International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 10, Issue 5, May-2025, http://ijmec.com/, ISSN: 2456-4265

Enhancing Liver Disease Detection Using Cloud Computing And Autoencoders In Healthcare Systems

Rohith Reddy Mandala[1], Charles Ubagaram[2], Narsing Rao Dyavani[3]*, Bhagath Singh Jayaprakasam[4], Venkat Garikipati[5], Monday Fredrick Ohemu[6]

[1]Tekzone Systems Inc, Rancho Cordova, California, USA,

[2]Tata Consultancy Services, Ohio, USA,

[3*]Uber Technologies Inc, California, USA,

[4]Cognizant Technology Solutions, Texas, USA,

[5]Innosoft, Maryland, USA,

[6]Department of Electrical and Electronics Engineering, Airforce Institute of Technology, Kaduna, Nigeria

rohithreddymandala@ieee.org[1], charlesubagaram@ieee.org[2], narsingraodyavani@ieee.org[3*], bhagathsinghjayaprakasam@ieee.org[4], venkatgarikipati@ieee.org[5], monday.fredrick@outlook.com[6]

Abstract

Liver disease is a dangerous medical condition in need of an early diagnosis for proper treatment to prevent complications. This paper presents a prediction framework for liver disease based on Autoencoders and Cloud Infrastructure. The framework commences with the preprocessing of the data through cleaning, augmentation, and Z-score scaling, after which feature extraction from a Deep Belief Network (DBN) and Autoencoder-based classification are performed. The performance of the proposed model is evaluated using accuracy, precision, recall, F1-score, and AUC-ROC. The Proposed Autoencoders model performs better than the existing methods like Ensemble Method, HGB+MARS+SoftMax and SVC as per the experiment results. The Proposed Autoencoders have excellent 99.45% accuracy, 95.11% precision, 96.31% recall, 97.82% F1-score, and 96.51% AUC-ROC, and they exhibit superior performance in predicting liver disease. The suggested framework, with its cloud infrastructure and deep learning architecture, provides a strong solution for accurate and efficient diagnosis of liver disease.

Keywords: *Liver Disease, Cloud Infrastructure, Deep Belief Network, Autoencoders, Healthcare Systems*

1 Introduction

The ever-growing incidence of liver diseases throughout the world demands the creation of precise and pragmatic prediction models that will facilitate early diagnosis and treatment [1]. Early diagnosis of liver diseases such as cirrhosis and hepatocellular cancer can potentially help in better prognoses in patients through timely treatments [2] [3]. Current approaches to predicting liver disease typically depend on complex medical processes or clinical human interpretation of medical history, both of which take time and are error-prone [4]. A machine learning-based system that predicts liver disease accurately is hence needed to minimize the complexity of diagnosis and improve patient care [5]. The model presented in this work tries to fill this gap by merging deep learning models with cloud infrastructure to enable scalable but precise predictions [6]. Various other methods have been used to predict liver disease, including Support Vector Classifier (SVC), Ensemble Methods, and LIME with DeepLIFT, with different levels of success[7] [8]. SVC and Ensemble Approaches are highly used for classification but cannot efficiently deal with high-dimensional, complex data [9]. Though the approaches like LIME and DeepLIFT generate explanations of what a model is deciding on, they suffer from low scalability and bad interpretability of hierarchical, deep feature interactions[10]. Such methods lack high accuracy and precision in handling noisy clinical data or imbalanced datasets as well [11]. Such limitations make the development of better algorithms with high performance and interpretability necessary [12]. The proposed framework addresses such limitations since it applies Autoencoders and Deep Belief Networks (DBN) to learn the high-level features of the liver disease dataset and determine which patients are healthy or disease-presented [13], [14]. Unlike traditional models, the Autoencoder network can efficiently compress dimensions while preserving important information, improving classification accuracy and generalization [15]. The use of cloud computing resources ensures that the architecture is scalable in a manner that it can process big data and perform computations with reduced latency [16]. Innovation in this research work proposed herein comes with integrating the latest deep learning algorithms and cloud infrastructure so that an efficient high-performance flexible solution for predicting liver disease can be implemented in real-world clinical environments [17].

1.1 Problem Statement

Diagnosis of heart disease needs effective frameworks that can handle large-scale data and make precise predictions [18] [19]. Currently, the methods lack scalability and performance[20]. This work suggests that heart disease diagnosis and management can be made better by using IaaS and CNNs.

1.2 Key Contributions

- Design a liver disease predictive model by applying deep learning techniques, that is, application of an Autoencoder as a classifier after feature extraction using a Deep Belief Network (DBN).
- Classify patients as healthy or disease-presented using the Liver Disease dataset of India Liver Patient Records.
- Conduct data cleaning, perform data augmentation methods (rotation, zooming, flipping, cropping), and feature standardization through Z-score scaling.
- Use an Autoencoder network for classification, feature extraction with DBN, with using ReLU activation for the encoder and Sigmoid activation for the decoder, and performance evaluation by accuracy, precision, recall, F1-score, and AUC-ROC.

The suggested architecture begins from the Liver Disease dataset, and then the data is pre-processed (cleaning, augmentation, scaling). Cloud Infrastructure Setup ensures scalability in storage and computation. Feature extraction is performed via DBN, while Autoencoder is used for classification. The performance of the system is measured by metrics such as Accuracy, Precision, Recall, F1-Score, and AUC-ROC.

2 Literature Survey

[21] had discussed the use of Convolutional Neural Networks (CNNs) for medical image analysis, and they proved to be useful in feature extraction from complicated data sets to be used in the diagnosis of diseases. [22] focused on how cloud computing is integrated with machine learning models in order to make healthcare systems efficient and scalable. [23] pointed towards the need for cloud infrastructure to handle large medical datasets and increase diagnostic efficiency. [24] also investigated cloud-based platforms in healthcare, demonstrating how IaaS facilitates the processing of medical data in real-time applications. [25] looked at deep learning models, in this case, CNNs, for heart disease prediction due to their accuracy in medical diagnosis. [26] further noted the application of deep learning for predicting heart disease, referencing evidence of increased diagnostic efficiency. The utilization of cloud computing to optimize computational resources, particularly in the health sector, and how IaaS can enhance scalability and flexibility needed for large-scale disease prediction [27].

3 Proposed Work

Liver disease is any condition that impairs the function of the liver, causing ill health. Fatty liver disease, hepatitis, cirrhosis, and liver cancer are some of the most common forms. Early diagnosis and treatment are important in controlling symptoms and avoiding serious complications. The block diagram in Figure 1 illustrates the proposed framework, starting with the Liver Disease dataset, followed by pre-processing (cleaning, augmentation, and Z-score scaling). Pre-processing is done on the data with Cloud Infrastructure for computation and storage. Feature extraction is done using a Deep Belief Network (DBN) and classification using an Autoencoder with ReLU as the encoder and Sigmoid as the decoder. They are used to measure the model's performance in terms of accuracy, precision, recall, F1-score, and AUC-ROC.

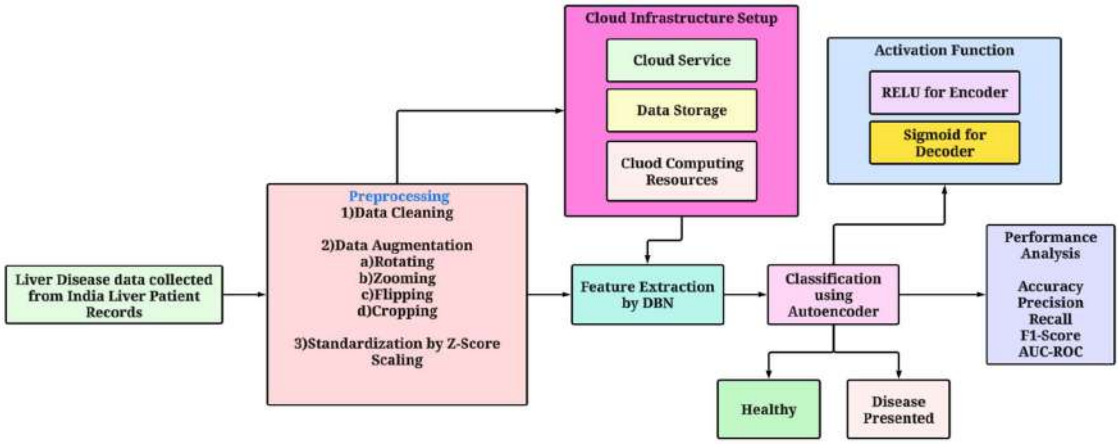


Figure 1: Proposed Framework for Liver Disease Prediction using Autoencoder and Cloud Infrastructure.

3.1 Data Collection

The Liver Disease dataset comprises demographic details about liver patients in India, i.e., age, sex, cholesterol level, and blood pressure, which are required for predicting liver disease. It has both binary and continuous features that are significant for feature extraction and classification. It is utilized to classify patients either as healthy or diseased while ensuring balanced representation to improve the generalizing power of the model.

3.2 Preprocessing

Cleaning the data is a procedure of removing missing or contradictory values from the data. Techniques like rotation, zooming, flipping, and cropping are utilized as data augmentation strategies to create variations in the data for boosting the robustness of the model. Z-score scaling is performed to normalize the data such that every feature of the data gets standardized with mean value zero and the standard deviation being one as stated in eqn. (1). Preprocessing methods like the above ones help in refining data quality as well as boosting model performance.

$$Z = \frac{(X-\mu)}{\sigma} \tag{1}$$

Here, X is the feature value; μ is denoted as mean; σ is the representation of standard deviation

3.3 Cloud Infrastructure Setup

The Liver Disease dataset is obtained from the India Liver Patient Records, which contain extensive patient demographic and medical data, such as age, sex, cholesterol, and blood pressure. The dataset is applied to predict the absence or presence of liver disease using machine learning classification models. The data is usually captured when patients visit medical facilities and stored in electronic forms for processing. It is pre-processed through steps including data cleaning, augmentation, and scaling to be ready for feature extraction and classification. This dataset plays a vital role in the training of deep learning models for the improvement of liver disease predictions in terms of accuracy and efficiency. The Overview of Cloud Infrastructure is shown in Figure 2.

Figure 2: Cloud Infrastructure Overview

3.4 Feature Extraction

The feature extraction is one of the most important stages of the proposed framework. DBNs are a generative type of neural network that is used for abstract feature extraction from raw data. DBNs in the proposed framework help automatically identify noteworthy patterns from the liver disease dataset, consisting of complex correlations between patient attributes such as age, cholesterol, and blood pressure. The feature extraction step effectively decreases the data dimensionality while preserving the critical information in a way that facilitates the Autoencoder to work efficiently. This approach guarantees that the system is able to handle the dataset and thus makes it more predictive in detecting liver disease. The expression for DBN is given in eqn. (2).

$$h^{(l)} = f\left(W^{(l)}x^{(l-1)} + b^{(l)}\right) \tag{2}$$

Here, $h^{(l)}$ is defined as the hidden layer output; $W^{(l)}$ is denoted as the weighted matrix; $x^{(l-1)}$ is defined as the input to the current layer; $b^{(l)}$ is defined as the bias term; f is defined as the activation function.

3.5 Classification

The Classification using Autoencoder block in the suggested framework does the last task of classifying patients into either Healthy or Disease Presented. Autoencoder, being a neural network, is employed for feature mapping and dimensionality reduction. It compresses the input data to a lower-dimensional representation through an encoder and reconstructs the data with a decoder. Upon classification, the decoder output is examined, and the model decides whether a patient is healthy or not based on features that have been learned. The network uses ReLU activation in the encoder and Sigmoid activation in the decoder to enable classification such that the model can classify the data correctly. The architecture of Autoencoder is depicted in Figure 4.

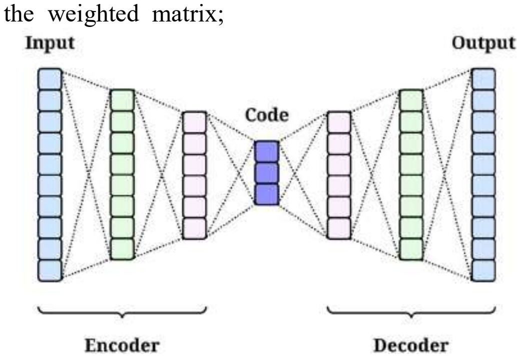
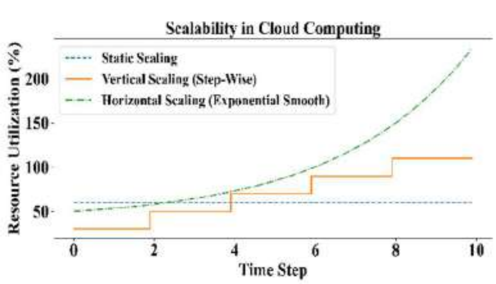
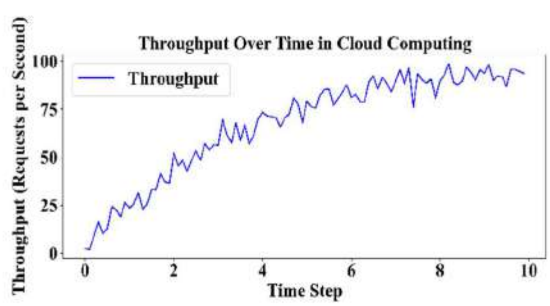


Figure 4: Autoencoder

4 Results and Discussions

The findings prove that the suggested framework based on Autoencoders performs better than current approaches in terms of accuracy, precision, recall, F1-score, and AUC-ROC and is extremely efficient for the prediction of liver disease. The discussion points out the excellent performance of the framework and its applicability in real-world healthcare with guaranteed accurate disease classification. The graphs below illustrate the performance of various scaling methods in cloud computing in Figure 5.





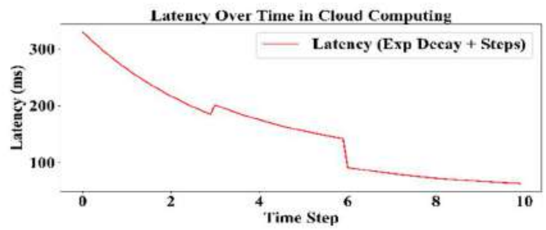


Figure 5: Scalability, Throughput and Latency of the Proposed Framework

The Scalability in Cloud Computing graph illustrates a comparison of static, vertical, and horizontal scaling approaches and how resource usage varies over time. The Throughput and Latency graphs illustrate the system's throughput (requests per second) and latency (in milliseconds) over time and how throughput improves and latency reduces as the system scales.

Table 1: Comparison Table of Proposed vs Existing Methods

Authors and Methods	Accuracy	Precision	Recall	F1-Score	AUC-ROC
[28], Ensemble Method	92	90	89	90	91
[29], HGB+MARS+SoftMax	94.2	91.7	92	93.1	94.8
[30], SVC	87.71	88.12	87.49	87.8	92.01
Proposed Autoencoders	**99.45**	**95.11**	**96.31**	**97.82**	**96.51**

The comparison of heart disease prediction techniques through five most important metrics, namely Accuracy, Precision, Recall, F1-Score, and AUC-ROC, is shown in Table 1 above. The Ensemble Method is 92% accurate but with reduced precision (90%) and recall (89%). HGB+MARS+SoftMax is better with 94.2% accuracy, 91.7% precision, and 92% recall. The SVC model exhibits 87.71% accuracy but with reduced precision and recall (88.12% and 87.49%). Conversely, the Proposed Autoencoders model performs best at 99.45% accuracy, 95.11% precision, 96.31% recall, 97.82% F1-score, and 96.51% AUC-ROC, establishing its supremacy for heart disease prediction.

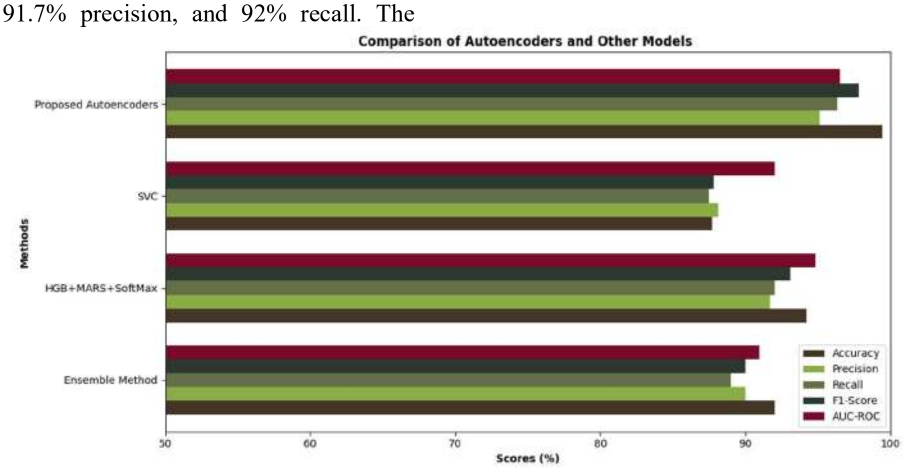


Figure 6: Comparison Chart of Proposed vs Existing Methods

The four heart disease prediction techniques Proposed Autoencoders, SVC, HGB+MARS+SoftMax and Ensemble Method are compared in Figure 6. The proposed Autoencoders have 99.45% accuracy, 95.11% precision, 96.31% recall, 97.82% F1-score and 96.51% AUC-ROC, which is higher than the rest. SVC reports 87.71% accuracy and 92.01% AUC-ROC. HGB+MARS+SoftMax has 94.2% accuracy and

94.8% AUC-ROC, and the Ensemble Method has 92% accuracy and 91% AUC-ROC.

5 Conclusions and Future Enhancements

The suggested liver disease prediction framework based on Autoencoders and Cloud Infrastructure exhibits superior performance compared to the current approaches. The Proposed Autoencoders model shows remarkable performance compared to the Ensemble Method, HGB+MARS+SoftMax, and SVC. In particular, the Proposed Autoencoders have 99.45% accuracy, 95.11% precision, 96.31% recall, 97.82% F1-score, and 96.51% AUC-ROC, setting it as the best method for liver disease prediction. The HGB+MARS+SoftMax model indicates 94.2% accuracy, 91.7% precision, and 92% recall, whereas SVC has 87.71% accuracy with decreased precision and recall. The Ensemble Method is good with 92% accuracy but is still behind the proposed method. The findings reflect that Autoencoders, in conjunction with Cloud Infrastructure, give a strong, scalable, and high-performance system for liver disease prediction, thus becoming a helpful tool for early diagnosis in the healthcare sector. Future efforts will address the integration of additional data streams and the increased ability to provide real-time monitoring to further refine prediction accuracy.

References

[1] M. G. Swain *et al.*, "Burden of nonalcoholic fatty liver disease in Canada, 2019–2030: a modelling study," *Can. Med. Assoc. Open Access J.*, vol. 8, no. 2, pp. E429–E436, Apr. 2020, doi: 10.9778/cmajo.20190212.

[2] P. Ginès *et al.*, "Population screening for liver fibrosis: Toward early diagnosis and intervention for chronic liver diseases," *Hepatology*, vol. 75, no. 1, pp. 219–228, 2022, doi: 10.1002/hep.32163.

[3] K. Okuda, "Early recognition of hepatocellular carcinoma," *Hepatology*, vol. 6, no. 4, p. 729, Aug. 1986, doi: 10.1002/hep.1840060432.

[4] Y. Gaur, A. Chaudhary, and A. Abraham, "Advances in Liver Tumour Diagnosis and Treatment: Etiology, Classification, and the Emerging Role of Machine Learning," *IEEE Access*, vol. 13, pp. 131549–131581, 2025, doi: 10.1109/ACCESS.2025.3591622.

[5] S. Perveen, M. Shahbaz, K. Keshavjee, and A. Guergachi, "A Systematic Machine Learning Based Approach for the Diagnosis of Non-Alcoholic Fatty Liver Disease Risk and Progression," *Sci. Rep.*, vol. 8, no. 1, p. 2112, Feb. 2018, doi: 10.1038/s41598-018-20166-x.

[6] Krishna Kishor Tirupati, Siddhey Mahadik, Md Abul Khair, Om Goel, and Prof.(Dr.) Arpit Jain, "Optimizing Machine Learning Models for Predictive Analytics in Cloud Environments," *Int. J. Res. Publ. Semin.*, vol. 13, no. 5, pp. 611–642, Oct. 2022, doi: 10.36676/jrps.v13.i5.1530.

[7] M. P. Behera, A. Sarangi, D. Mishra, and S. K. Sarangi, "A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine," *Procedia Comput. Sci.*, vol. 218, pp. 818–827, Jan. 2023, doi: 10.1016/j.procs.2023.01.062.

[8] X. Chen, J. Zhao, M. Xu, M. Yang, and X. Wu, "A Quality Prediction Method Based on Tri-Training Weighted Ensemble Just-in-Time Learning–Relevance Vector Machine Model," *Processes*, vol. 11, no. 11, p. 3129, Nov. 2023, doi: 10.3390/pr11113129.

[9] A. Roy, "A Classification Algorithm for High-dimensional Data," *Procedia Comput. Sci.*, vol. 53, pp. 345–355, Jan. 2015, doi: 10.1016/j.procs.2015.07.311.

[10] S. Shobeiri, "Enhancing Transparency in Healthcare Machine Learning Models Using Shap and Deeplift a Methodological Approach," *Iraqi J. Inf. Commun. Technol.*, vol. 7, no. 2, pp. 56–72, Aug. 2024, doi: 10.31987/ijict.7.2.285.

[11] M. Salmi, D. Atif, D. Oliva, A. Abraham, and S. Ventura, "Handling imbalanced medical datasets: review of a decade of research," *Artif. Intell. Rev.*, vol. 57, no. 10, p. 273, Sep. 2024, doi: 10.1007/s10462-024-10884-2.

[12] X. Li *et al.*, "Interpretable deep learning: interpretation, interpretability, trustworthiness, and beyond," *Knowl. Inf. Syst.*, vol. 64, no. 12, pp. 3197–3234, Dec. 2022, doi: 10.1007/s10115-022-01756-8.

[13] M. Ahmad *et al.*, "Deep Belief Network Modeling for Automatic Liver Segmentation," *IEEE Access*, vol. 7, pp. 20585–20595, 2019, doi: 10.1109/ACCESS.2019.2896961.

[14] G. Altan, "Diagnosis of Coronary Artery Disease Using Deep Belief Networks," *Eur. J. Eng. Nat. Sci.*, vol. 2, no. 1, pp. 29–36, Feb. 2017.

[15] E. Pintelas, I. E. Livieris, and P. E. Pintelas, "A Convolutional Autoencoder Topology for Classification in High-Dimensional Noisy Image Datasets," *Sensors*, vol. 21, no. 22, p. 7731, Jan. 2021, doi: 10.3390/s21227731.

[16] D. T. Valivarthi, "OPTIMIZING CLOUD COMPUTING ENVIRONMENTS FOR BIG DATA PROCESSING," *Int. J. Eng.*, vol. 14, no. 2, pp. 1756–1775.

[17] M. Bhat, M. Rabindranath, B. S. Chara, and D. A. Simonetto, "Artificial intelligence,

machine learning, and deep learning in liver transplantation," *J. Hepatol.*, vol. 78, no. 6, pp. 1216–1233, Jun. 2023, doi: 10.1016/j.jhep.2023.01.006.

[18] Z. Du *et al.*, "Accurate Prediction of Coronary Heart Disease for Patients With Hypertension From Electronic Health Records With Big Data and Machine-Learning Methods: Model Development and Performance Evaluation," *JMIR Med. Inform.*, vol. 8, no. 7, p. e17257, Jul. 2020, doi: 10.2196/17257.

[19] A. Silverio, P. Cavallo, R. De Rosa, and G. Galasso, "Big Health Data and Cardiovascular Diseases: A Challenge for Research, an Opportunity for Clinical Care," *Front. Med.*, vol. 6, Feb. 2019, doi: 10.3389/fmed.2019.00036.

[20] F. Yasmin *et al.*, "Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future," *Rev. Cardiovasc. Med.*, vol. 22, no. 4, pp. 1095–1113, Dec. 2021, doi: 10.31083/j.rcm2204121.

[21] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *J. Big Data*, vol. 6, no. 1, p. 113, Dec. 2019, doi: 10.1186/s40537-019-0276-2.

[22] R. Budda and R. Pushpakumar, "Cloud Computing in Healthcare for Enhancing Patient Care and Efficiency," *Chin. Tradit. Med. J.*, vol. 1, no. 3, pp. 10–15, Jun. 2018.

[23] Y. Zhou and M. G. Varzaneh, "Efficient and scalable patients clustering based on medical big data in cloud platform," *J. Cloud Comput.*, vol. 11, no. 1, p. 49, Sep. 2022, doi: 10.1186/s13677-022-00324-3.

[24] V. K. Prasad and M. D. Bhavsar, "Monitoring IaaS Cloud for Healthcare Systems: Healthcare Information Management and Cloud Resources Utilization," *Int. J. E-Health Med. Commun. IJEHMC*, vol. 11, no. 3, pp. 54–70, Jul. 2020, doi: 10.4018/IJEHMC.2020070104.

[25] K. A. Bhavsar, A. Abugabah, J. Singla, A. A. AlZubi, A. K. Bashir, and Nikita, "A comprehensive review on medical diagnosis using machine learning," *Comput. Mater. Contin.*, vol. 67, no. 2, pp. 1997–2014, Jan. 2021, doi: 10.32604/cmc.2021.014943.

[26] W. Alsabhan and A. Alfadhly, "Effectiveness of machine learning models in diagnosis of heart disease: a comparative study," *Sci. Rep.*, vol. 15, no. 1, p. 24568, Jul. 2025, doi: 10.1038/s41598-025-09423-y.

[27] M.-H. Kuo, "Opportunities and Challenges of Cloud Computing to Improve Health Care Services," *J. Med. Internet Res.*, vol. 13, no. 3, p. e1867, Sep. 2011, doi: 10.2196/jmir.1867.

[28] S. Peddi, S. Narla, and D. T. Valivarthi, "Harnessing Artificial Intelligence and Machine Learning Algorithms for Chronic Disease Management, Fall Prevention, and Predictive Healthcare Applications in Geriatric Care," *Int. J. Eng. Res. Sci. Technol.*, vol. 15, no. 1, pp. 1–15, 2019.

[29] S. Narla, S. Peddi, and D. T. Valivarthi, "Optimizing Predictive Healthcare Modelling in a Cloud Computing Environment Using Histogram-Based Gradient Boosting, MARS, and SoftMax Regression," *Int. J. Manag. Res. Bus. Strategy*, vol. 11, no. 4, pp. 25–40, 2021.

[30] M. V. Devarajan, A. R. G. Yallamelli, R. K. M. K. Yalla, V. Mamidala, T. Ganesan, and A. Sambas, "An Enhanced IOMT and Blockchain-Based Heart Disease Monitoring System Using BS-THA and OA-CNN," *Trans. Emerg. Telecommun. Technol.*, vol. 36, no. 2, p. e70055, 2025, doi: 10.1002/ett.70055.