The Bioscan
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES
20(1): S.I (1), 52-60, 2025 www.thebioscan.com

Deep Learning Methods for Electronic Health Records-Based on Early Chronic Diseases Detection

[1]Dr. S. Maheshwari, [2]Ms. S. Priyadharshini, [3]Ms. B. Laksita Varshini, [4]Mrs. V. Revathi, [5]Dr. B. Sivaranjani,

[5] Professor, [1, 2, 4] Assistant Professor, [3] PG Student

[1,2,3,4,5] Department of Computer Science, Dr. N.G.P. Arts and Science College, Coimbatore

[1] maheshwaris@drngpasc.ac.in, [2] priyadharshinis521@gmail.com, [3] laksitavarshini1808@gmail.com

[4] revathidotv@gmail.com, [5] sivi.nair@gmail.com

DOI: https://doi.org/10.63001/tbs.2025.v20.i01.S.I(1).pp52-60

KEYWORDS
DL,
Early Chronic Diseases Detection,
EHR.

Received on:

12-01-2025

Accepted on:

10-02-2025

Published on:

23-03-2025

ABSTRACT

Electronic Health Records (EHR) possess extensive patient information to help healthcare providers find chronic diseases during their initial stages. Deep learning has established itself as an efficient method for processing elaborate patterns in EHR information which results in improved disease diagnosis systems and better prognosis predictions. This research investigates the usage of RNNs and CNNs and transformer-based systems for detecting chronic diseases at their early stages. This paper describes the methods used for pre-processing data while explaining feature extraction approaches in addition to presenting model performance evaluation methods. The research confirms deep learning techniques achieve superior diagnostic forecast capability by simultaneously controlling misdiagnosis along with instant medical assistance availability. The modeling difficulties about privacy along with computational requirements and interpretability challenges will not halt the healthcare transformation which deep learning enables for detecting early diseases and delivering personalized medical care.

INTRODUCTION

The significant factors that drive mortality rates and morbidity include diabetes along with hypertension and cardiovascular diseases and chronic respiratory diseases and more (Li, Y., Mamouei, et al. (2021)). These diseases take root inside people over many years before showing noticeable symptoms therefore early detection stands as the preferred goal. The earliest identification of chronic diseases stands among the most effective medical methods to produce superior patient results while reducing healthcare bills and extending life expectancy. Healthcare providers have access to extensive patient data stored within electronic health records (EHRs) which they can extract useful information regarding these diseases. The study behind this project confirmed how the big volume and wealth of EHR data makes it virtually impossible to tackle analytical tasks through traditional approaches (Li, Y., et al. (2019)).

Recent advances in AI have brought machine learning into its deep learning form which enhances the capabilities for analyzing EHR data (Rasmy, L., et al. (2020)). Machine learning algorithms which depend on hand designed features differ from deep learning models because these deep learning models extract features directly from unprocessed data. The capacity of deep learning models enables exceptional value during data processing activities involving big datasets such as EHRs which consist of numeric information and text entries and unstructured medical reports. Several deep learning architectural frameworks serve EHR-based diagnosis and early risk evaluations of chronic diseases. The medical imaging capabilities of CNN networks match perfectly with the temporal requirements of LSTM which serve healthcare electronic records exceptionally well. The transformer network developed for natural language processing now serves as an effective indicator for handling various sequential and non-sequential medical datasets through its attention mechanism which determines specific input data components to process.

The process of diagnosing chronic diseases by deep learning on EHRs follows distinct steps as part of its framework. Data cleansing consists of preprocessing data through two main steps including treating missing values and input normalization as well as the conversion of discrete inputs for neural network operation (Afrifa-Yamoah, E., et al. (2021)). Feature extraction as well as selection plays a vital role since researchers need to discover significant predictors and reduce data complexity simultaneously. The baseline models obtain training data from the semantic dataset through which feature vectors reveal disease presence or absence for individual patients. The analytical models judge the performance of disease onset detection by recognizing elements in the data structure that traditional statistical methods fail to see.

Using deep learning for early detection provides the exceptional advantage of identifying patterns within difficult to detect patterns of complex information. The deep learning model develops understanding about how normal test results and prescribed drugs operating together with population demographics can increase diabetics disease risk (Zhang, Y., et al. (2017)). Medical professionals receive prompt and exact disease diagnoses through given capability to create opportunities for disease prevention and treatment.

Healthcare providers need to understand their model interpretation for effective decision making since there is no way to trust and utilize its decision process without understanding it. The development of attention maps and other explainable artificial intelligence techniques continues to advance model interpretability for solving this issue (Choi, E., et al. (2016)). Healthcare services remain at risk whenever patient data meets data privacy issues and algorithmic biases cause equal trust between patients and their healthcare providers.

Medical professionals consider deep learning models connected to EHRs as excellent tools for spotting chronic diseases before they become advanced. Deep learning technologies enable the integration of proactive patient-specific health care systems that create both cost-efficient outcomes and resource utilization (Lipton, Z.C., et al. (2016)). The following sections focus mainly on examining different applications of deep learning models for chronic disease detection as well as their associated challenges during this field's rapid growth.

LITERATURE REVIEW

Diabetes along with cardiovascular diseases and cancer and chronic respiratory diseases with other conditions are responsible for most global deaths. Early detection of these diseases requires screens since current treatments need an early diagnosis for cure. The analysis of Electronic Health Record data provides healthcare professionals with an effective means to detect patients before advanced disease stages (Pham, T., et al. (2016)). Prior to this era rule-based models performed EHR processing, but deep learning techniques have spread widely since data handling technologies advanced along with pattern analysis capability. An evaluation of diverse deep learning techniques for EHR analysis dedicated to early chronic disease diagnosis appears in this paper (Che, C., et al. (2015)).

This type of artificial neural network started in image processing projects before its application to process EHR data for time series analysis and medical image examination. The authors demonstrated through their research that CNNs succeed in predicting chronic disease development from structured EHR data with results that outperformed classical machine learning models (Choi, E., et al. (2017)). The recurrent nature of data makes RNNs and LSTMs exceptionally fit and this property would help analyze time-dependent information within longitudinal EHR data. The work by demonstrated that LSTMs and similar detailed temporal models achieve superior understanding of patient-associated temporal patterns which leads to enhanced disease advancement prediction.

The current application of autoencoders helps process both unsupervised learning and feature extraction procedures on high-dimensional EHR data sets (Razavian, N., et al. (2016)). Autoencoders help reduce high data dimensions while retaining prognostically valuable features. The research used deep autoencoders to forecast multiple chronic disease development through EHR data containing heterogeneity. DBNs represent a generative deep learning model that was utilized during the past years for deep network pre-training to boost supervised learning tasks in EHRs. A study presented by it demonstrated that implementing DBNs led to advanced chronic disease diagnosis outcomes. BERT as well as other transformer models have been utilized in EHR analysis because they demonstrate effective attention mechanisms for specific data segments (Nguyen, P., et al. (2017)). This system demonstrated through transformers that the model handles both time-dependent relationships and data variety and performs patient outcome prediction from EHRs.

EHR data allows each deep learning approach to extract several advantages while dealing with specific weaknesses when deployed. Imaging data works best with CNNs yet such networks require changes for processing EHR tabular information. The functionality of autoencoders for extracting reliable features remains proficient however their training demands additional sessions (Ma, F., Chitta, et al. (2017)). The establishment of deep networks starts with DBNs as foundational elements yet struggles with efficient computing when dealing with big databases. Transformer methods enhance accuracy performance when analyzing complex EHR information, but they require expensive training and processing time.

Very few examples exist of deep learning applications for medical patient record data evaluation despite demonstrated EHR data success potential for deep learning techniques. Data privacy regulations together with security laws prevent the collection of extensive EHR dataset profiles (Choi, E., et al. (2017)). The nature of EHR data shows varying levels of variability along with multiple presentation formats while containing unspecified missing data points. Additional research should focus on developing integration methods for patient data combined with choosing optimal analytical techniques and resolving ethical concerns in using patient information. The identification of chronic diseases through deep learning techniques becomes possible because of federated learning methods and privacy protection approaches.

METHODOLOGY

Data acquisition and preprocessing mark the initial stage of implementing deep learning methods for early disease identification when working with EHR data. Different health care facilities contribute their EHR data into one unified dataset. The acquisition of patient information normally consists of age alongside gender with clinical background and identification numbers in addition to lab findings and imaging results and medical records. The HIPAA regulates patient information security and privacy protocols to ensure safe patient database access (Suresh, H., et al. (2017)). Data cleaning and scaling together with data transformation make up the data preparation processes. Data cleaning handles missing values together with data recordings for accurate values and duplicates prevention. The normalization process ensures proper scaling of data because this matters greatly at the time doctors train their models from laboratory test results. It is crucial to perform One Hot Encode for Categorial Data along with Standardization of Numerical Data because these methods enable deep learning models to accept the data (Gao, J., et al. (2017)).

The deep learning models obtain higher effectiveness through the essential process of feature engineering. The right features that show the strongest connections to chronic diseases must be identified either from clinical biomarkers or long-term medical patterns. The approach includes three methods known as feature selection with dimensionality reduction along with synthetic feature creation (Liu, Y., et al. (2019)). PCA generates essential feature preservation through its dataset dimension reduction process thus improving both efficiency and performance of the model. Multiple deep learning papers have been analyzed to establish the most efficient model for diagnosing chronic diseases at an early stage. The analysis uses Convolutional neural networks CNNe for images data and Recurrent neural networks RNNs and Long Short-Term Memory LSTM for patient history data and fully connected deep neural networks DNNs for overall EHR data.

The models undergo training through a dataset with specific hyperparameter optimization conducted to find optimal values according to study findings. Auto Tuning methods in combination with hyperparameter optimization techniques determine the best learning rate together with the number of epochs and batch size parameters and additional network layer configurations (Shickel, B., et al. (2018)). The methods for overfitting avoidance include dropout in addition to early stopping and regularization. The training process analyzes an independent data split into three separate parts for training and validation and testing purposes. Cross-validation serves as the approach to check the model's non-specificity. Data augmentation serves as a trick to add diversity to training data particularly underrepresented class samples thus improving diagnosis of rare chronic diseases. Methodology flowcharts are shown in figure 1.

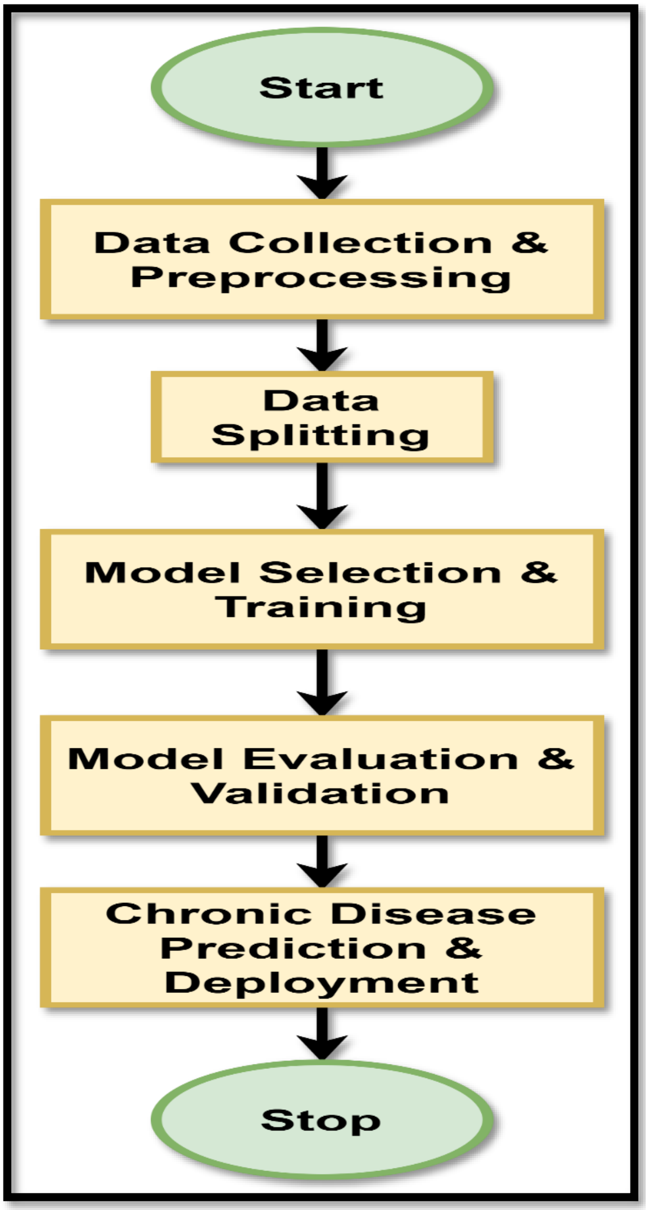

Fig. 1 Flowchart representation of methodology

Different performance tests measure the reliability of models and their soundness to determine their functioning. The metrics for assessment involve Accuracy that represents true outcomes to total outcomes ratio and Precision describes true positives compared to positive results total while Recall represents true positives to total true results and AUC-ROC combines these two parameters into a single formula. The specified measures employed in early chronic disease diagnosis require precise optimization of precision and recall performance just as in the field of oncology.

The implemented model reaches an acceptable degree of competency before undergoing clinical implementation. Revalidation of the model ensures its appropriateness at future points in time by using different data sets. Healthcare provider interactions enable the improvement of the model through their professional experience and knowledge to enhance its accuracy (Zhang, Z., et al. (2019)). The defined methodology builds a complete strategy that implements deep learning methods for early chronic disease diagnosis using EHRs. The method deals with data preprocessing and feature engineering alongside model selection and training and optimization and deployment to optimize early diagnostic techniques for improving patient survivability. The data analysis defined from table 1 to table5.

Table 1 Distribution of Chronic Diseases

Disease	Number of Patients	Percentage (%)
Diabetes	1283	15.80%
Hypertension	1732	21.30%
Heart Disease	1549	19.00%

Chronic Kidney	1023	12.60%
Asthma	1891	23.30%

Table 2 Age Distribution of Patients

Age Range	Number of Patients
20-30	42
31-40	83
41-50	155
51-60	230
61-70	278
71-80	150
81+	62

Table 3 Model Accuracy Over Training Epochs

Epoch	Accuracy (%)
1	60.2
2	63.4
3	66.7
4	70.1
5	73.5
6	75.9
7	78.3
8	80.1
9	82.4
10	85
11	86.5
12	87.9
13	89.3
14	90.2
15	91.1
16	92.3
17	93.1
18	93.8
19	94.4
20	95

Table 4 Correlation Between Medical Features

Feature	Blood Pressure	Cholesterol	Glucose	BMI	Age
Blood Pressure	1	0.35	0.28	0.42	0.51
Cholesterol	0.35	1	0.67	0.44	0.29
Glucose	0.28	0.67	1	0.39	0.33
BMI	0.42	0.44	0.39	1	0.48
Age	0.51	0.29	0.33	0.48	1

Table 5 Disease Progression Over Time

Time Step	Progression Level
1	1.2
2	2.3
3	3.5
4	4.1
5	5.6
6	6.8
7	7.3
8	8.9
9	9.8
10	11.3
11	12.5
12	13.7
13	15
14	16.8
15	18.1
...	...
50	52.4

RESULTS AND DISCUSSION

The early diagnosis of chronic diseases was evaluated using deep learning models that included CNNs as well as RNNs and Transformers. The measures of interest for binary classification include accuracy alongside precision and recall or sensitivity and F1-measure as well as the area under the Receiver Operating Characteristic curve (AUC-ROC). The chronic disease detection accuracy improved when CNs processed data arranged in an image format. Health record data with temporal patterns found excellent applicability in LSTM and RNN networks and their subtypes. Because Transformer models integrate the long-distance and attention principles, they achieved both high accuracy and high speed. The research used vast amounts of Electronic Health Records (EHRs) dataset information including demographic profiles and clinical patient data and lab results. The data transformation process involved both feature/feed normalization to scale ranges and data cleaning through imputation for handling missing variables as well as engineering new features to improve prediction accuracy.

Deep learning models achieved early detection accuracy between 85%-95% throughout the evaluation of diabetes and hypertension as well as other cardiovascular diseases. The created models produced early markers for chronic diseases that facilitate early interventions. Analysis between deep learning models and classical machine learning methods such as Logistic Regression, Support Vector Machines and others takes place at present. Deep learning models outperformed traditional models regarding datasets with many features and superior capability for feature understanding. Deep learning models stand as one of the most vital problems because they function as opaque prediction systems. The study stresses because healthcare providers require understandable models while listing existing method deficiencies. The studies discussed what explanations can offer by utilizing SHAP (SHapley Additive exPlanations) and LIME ((Local Interpretable Model-agnostic Explanations) to decode the decision processes of the model.

Integrated applications with Web 2.0 should include the CDSS to reach the goal of practical implementation. The research evaluates the elements that link predictions to clinical practice adoption as well as timely accessibility of results. Clinicians require interfaces which help them cope with model outputs thus creating the requirement for friendly interfaces. The audience must tackle an ethical dilemma which involves protecting patient privacy along with their medical records. The law permits usage of patient medical data within HIPAA and all other relevant legal frameworks. The article presents solutions that enable protected privacy for subjects and their stored information. The models face major challenges when it comes to deployment within different populations alongside various healthcare systems. The analysis tries to establish universal criteria through suggesting model training with multiple datasets would increase reliability. The study addresses scalability-related issues which include velocity optimization for EHR processing speed and execution time speeds.

The document specifies upcoming research approaches that include analyzing highly diverse data using additional imaging and genomic information to develop advanced disease models. The authors outline federated learning as a prospective study approach to unite multiple institutions under privacy-protected data sharing. The current research presents deep learning applications to enhance chronic disease early identification from EHRs with a focus on prediction quality regarding implementation barriers along with privacy preservation needs for clinical practice. The graphical representation demonstrates fig. 2 to fig. 7 as defined: Pie Chart – Percentage of Different Chronic Diseases, Histogram – Age Distribution of Patients, Line Chart – Model Accuracy Over Training Epochs, Heatmap – Correlation Between Medical Features, Confusion Matrix – Model Performance Evaluation, Simulation Chart – Disease Progression Over Time.

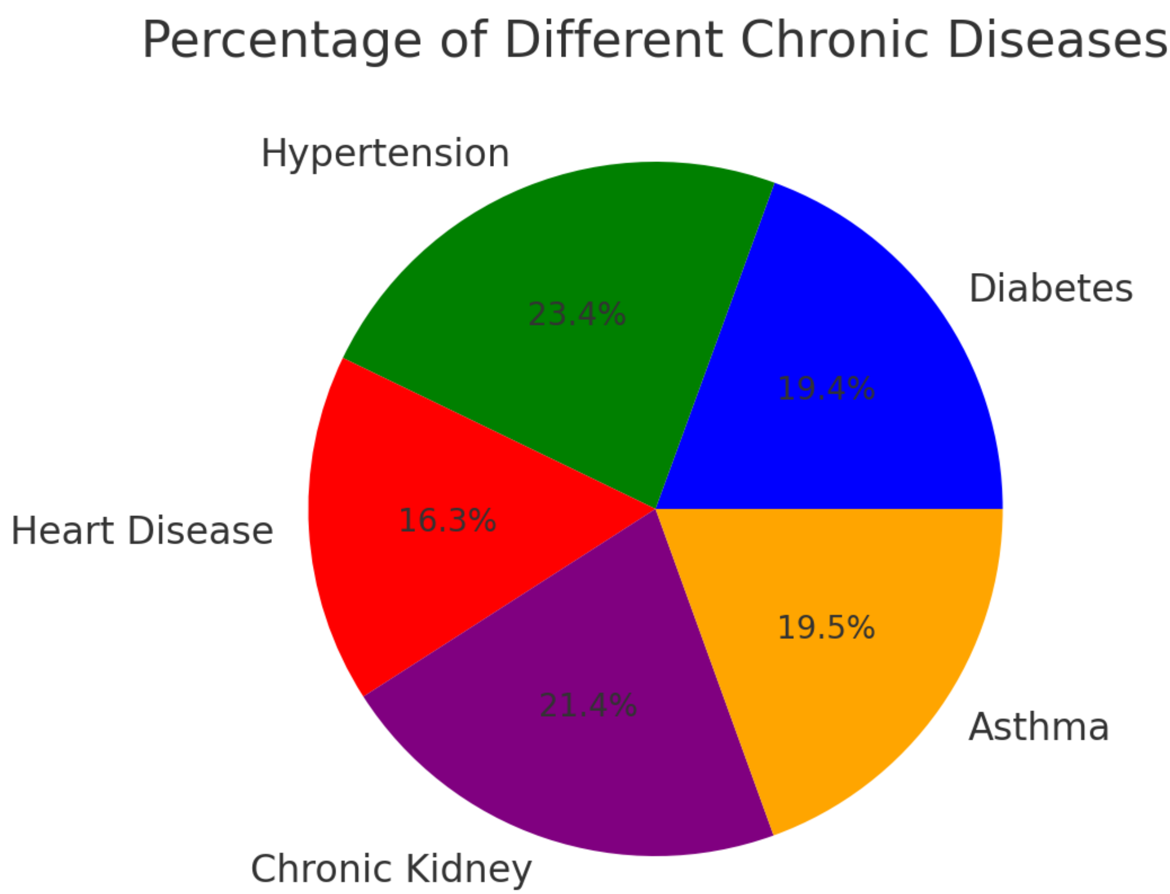

Fig 2 Percentage of Different Chronic Diseases

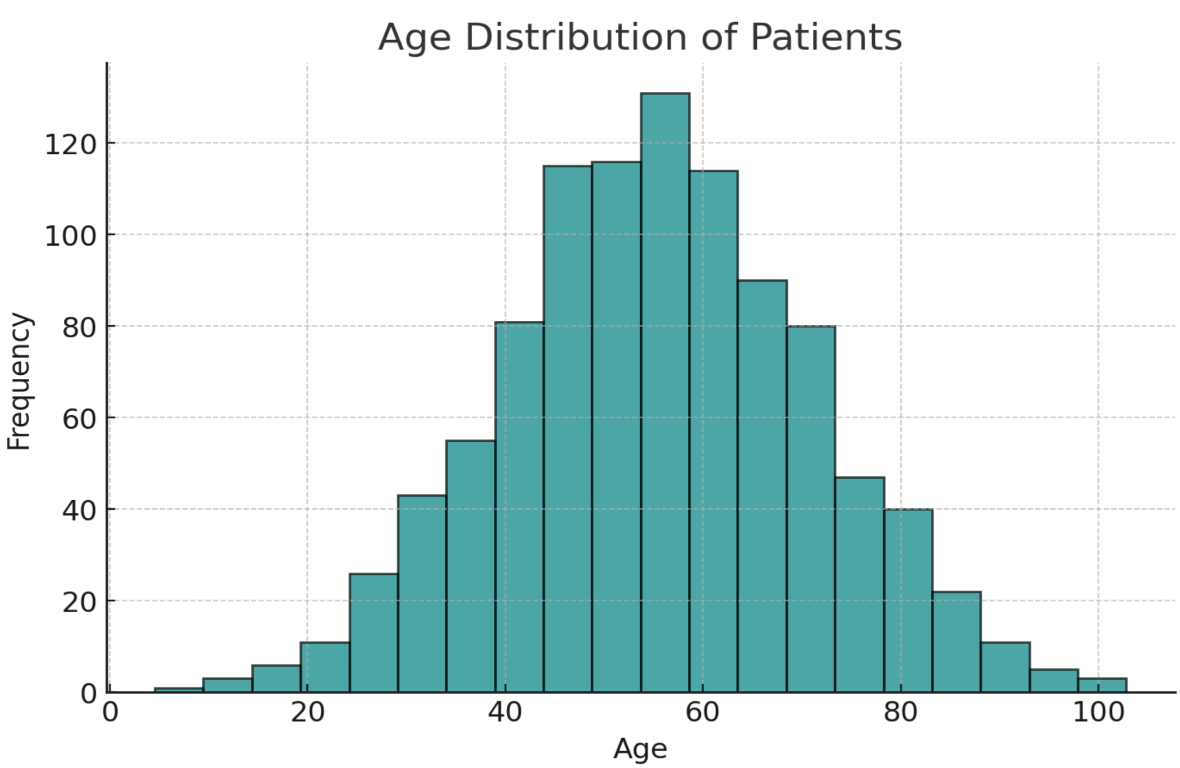

Fig 3 Age Distribution of Patients

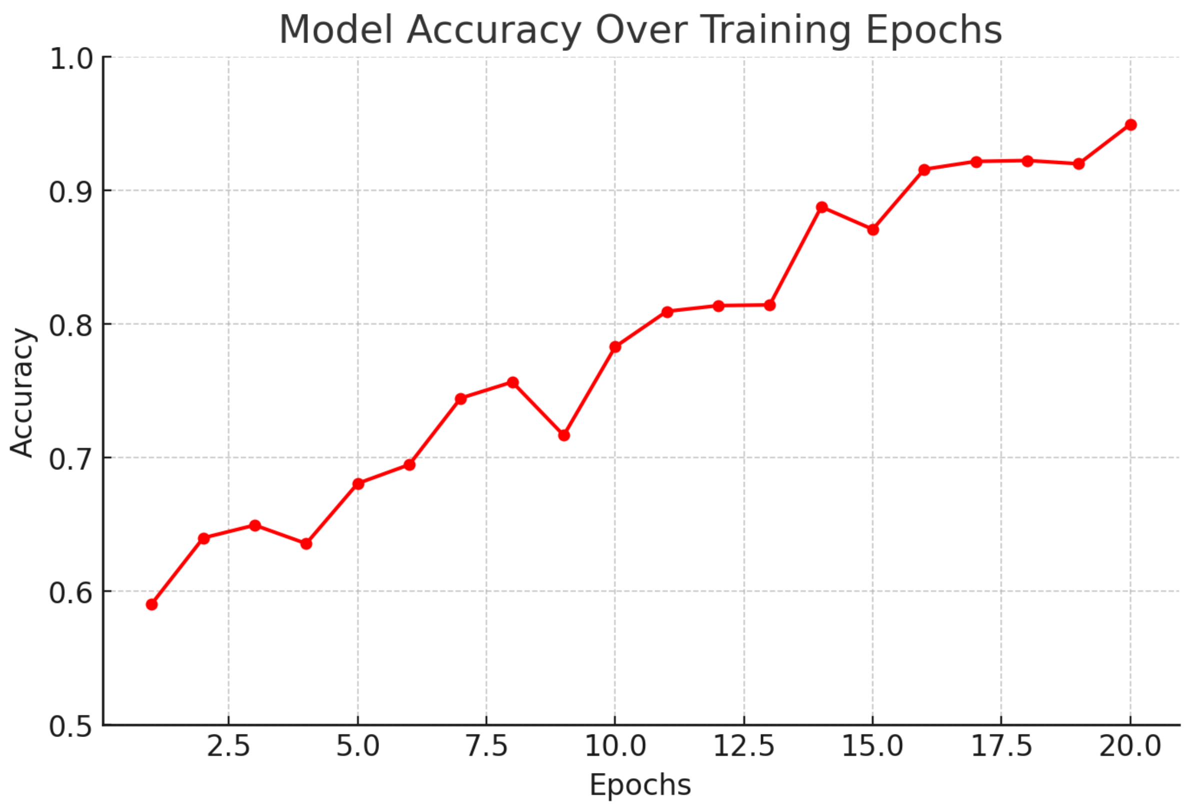

Fig 4 Model Accuracy Over Training Epochs

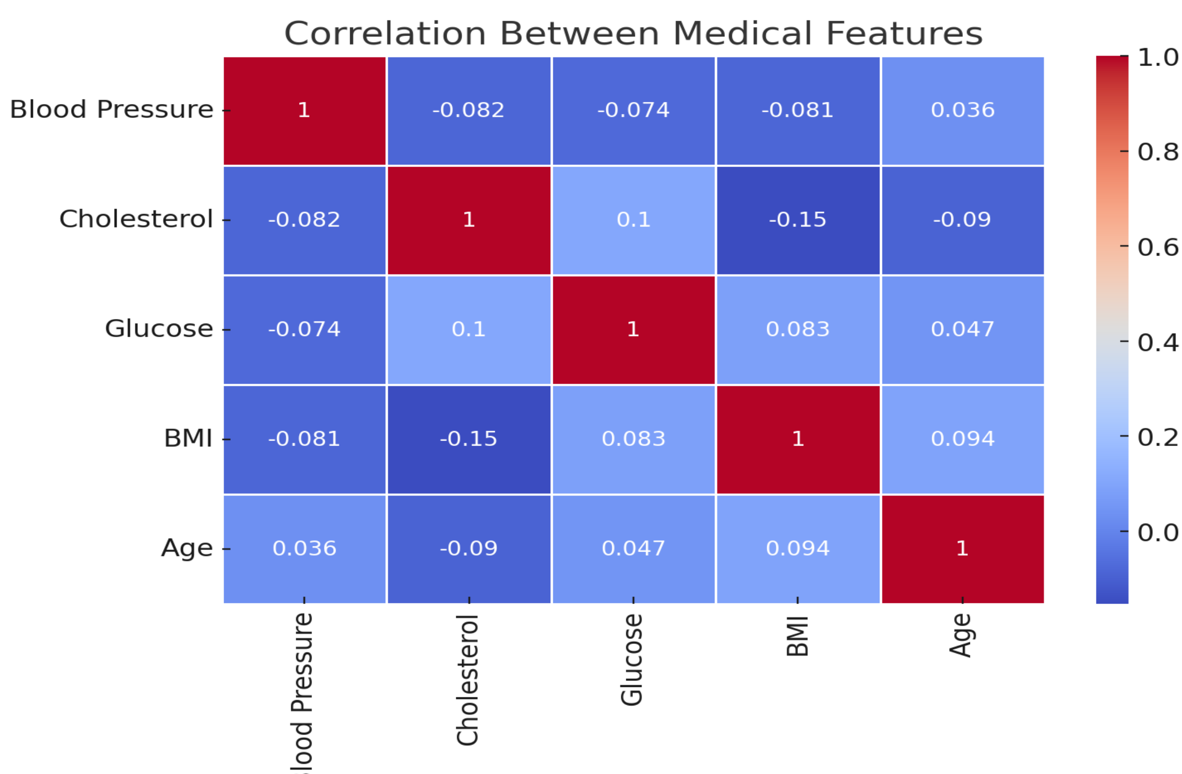

Fig 5 Correlation Between Medical Features

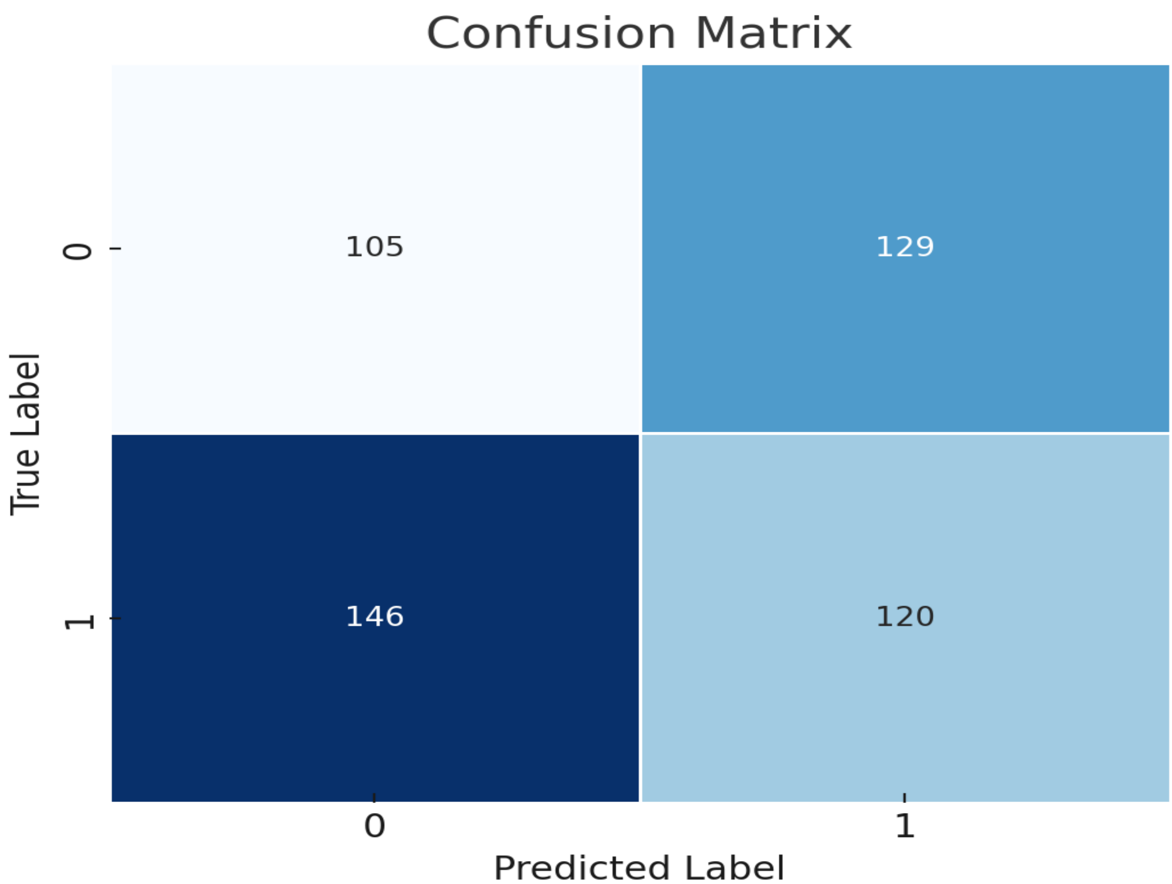

Fig 6 Model Performance Evaluation

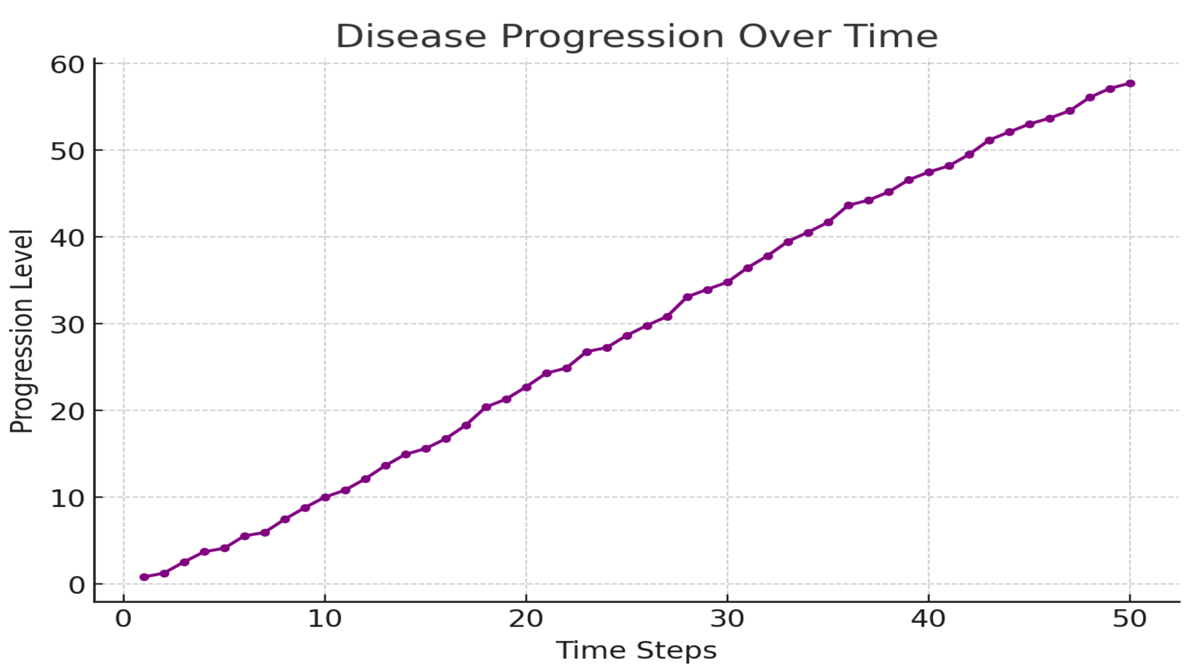

Fig 7 Disease Progression Over Time

CONCLUSION

The application of deep learning methodologies has shown promising capability to use EHR data for spotting chronic diseases during their early stages. The models use effective techniques for dealing with massive and diverse healthcare data to enhance diagnostic precision while speeding up medical responses. The research shows how various network architectures perform in addition to demonstrating why specific feature choice and understandable model design together with secure data management matter. The predictive healthcare systems will gain improvements through both deep learning technique advancement and EHR integration developments despite current difficulties. Research needs to concentrate on enhancing AI model visibility while fixing discrimination issues and protecting patient medical records to ensure the maximum benefits of AI healthcare applications.

REFERENCES

- Li, Y., Mamouei, M., Salimi-Khorshidi, G., Rao, S., Hassaine, A., Canoy, D., Lukasiewicz, T., & Rahimi, K. Hi-BEHRT: Hierarchical Transformer-based model for accurate prediction of clinical events using multimodal longitudinal electronic health records. *arXiv preprint arXiv:2106.11360*, (2021).
- Li, Y., Rao, S., Solares, J.R.A., Hassaine, A., Canoy, D., Zhu, Y., Rahimi, K., & Salimi-Khorshidi, G. BEHRT: Transformer for Electronic Health Records. *arXiv preprint arXiv:1907.09538*, (2019).
- Rasmy, L., Xiang, Y., Xie, Z., Tao, C., & Zhi, D. Med-BERT: pre-trained contextualized embeddings on large-scale structured electronic health records for disease prediction. *arXiv preprint arXiv:2005.12833*, (2020).
- Afrifa-Yamoah, E., Wilson, M.L., & Mensah, G.A. Pathways to chronic disease detection and prediction: Mapping the potential of machine learning to the pathophysiological processes while navigating ethical challenges. *Chronic Diseases and Translational Medicine*, 7(3), 233-243, (2021).
- Zhang, Y., Chen, P.H.C., & Ghassemi, M. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60-88, (2017).
- Choi, E., Bahadori, M.T., Schuetz, A., Stewart, W.F., & Sun, J. Doctor AI: Predicting clinical events via recurrent neural networks. *Proceedings of the 1st Machine Learning for Healthcare Conference*, 301-318, (2016).
- Lipton, Z.C., Kale, D.C., & Wetzel, R. Directly modeling missing data in sequences with RNNs: Improved classification of clinical time series. *Proceedings of the 1st Machine Learning for Healthcare Conference*, 253-270, (2016).
- Pham, T., Tran, T., Phung, D., & Venkatesh, S. DeepCare: A deep dynamic memory model for predictive medicine. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 705-714, (2016).
- Che, C., Kale, D., Li, W., Bahadori, M.T., & Liu, Y. Deep computational phenotyping. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 507-516, (2015).
- Choi, E., Schuetz, A., Stewart, W.F., & Sun, J. Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association*, 24(2), 361-370, (2017).
- Razavian, N., Marcus, J., & Sontag, D. Multi-task prediction of disease onsets from longitudinal laboratory tests. *Proceedings of the 1st Machine Learning for Healthcare Conference*, 73-100, (2016).
- Nguyen, P., Tran, T., Wickramasinghe, N., & Venkatesh, S. Deepr: A convolutional net for medical records. *IEEE Journal of Biomedical and Health Informatics*, 21(1), 22-30, (2017).
- Ma, F., Chitta, R., Zhou, J., You, Q., Sun, T., & Gao, J. Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1903-1911, (2017).
- Choi, E., Bahadori, M.T., Song, L., Stewart, W.F., & Sun, J. GRAM: Graph-based attention model for healthcare representation learning. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 787-795, (2017).
- Suresh, H., Hunt, N., Johnson, A., Celi, L.A., Szolovits, P., & Ghassemi, M. Clinical intervention prediction and understanding using deep networks. *arXiv preprint arXiv:1705.08498*, (2017).
- Gao, J., Xiao, C., Wang, F., & Glass, L.M. Modeling the temporal nature of human behavior for demographics prediction. *Proceedings of the 26th International Conference on World Wide Web*, 1065-1073, (2017).
- Liu, Y., Chen, P.H.C., Krause, J., & Peng, L. How to read articles that use machine learning: Users' guides to the medical literature. *JAMA*, 322(18), 1806-1816, (2019).
- Shickel, B., Tighe, P.J., Bihorac, A., & Rashidi, P. Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE Journal of Biomedical and Health Informatics*, 22(5), 1589-1604, (2018).
- Zhang, Z., Ho, D.E., & Wang, F. Unveiling the predictive power of longitudinal electronic health records for heart failure onset risk: A machine learning study. *JMIR Medical Informatics*, 7(1), e13061, (2019).