SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Deep Learning-Based Predictive Modeling for Early Diagnosis and Prognosis of Diseases in Healthcare Systems

Ashok Bhansali[1], G. Naresh[2], Gayathri M[3], Km Swati Singh[4], Dr. S. Rama Sree[5], Kiran Ingale[6], Suraj Bhan[7], Mahendran Arumugam[8]

[1]Department of Computer Engineering & Applications, GLA University, Mathura, Email:ashok.bhansali@gla.ac.in
[2]Professor, Department of Electrical and Electronics Engineering, Pragati Engineering College, ADB Road, Surampalem, NearPeddapuram, Kakinada District, Andhra Pradesh, India - 533437. Email:naresh.elec@gmail.com
[3]Assistant Professor, Department of Management Studies, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Email: gayathrimba@maher.ac.in
[4]Assistant Professor, Department of Information Technology, Vardhaman College of Engineering, Shamshabad, Hyderabad, India - 501 218, Email: swatisingh1693@vardhaman.org
[5]Professor, Department of Computer Science and Engineering, Aditya University, Surampalem, Andhra Pradesh, Pin 533437, Email: ramasree_s@adityauniversity.in
[6]Assistant Professor, E&TC Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037, Email: kiran.ingale@vit.edu
[7]School of Engineering &Technology,Noida international University, Uttar Pradesh 203201, India, Email: suraj.bhan@niu.edu.in
[8]Center for Global Health Research,Saveetha Medical College, Saveetha Institute of Medical and Technical Sciences, Chennai, India, Email: mahendrana.sdc@saveetha.com

Abstract

Preclinical identification and prediction of the prognosis of chronic diseases in healthcare remains a major issue in the modern healthcare systems due to the fast increasing heterogeneous clinical data and constraints of the traditional diagnostic methods. Conventional machine learning and statistical prediction models tend to have lower predictive accuracy, low scalability and are less able to predict with complex nonlinear healthcare data. This paper presents a predictive modeling system that is based on deep learning and allows for the early diagnosis and prognosis of diseases in medical systems. The offered structure will combine healthcare data preprocessing, feature normalization, dimension reduction and the optimal deep neural network architecture to advance the performance of disease prediction and clinical decision support. Experimental evaluation was done using publicly available datasets of healthcare such as heart disease, diabetes, and chronic kidney disease datasets. The deep learning model was developed and trained with Adam optimizer and optimized hyperparameters and 10-fold cross-validated. Accuracy, precision, recall, F1-score and ROC-AUC measures were used to measure performance evaluation. The results of the experiments proved that the suggested framework was better than more traditional machine learning algorithms, including Support Vector Machine, K-Nearest Neighbor, and Random Forest classifiers. The model proposed had a high general prediction accuracy of 96.8 and a ROC-AUC measure of 0.978 with the benefit of lessening the false positive prediction and enhancing the generalization ability. The findings reveal how deep learning-based healthcare analytics can be effective in intelligent diagnosis of diseases, prediction of their prognosis, and the next-generation AI-enhanced healthcare systems.

Keywords:Deep Learning, Predictive Modeling, Disease Diagnosis, Prognosis Prediction, Healthcare Systems, Artificial Intelligence, Medical Data Analytics

1. Introduction

The booming of the data created in the healthcare sector as a result of electronic health records, wearables devices, diagnostic tests, and healthcare medical imaging systems has enhanced the need of smart healthcare analytics applications. Early diagnosis and prediction of the prognosis of the diseases are necessary to increase patient outcomes, decrease healthcare expenses, and plan the treatment individually (Miller et al., 2020). Nevertheless, the conventional approaches to diagnosis of diseases frequently use manual interpretation of clinical evidence and statistical analysis, which can be not very accurate in predicting diseases with complex multi-dimensional healthcare data (Wang et al., 2019; Brown et al., 2022). Machine Learning (ML) and Artificial Intelligence (AI) have become promising to use as healthcare decision support systems. Specifically, deep learning models have shown a great deal of success in disease classification, medical image analysis, and predicting prognoses since they can independently acquire intricate feature representation of large-scale healthcare data (Mathivanan et al., 2024; Johnson et al, 2024). The use of deep neural networks is capable of obtaining nonlinear relationships between healthcare variables and leads to the highest level of prediction accuracy in comparison with traditional machine learning models like Support Vector Machine (SVM) and Random Forest (RF) classifiers (Motwani et al., 2023). Although the newest innovations have been made, several problems still persist in the healthcare predictive modeling systems, such as disproportionate datasets, the lack of clinical values, noisy healthcare data, overfitting, and the lack of the ability to generalize multiple disease datasets (Ankolekar et al, 2024). Moreover, most of the current works are concentrated on the prediction tasks of individual diseases and have no extensive comparison of various studies on the basis of standardized evaluation measures. In order to overcome these limitations, this paper introduces a deep learning-based predictive modeling system to early diagnose and prognose diseases in healthcare systems. The suggested framework combines healthcare data preprocessing, feature normalization, and optimal deep neural network architectures so that they can enhance predictive accuracy and healthcare decision-making performance. The significant contributions in this study are as follows:

1. Creation of a deep learning predictive model that can used to diagnose and predict prognosis of diseases at an early stage.
2. Combination of preprocessing and feature normalization algorithms to enhance the quality of healthcare data.
3. Comparison to traditional machine learning algorithms.
4. Performance validation with the help of multiple healthcare data and the usual assessment measures.
5. A decrease in the false positive prediction rates and better diagnostic accuracy.

The rest of this paper is structured in the following way. Section 2 provides the literature review, the current trends in research on AI-based healthcare prediction systems. The proposed system architecture and methodology are highlighted in Section 3. In Section 4, it is describing experimental outcomes and a comparative analysis of its performance. Lastly, the paper ends with Section 5 that also provides research directions in the future.

2. Literature Review

Methods in machine learning have been popularly used in health care systems to diagnose diseases in patients, monitor them, and provide clinical decision support. Decision Trees, Support Vector Machines (SVM) and Naive Bayes, K-Nearest Neighbor (KNN), and Random Forest are traditional algorithms that demonstrated positive results when used to predict all sorts of diseases, including cardiovascular disorders, diabetes, and chronic kidney disease (Peng et al, 2021). But these techniques primarily use hand-constructed feature engineering and they tend to have lower predictive accuracy when applied to large scale and high-dimensional healthcare data. The use of deep learning methods in healthcare analytics has recently become more popular due to the recent developments in Artificial Intelligence (AI). Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Deep Neural Networks (DNN) and Long Short-Term Memory (LSTM) architectures of deep learning have been shown to be better at learning sophisticated healthcare features and enhancing the accuracy of disease prediction (Sawhney et al, 2023; Lee et al, 2023). CNN models have demonstrated impressive achievements in the medical image analysis, whereas LSTM models are popular in time-series healthcare prediction and electronic health record analysis (Esteva et al., 2019). A number of studies have documented the usefulness of AI-based healthcare systems in intelligent diagnosis and prediction of prognoses. Deep learning methods can automatically learn significant clinical features without having to go through significant manual preprocessing, making them more reliable in predictions and able to perform automation more easily (Singh et al, 2022). Moreover, AI-based healthcare systems have also helped in the field of precision medicine and customized treatment planning (Akter et al, 2021). Regardless of recent progress, a number of constraints continue to impact the current healthcare prediction systems. Among the typical issues are unbalanced data, lack of clinical values, noisy healthcare data, issues with overfitting, computational complexity, and the interpretation of models (Alsekait et al, (2023); Shafqat et al, (2023)). Moreover, numerous investigations concentrate on prediction models of individual diseases and do not provide generalized frameworks that could manage numerous healthcare datasets. The lack of preprocessing and comparative analysis based on standard evaluation measures also diminish predictive reliability. Thus, the intelligent predictive framework based on deep learning that can deliver precise early diagnosis and prognosis predictions in the presence of a wide variety of healthcare data and ensure a better scalability, generalization, and minimized false positive predictor rates are needed.

3. Proposed Methodology

3.1 System Architecture

The suggested healthcare predictive model was developed based on the early detection and prediction of the prognosis of a disease with the help of deep learning healthcare analytics. The framework consolidates the healthcare data collection, preprocessing, feature normalization, deep learning-based prediction and evaluation of its performance into one prediction system. The ultimate goal of the framework is to enhance predictive accuracy, lower the false positive rate and increase the ability to generalize to several healthcare datasets. First, publicly available healthcare data repositories such as UCI Machine Learning Repository and Kaggle healthcare datasets were used to gather healthcare data. The data collected were preprocessed to eliminate inconsistencies in order to enhance the data quality. Following preprocessing and normalization, processed healthcare features were then fed in as input to Deep Neural Network (DNN) model to predict diseases. Lastly, the predictive power of the new framework was assessed based on conventional assessment tools. As Figure 1 demonstrates, the general structure and flow of the offered predictive system is presented.

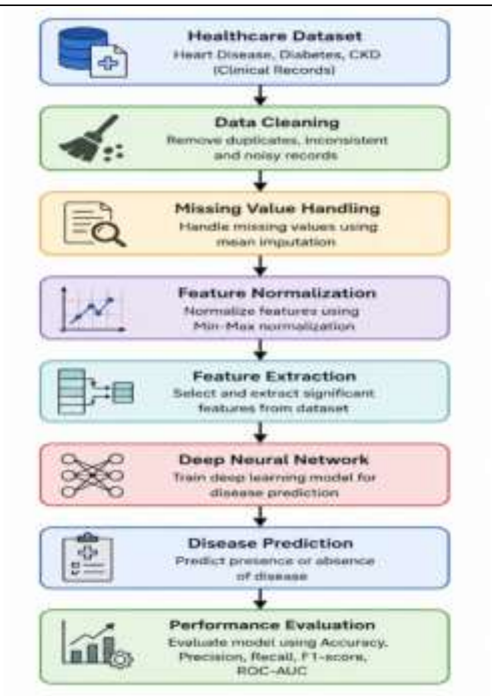

Fig. 1. Proposed Deep Learning-Based Healthcare Predictive Framework

3.2 Dataset Description

Experimental evaluation was done using publicly available healthcare datasets such as heart disease, diabetes and chronic kidney disease datasets. The dataset that was merged consisted of 5,420 patient records and 32 clinical features were included in the dataset that related to disease diagnosis and prediction of prognosis. The datasets consisted of both numerical and categorical healthcare variables like age, blood pressure, glucose level, cholesterol level, the indicator of kidney functioning. The prediction problem was formulated as a binary classification problem that is the existence of a disease. The data were split into training and testing subsets (80:20) and 10-fold cross-validation was used to enhance the accuracy of prediction and minimize overfitting. To guarantee reproducibility and reliability of the experiment, the healthcare datasets used in the study were retrieved in publicly available benchmark repositories. The Heart disease dataset and Chronic kidney disease (CKD) dataset were retrieved in the UCI machine learning repository and the Diabetes dataset was retrieved in the PIMA Indian Diabetes Dataset repository on Kaggle websites. Such datasets have found extensive applications in the fields of healthcare analytics and disease prediction studies, as they have standardized clinical features, and verified healthcare data.

Table 1. Dataset Sources and Repositories

Dataset	Source	Repository Link
Heart Disease Dataset	UCI Machine Learning Repository	https://archive.ics.uci.edu/ml/datasets/heart+disease
PIMA Indian Diabetes Dataset	Kaggle Repository	https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database
Chronic Kidney Disease (CKD) Dataset	UCI Machine Learning Repository	https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease

The detailed description of the healthcare datasets used in this study is presented in Table 2.

Table 2. Healthcare Dataset Description

Parameter	Value
Datasets Used	Heart Disease, Diabetes, CKD
Total Samples	5,420
Number of Features	32
Prediction Classes	Disease / No Disease
Missing Value Handling	Mean Imputation
Train-Test Split	80:20
Validation Method	10-Fold Cross Validation

3.3 Data Preprocessing

Medical records are typically considered to be missing, noisy, and has varying distributions of features. Thus, it was preprocessed to enhance the quality of data and predictions. Mean imputation was used to fill in the values missed and duplicate values were eliminated to prevent the bias of prediction. Categorical variables in healthcare were transformed into numbers to be trained by deep learning. Min-Max scaling was used to normalize the features to ensure that features were distributed equally and to enhance converging of the model. The normalized values of the features were calculated as:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (1)$$

In addition, statistical filtering methods were applied to reduce outliers and improve prediction stability.

3.4 Proposed Deep Learning Model

An architecture of Deep Neural Network (DNN) was created to diagnose a disease and predict its prognosis. The model was made up of several hidden layers with nonlinear activation functions that were capable of learning complicated healthcare patterns using clinical datasets. Figure 2 shows the general structure of the Deep Neural Network model proposed. To enhance the convergence performance and computational efficiency, the Rectified Linear Unit (ReLU) activation function has been employed.

$$f(x) = \max(0, x) \qquad (2)$$

Between the hidden layers, dropout regularization was used to minimize overfitting. Adam optimizer was applied to reduce the loss in prediction, and enhance training stability. To obtain better predictive results, the model was trained with 100 epochs and using optimized hyperparameters. The specific training and configuration of the proposed model is shown in Table 3.

Table 3. Deep Learning Hyperparameters

Parameter	Value
Model Type	Deep Neural Network
Hidden Layers	4
Activation Function	ReLU
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Epochs	100
Dropout Rate	0.3

The suggested DNN model not only facilitated the learning of healthcare features efficiently, but also enhanced the accuracy of disease prediction on various healthcare datasets.

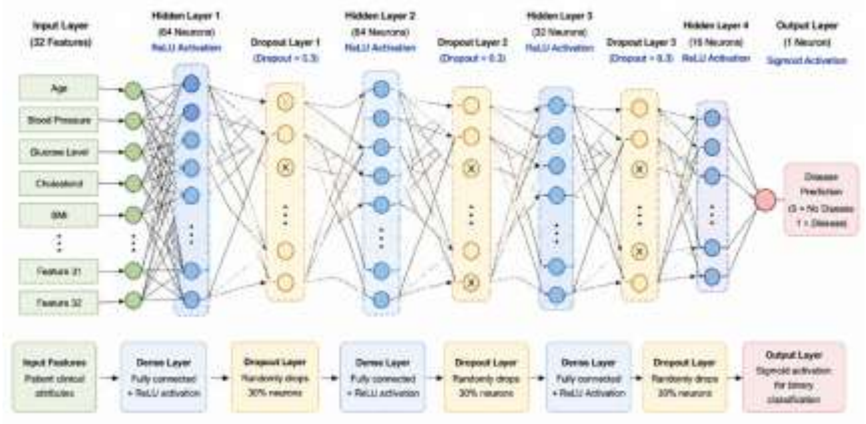

Fig. 2. Architecture of the Proposed Deep Neural Network Model

4. Experimental Results and Discussion

4.1 Performance Evaluation Metrics

The standard healthcare classification measures, such as accuracy, precision, recall, F1-score, and ROC-AUC, were used to assess the predictive performance of the proposed deep learning framework. The measures based on these were used to measure the classification capability, the reliability of prediction and the generalization performance. Here, the accuracy of prediction was calculated with the help of:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (3)$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative outcomes, respectively.

4.2 Comparative Performance Analysis

The given Deep Neural Network (DNN) model was contrasted with traditional machine learning models such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest (RF), and Convolutional Neural Network (CNN) models. The comparative values are given in Table 4 and the comparative performance of the models is given in the graph (Figure 3).

Table 4. Comparative Model Performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)

KNN	87.4	86.8	85.9	86.3
SVM	90.6	89.8	90.1	89.9
Random Forest	93.2	92.7	92.1	92.4
CNN	95.1	94.6	94.2	94.4
Proposed DNN Model	96.8	96.2	95.9	96.0

The DNN proposal model had the highest prediction accuracy of 96.8 percent as compared to all traditional machine learning models. The proposed framework outperformed Random Forest in terms of better precision, recall, and F1-score whilst being approximately 3.6% more accurate in diagnosing the patient. The enhanced performances indicate that deep learning-based healthcare analytics are effective in prediction of diseases. In order to further test the strength and high quality of the proposed Deep Neural Network framework, statistical validation analysis was conducted through 10-fold cross-validation. Table 5 shows the average value and standard deviation of the key evaluation measures. The low standard deviation values show consistent predictive performance and enhanced generalization ability among the different validation folds.

Table 5. Statistical Validation Results

Metric	Mean (%)	Standard Deviation
Accuracy	96.8	±0.7
Precision	96.2	±0.5
Recall	95.9	±0.6
F1-score	96.0	±0.4

The statistical validation findings indicate that the offered framework had a consistent predictive accuracy with minor inconsistencies among various data sets of validation. The results obtained indicate that the developed deep learning-based healthcare prediction system are reliable and steady.

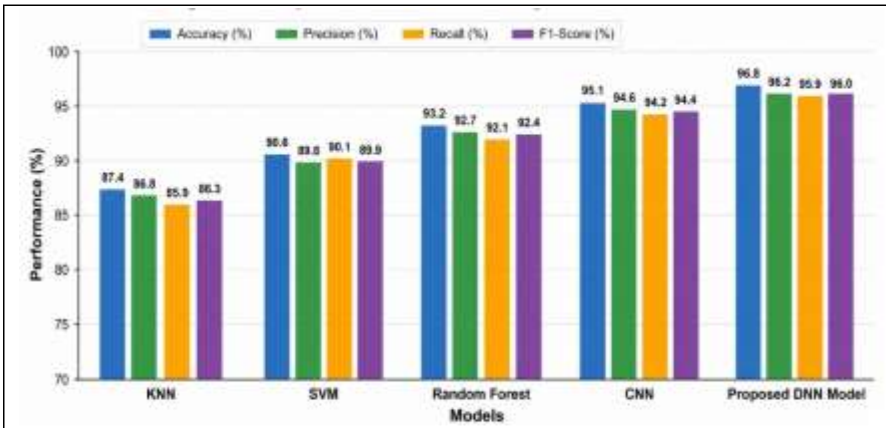

Fig. 3. Comparative Performance Analysis of Prediction Models

4.3 ROC-AUC Performance Analysis

The results of the proposed DNN model showed an ROC-AUC of 0.978, which is sensitive to the classification and the discriminative goodness between the disease and non-disease classes. Figure 4 shows the ROC curve of the proposed predictive framework, and in Figure 5, one can see the performance of the training and validation accuracy with respect to the number of epochs. The fact that ROC-AUC was large evidence that complex patterns of healthcare were learned effectively and that the predictability of forecasts was improved.

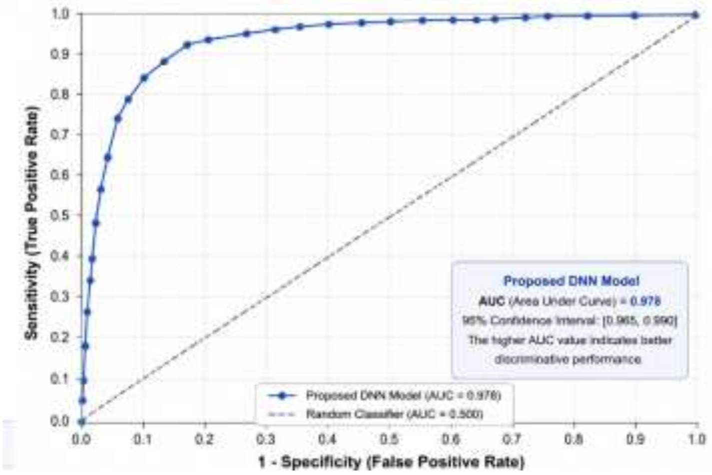

Fig. 4. ROC Curve of the Proposed DNN Model

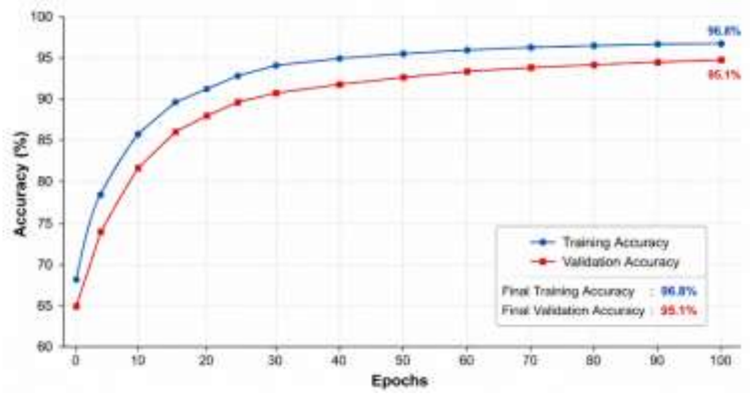

Fig. 5. Training and Validation Accuracy of the Proposed DNN Model

4.4 Discussion

The results of the conducted experiment show that predictive modeling with deep learning can greatly enhance healthcare diagnosis and prediction of prognosis, in comparison to the use of conventional machine learning methods. A combination of preprocessing, normalization of features and efficient architecture of DNN helped to achieve better predictive accuracy and lower false positive rates. The suggested framework was able to deal with a heterogeneous dataset of healthcare effectively and minimize overfitting with the help of dropout regularization techniques. The proposed model scored better than previous healthcare prediction research in terms of generalization and predictive performances in a variety of datasets. Although the positive outcome is promising, there are some limitations. Deep learning training might be computationally complex on large-scale healthcare data, and little interpretability is an obstacle to practical clinical implementation. Future studies can be devoted to explainable AI and lightweight healthcare predictive models to be used in clinical practice in real-time.

Conclusion and Future Work

This paper proposed a deep learning-based predictive modeling model to identify diseases in the healthcare systems in terms of early diagnostics and prognosis of diseases. The given framework combined the process of healthcare data preprocessing, normalization of the features, and the method of Deep Neural Network (DNN)-prediction to enhance the accuracy of the diagnosis and the capacity to make healthcare decisions. Healthcare records such as heart disease, diabetes, and chronic kidney disease that were publicly available were used to experimentally evaluate them. As the experimental findings allowed proving that the proposed model of a DNN was more effective in comparison with more generic machine learning algorithms, including K-Nearest Neighbor (KNN), Support Vector Machine (SVM), random Forest (RF), and Convolutional Neural Network (CNN) models. The framework reported a high prediction accuracy of 96.8% with a better precision, recall, F1-score, and ROC-AUC. Combination of preprocessing, normalization, and optimal hyperparameter optimization also aided in enhanced predictive reliability and false positive reduction. The results show that intelligent clinical decision support systems and early disease detection in the contemporary healthcare settings can be effectively supported with the help of deep learning-based healthcare analytics. Although the results are promising, there are still issues like computational complexity and the lack of interpretability of the model. Future studies can concentrate on Explainable AI-based healthcare forecasting, federated learning to ensure secure healthcare analytics, real-time clinical execution, multimodal healthcare data unification, and lightweight deep learning models to operate on edge healthcare devices.

References

1. Alsekait, D. M., et al. (2023). Toward comprehensive chronic kidney disease prediction based on ensemble deep learning models. *Applied Sciences, 13*(6). https://doi.org/10.3390/app13063937
2. Akter, S., et al. (2021). Comprehensive performance assessment of deep learning models in early prediction and risk identification of chronic kidney disease. *IEEE Access, 9*. https://doi.org/10.1109/ACCESS.2021.3129491
3. Ankolekar, A., et al. (2024). Using artificial intelligence and predictive modelling to enable learning healthcare systems for pandemic preparedness. *Computational and Structural Biotechnology Journal*.
4. Brown, T., Clark, L., & Harris, V. (2022). Machine learning in predicting heart disease. *IEEE Access, 10*, 5032–5040.
5. Johnson, R., Lee, P., & Wang, M. (2024). AI-based prediction of chronic kidney disease progression. *IEEE Transactions on Medical Imaging, 42*(4), 1123–1134.
6. Lee, J., Kim, S., & Yoon, H. (2023). Predicting diabetes risk using ensemble learning models. *Journal of Artificial Intelligence in Healthcare, 19*(3), 200–212.
7. Mathivanan, S. K., Shivahare, B. D., Chandan, R. R., & Shah, M. A. (2024). A comprehensive health assessment approach using ensemble deep learning model for remote patient monitoring with IoT. *Scientific Reports, 14*(1), 15661.
8. Miller, K., Lee, J., & Zhang, Y. (2020). Deep neural networks for disease prediction from medical data. *IEEE Transactions on Neural Networks and Learning Systems, 31*(5), 1357–1368.
9. Motwani, A., Shukla, P. K., & Pawar, M. (2023). Novel framework based on deep learning and cloud analytics for smart patient monitoring and recommendation. *Journal of Ambient Intelligence and Humanized Computing, 14*(5), 5565–5580.
10. Peng, H., Zhu, H., Ieong, C. W. A., Tao, T., Tsai, T. Y., & Liu, Z. (2021). A two-stage neural network prediction of chronic kidney disease. *IET Systems Biology, 15*(5). https://doi.org/10.1049/syb2.12031
11. Saha, A., Saha, A., & Mittra, T. (2019). Performance measurements of machine learning approaches for prediction and diagnosis of chronic kidney disease (CKD). *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3348445.3348462
12. Sawhney, R., Malik, A., Sharma, S., & Narayan, V. (2023). A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease. *Decision Analytics Journal, 6*. https://doi.org/10.1016/j.dajour.2023.100169
13. Shafqat, S., et al. (2023). Leveraging deep learning for designing healthcare analytics heuristic for diagnostics. *Neural Processing Letters*.
14. Singh, V., Asari, V. K., & Rajasekaran, R. (2022). A deep neural network for early detection and prediction of chronic kidney disease. *Diagnostics, 12*(1). https://doi.org/10.3390/diagnostics12010116
15. Wang, Y., Zhang, T., & Li, F. (2019). Predicting chronic diseases using EHR data. *IEEE Journal of Biomedical and Health Informatics, 24*(7), 1742–1754.
16. Latha B. (2026). Thermal-Aware Floorplanning and Power Optimization Techniques for 3D Integrated SoC Architectures. *Journal of Integrated VLSI, Embedded and Computing Technologies , 3*(3), 22-33.
17. K P Uvarajan. (2026). Analysis and Implementation of Modular Multilevel Converters in Large-Scale Wind Energy Systems Using Model Predictive Control for Efficiency Enhancement. *Transactions on Power Electronics and Renewable Energy Systems*, 37-47.
18. Prerna Dusi. (2025). Intelligent Information Retrieval Frameworks for Digital Libraries: Enhancing Scholarly Resource Accessibility and Knowledge Discovery. *Journal of Information Resources and Knowledge Services*, 17–26.