pp. 229–236, Oct. 2011, doi: 10.1002/j.2051-5545.2011.tb00062.x. Available: https://doi.org/10.1002/j.2051-5545.2011.tb00062.x .

[59] J. Collmann and T. Cooper, "Breaching the Security of the Kaiser Permanente Internet Patient Portal: the Organizational Foundations of Information Security," *Journal of the American Medical Informatics Association*, vol. 14, no. 2, pp. 239–243, Mar. 2007, doi: 10.1197/jamia.m2195. Available: https://doi.org/10.1197/jamia.m2195."

[60] Y. Dong, X. Chen, L. Shen, and D. Wang, "Privacy-Preserving distributed machine learning based on secret sharing," in *Lecture notes in computer science*, 2020, pp. 684–702. doi: 10.1007/978-3-030-41579-2_40. Available: https://doi.org/10.1007/978-3-030-41579-2_40.
[61] T. Wu, C. Zhao, and Y.-J. A. Zhang, "Privacy-Preserving distributed optimal power flow with partially homomorphic encryption," IEEE Transactions on Smart Grid, vol. 12, no. 5, pp. 4506–4521, Sep. 2021, doi: 10.1109/tsg.2021.3084934. Available: https://doi.org/10.1109/tsg.2021.3084934.
[62] S. Meftah, B. H. M. Tan, C. F. Mun, K. M. M. Aung, B. Veeravalli, and V. Chandrasekhar, "DOReN: Toward Efficient Deep Convolutional Neural Networks with Fully Homomorphic Encryption," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 3740–3752, Jan. 2021, doi: 10.1109/tifs.2021.3090959. Available: https://doi.org/10.1109/tifs.2021.3090959.
[63] J. Zhang, H. Liu, and L. Ni, "A secure Energy-Saving communication and encrypted storage model based on RC4 for EHR," *IEEE Access*, vol. 8, pp. 38995–39012, Jan. 2020, doi: 10.1109/access.2020.2975208. Available: https://doi.org/10.1109/access.2020.2975208.
[64] K. Rabieh, K. Akkaya, U. Karabiyik, and J. Qamruddin, "A secure and cloud-based medical records access scheme for on-road emergencies," *15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, Jan. 2018, doi: 10.1109/ccnc.2018.8319175. Available: https://doi.org/10.1109/ccnc.2018.8319175.
[65] M. F. F. Khan and K. Sakamura, "A distributed approach to delegation of access rights for electronic health records," *2020 International Conference on Electronics, Information, and Communication (ICEIC)*, Jan. 2020, doi: 10.1109/iceic49074.2020.9051092. Available: https://doi.org/10.1109/iceic49074.2020.9051092.
[66] Y. Zhang, C. Xu, X. Liang, H. Li, Y. Mu, and X. Zhang, "Efficient Public Verification of Data Integrity for Cloud Storage Systems from Indistinguishability Obfuscation," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 3, pp. 676–688, Mar. 2017, doi: 10.1109/tifs.2016.2631951. Available: https://doi.org/10.1109/tifs.2016.2631951.
[67] I. S. Abdulhameed, I. Al-Mejibli, and J. R. Naif, "Improving security and privacy for health information and images," *Webology*, vol. 19, no. 1, pp. 2435–2457, Jan. 2022, doi: 10.14704/web/v19i1/web19164. Available: https://doi.org/10.14704/web/v19i1/web19164.

[68] S. K. Nayak and S. Tripathy, "Privacy Preserving Provable Data Possession for Cloud Based Electronic Health Record System," *IEEE Trustcom/ BigDataSE/ISPA, Aug 2016, Pp. 860–867.*, Aug. 2016, doi: 10.1109/trustcom.2016.0149. Available: https://doi.org/10.1109/trustcom.2016.0149.
[69] S. Chenthara, K. Ahmed, H. Wang, and F. Whittaker, "Security and Privacy-Preserving challenges of e-Health solutions in cloud Computing," *IEEE Access*, vol. 7, pp. 74361–74382, Jan. 2019, doi: 10.1109/access.2019.2919982. Available: https://doi.org/10.1109/access.2019.2919982.
[70] P. C. Tang and C. J. McDonald, "Electronic health record systems," in *Computers in health care*, 2006, pp. 447–475. doi: 10.1007/0-387-36278-9_12. Available: https://doi.org/10.1007/0-387-36278-9_12.
[71] J. Vijayashree and H. P. Sultana, "A Machine Learning Framework for Feature Selection in Heart Disease Classification Using Improved Particle Swarm Optimization with Support Vector Machine Classifier," Programming and Computer Software, vol. 44, no. 6, pp. 388–397, Nov. 2018, doi: 10.1134/s0361768818060129. Available: https://doi.org/10.1134/s0361768818060129.
[72] J. Vijayashree and H. P. Sultana, "Heart disease classification using hybridized Ruzzo-Tompa memetic based deep trained Neocognitron neural network," Health and Technology, vol. 10, no. 1, pp. 207–216, Jan. 2019, doi: 10.1007/s12553-018-00292-2. Available: https://doi.org/10.1007/s12553-018-00292-2.