SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Edge Intelligence Systems: Optimizing Latency, Accuracy, and Resource Efficiency

Seema Verma[1], Prashant Anerao[2], Arivukkodi R[3], Dr. Shashikant Patil[4], Samundeeswari K[5], Piyush Pal[6], Mamatha Vayelapelli[7], Rahul Bhatt[8]

[1] Assistant Professor, Department of Computer Science & Application, Vivekananda Global University, Jaipur, Rajasthan, India. Email: seema.verma@vgu.ac.in ORCID: 0000-0002-9278-7343
[2] Department of Mechanical Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra 411037, India. Email: prashant.anerao@vit.edu
[3] Assistant Professor, Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, India. Email: arivukodir@maher.ac.in
[4] Professor, Department of uGDX School of Technology, Atlas SkillTech University, Mumbai, Maharashtra, India. Email: shashikant.patil@atlasuniversity.edu.in ORCID: 0000-0002-8835-908X
[5] Assistant Professor, Department of Commerce, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, India. Email: kscommerce@maher.ac.in
[6] Assistant Professor, Department of School of Engineering & Technology, Noida International University, Greater Noida, Uttar Pradesh, India. Email: piyush.pal@niu.edu.in
[7] Assistant Professor, Department of Computer Science & IT, Arka Jain University, Jamshedpur, Jharkhand, India. Email: mamatha.v@arkajainuniversity.ac.in
ORCID: 0000-0002-4694-0503
[8] Assistant Professor, Department of Computer Science & Engineering, Dev Bhoomi Uttarakhand University, Dehradun, Uttarakhand, India. Email: socse.rahul@dbuu.ac.in ORCID: 0009-0004-0486-9414

Abstract

Edge intelligence systems have emerged as an effective solution for improving computational efficiency and reducing processing overhead in smart healthcare Internet of Things (IoT) environments by enabling localized data analysis at edge devices. This research proposes a Sterna Migration-Optimized Lightweight Convolutional Neural Network (SM-L-CNN) method to enhance edge intelligence performance in smart healthcare IoT systems. Healthcare data are collected from wearable sensors, patient monitoring devices, and publicly available medical IoT datasets containing physiological signals and clinical parameters. The collected data undergo preprocessing using min–max normalization to standardize feature values and ensure uniform data representation across heterogeneous sensor inputs. Feature extraction is performed using Discrete Wavelet Transform (DWT) to decompose physiological time-series signals into multi-scale frequency components and capture essential health-related patterns. The extracted features are processed using a lightweight convolutional neural network deployed at the edge layer, while the Sterna Migration (SM) optimization algorithm is employed to optimize network parameters, filter weights, and resource allocation to reduce computational overhead and latency. The proposed SM-L-CNN framework improves classification accuracy, minimizes processing delay, and enhances energy-efficient resource utilization under constrained edge environments. Experimental results demonstrate superior performance in latency reduction, prediction accuracy of 0.953, precision of 0.925, recall of 0.958, and F1-score of 0.914. Computational efficiency compared to conventional edge intelligence models, providing a scalable and intelligent solution for smart healthcare monitoring and edge-based medical IoT applications.

Keywords: Internet of Things (IoT), Deep learning (DL), Edge intelligence, Healthcare, Optimization

1. Introduction

The smart healthcare and machine sectors are undergoing a significant transformation driven by innovations, advanced data analytics, personalized services, and artificial intelligence [1]. These advancements in healthcare applications have fundamentally reshaped how we perceive and deliver medical care. Smart sensors, whether wearable or embedded within devices, are capable of continuously collecting instantaneous data on behavior, movement, and location [2].An increasing need for personalized medical treatment, advancements in the creation of small, adaptable sensing systems, and the spread of technology in contemporary culture have all contributed to an expansion in the healthcare and wellness sector [3]. A Wireless Sensor Network (WSN) of the Internet of Things (IoT) comprises multiple wirelessly connected sensor nodes that collaborate and can self-organize, with topologies that can be deterministic or structure-less [4].Smart hospitals have been made possible through the application of IoT technologies in the healthcare industry. However, there are not many specific examples of fragmented infrastructures and inefficiencies [5]. The primary cause of the sensitivity of device latency during these systems is that medical systems, including healthcare monitoring, etc., require computations on massive volumes of health data [6].Smart sensors, whether wearable or embedded within devices, are capable of continuously collecting data on user behavior, movement patterns, and location, thereby enabling more accurate monitoring and analysis [7, 8].

Research Aim: To enhance edge intelligence in smart healthcare IoT systems, this research develops the Sterna Migration-Optimized Lightweight Convolutional Neural Network (SM-L-CNN) architecture. It aims to reduce latency, accuracy, and Resource efficiency, computing overhead, and energy consumption for instantaneous processing of diverse healthcare data at edge devices, while improving classification accuracy and resource utilization through optimal feature extraction and model parameter adjustment.

Research Organization: Section 1 discusses the introduction of Smart Healthcare IoT Systems and the background. Section 2 researches existing literature, highlighting integration gaps. Section 3 details the methodology, emphasizing data collection and the SM-L-CNN proposed model. Section 4 presents findings of the research. Section 5 concludes with findings, limitations, and future scope of the research.

2. Related Works

This Section reviews prior IoT healthcare research focusing on CNN/ML edge computing, achieving moderate accuracy but suffering from latency, energy, and scalability limitations in instantaneous applications systems. The instantaneous heart disease prediction via IoT and fog computing technology was developed in [9]. The method comprised several steps, such as data gathering, pre-processing, feature extraction, and classification through Q-Reinforcement Learning over Mist (QRLM) Algorithm in machine learning techniques at the fog layer. The result shows the accuracy of the system was 91.2%. Limitations were data-set dependency and lower robustness to noise from sensors. Improving the prevention of non-communicable diseases through smart health architecture with IoT, edge computing, and AI was investigated [10]. This model makes use of 12 regression models to predict glucose concentration without invasive procedures, with the help of the proposed edge-based cloud layer. It turns out that CatBoost is the best among them with an R^2 of 0.81. Limitations involve dataset dependency, generalization, and the lack of deployment verification. Healthcare monitoring for adults using IoT, Digital Twins (DT), and blockchain technology was explored [11]. Convolutional Neural Network (CNN) identifies anomalies on the fly, while Reputation-based Byzantine Fault Tolerance (RBFT) blockchain ensures secure data transmission. However, there are issues that arise with the methodology, including heavy computation, high latency, and poor scalability. An IoT-fog healthcare framework, where wearables detect cardiovascular and diabetes diseases, was analyzed [12]. An artificial intelligence framework, tuned via Crow Search Optimization (CSO), is employed to optimize the weights and biases used in classification. 97.26% and 96.16% accuracy are attained in the diagnoses of diabetes and heart diseases, respectively. The model exhibits risks of misdiagnoses in cases involving multiple diseases and lacks generalization between datasets. A frameworks for intelligent health care services utilizing an edge-cloud continuum for the management of voluminous healthcare data was examined [13]. It uses the proposed cloud computing method for electronic health records, medical images, and medical monitoring data. The system attains an accuracy rate of 88-92%, which is proof of effective disease management, but some challenges include network dependency, challenges with rural connectivity, and security challenges. Increased privacy and security in federated learning by means of edge intelligence, a framework named Serverless privacy edge intelligence-enabled federated learning (SPEI-FL)was investigated [14]. It uses various methods such as federated edge learning, authentication processes, and serverless computing, enabling it to reach high accuracy levels of around 90% as opposed to only 81% by other frameworks. The proposed architecture remains vulnerable to intelligent cyber attacks and high computation costs at the edge.Instantaneous health tracking, heart disease prediction, and medication dispensing are realized via IoT and edge computing, which was developed [15]. A Deep Neural Network (DNN) is used to perform inference of patient data via the edge-cloud paradigm; the system achieves 96.15% accuracy. It relies extensively on edge architecture and is not scalable to different types of data and environments. The processing of huge amounts of medical sensor data, instantaneous improvements, security, and privacy were investigated through the use of an edge-cloud ML framework [16]. The technique employs edge computing with machine learning, task offloading, and signal filtering for signal detection. It attains about 85-90% efficiency improvement and lower latency and resource efficiency, but its challenges include high communication cost, higher energy consumption, and high reliance on cloud availability.Heart Disease Prediction with the Aid of IoT Devices Using Cloud-Based Electronic Health Records for Proactive Healthcare was investigated [17]. It uses a bidirectional long short-term memory (Bi-LSTM) model to research the time series dataset for heart disease prediction; this technique attains a significant accuracy. It demands extensive computing power and is reliant on extensive labeled training datasets. Deep Learning-Based Diagnosis of Medical Images Using 6G IoT Internet of Medical Things (IoMT) Network was developed [18]. The preprocessing of computed tomography (CT) images is done, followed by the extraction of features from the MobileNetV3 model. Using optimization by the arithmetic optimization algorithm based on the Hunger Games Search (AOA+HGS), with an accuracy rate of 87.30%. It suffers from high computational complexity and limited instantaneous scalability. An AI-based IoT Wearable System for Instantaneous Prediction and Assistance of Diabetic Patients Using Wristband Sensors was investigated [19]. Random Forest (RF) and Decision Tree (DT) were found to be most accurate with NIR glucose sensing, ML, and Fuzzy Logic Decision System. Restrictions include small sample size, dependency on sensors, and lack of generalizability. AI-driven Internet of Things and Cyber-Physical Systems (IoT-CPS) system development for disease and falls detection utilizing database and sensor data from wearable devices was studied [20]. The above technique based on AI-Enabled IOT-CPS has been proposed, and it provides better accuracy than other conventional methods (85.1), but it faces many challenges, such as security issues, ethics issues, scalability issues, and no real-time clinical use cases.

3. Methodology

The methodology focuses on the use of efficient health care monitoring through edge devices using data processing and lightweight deep learning approaches. The process begins with wearable data collection, followed by Min-Max normalization, DWT-based feature extraction, LCNN classification, and performance enhancement using the SMO algorithm for improved accuracy in healthcare monitoring systems. Figure 1 illustrates the method workflow of a proposed SM-L-CNN framework for a healthcare system.

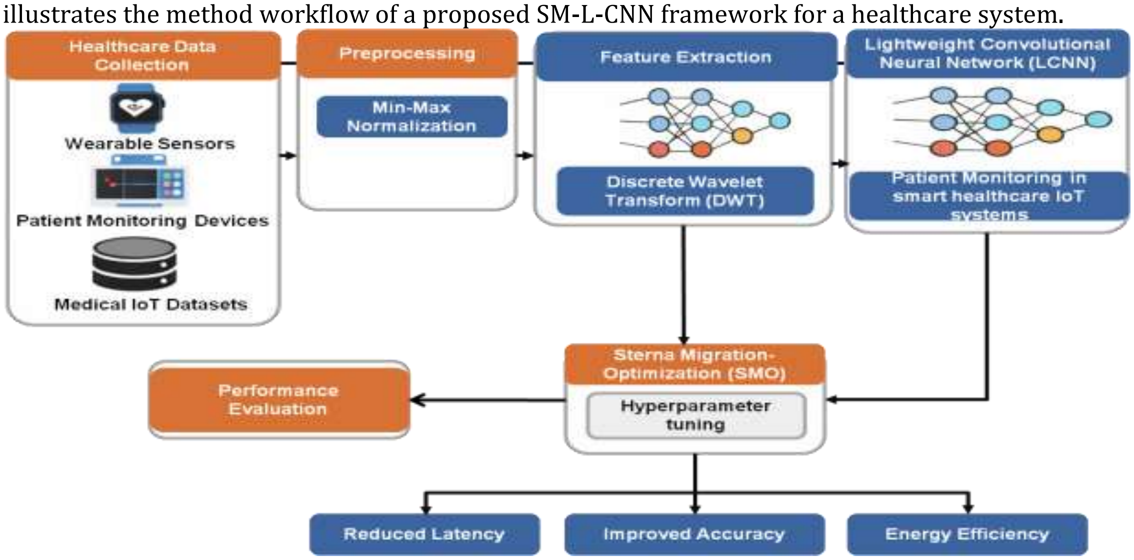


Figure 1: Method flow of the proposed SM-L-CNN IoT framework for intelligent healthcare systems

3.1. Data Collection

The designed SM-L-CNN IoT edge intelligence system for intelligent healthcare application makes use of the Healthcare IoT Dataset, which consists of 4000 data points with 26 mixed attributes acquired from wearable sensors, patient monitoring equipment, and edge-enabled IoT devices in healthcare. The dataset is partitioned into an 80:20 ratio for training and testing purposes, respectively. The dataset comprises physiological attributes (body temperature, heart rate, blood pressure, oxygen saturation, and respiration rate), motion-related attributes (acceleration and orientation attributes), signal attributes (low frequency energy, mid frequency energy, high frequency energy, low frequency entropy, mid frequency entropy, high frequency entropy), and edge computing attributes (Latency, Accuracy and Resource efficiency, CPU utilization, memory utilization, energy consumption, and network conditions). The dataset is accessible at Kaggle (https://www.kaggle.com/datasets/zara2099/edge-healthcare-iot-intelligence-dataset/data).

3.2. Data Preprocessing Using Min-Max Normalization

In smart healthcare IoT systems, heterogeneous sensor data are normalized using Min–Max scaling to ensure consistent feature representation. This reduces scale variations, improves model stability, and enhances convergence efficiency in the edge-deployed L-CNN. The transformation is defined in Equation (1).

$$W_{minmax} = \frac{W - W_n}{W_N - W_n} \quad (1)$$

W_{minmax} is the normalized input feature value considered as a standardized healthcare input for the SM-L-CNN model. W stands for the original raw input feature value collected from the wearable sensors and IoT devices. W_n and W_N stand for the minimum and maximum values of a feature, respectively. These values are utilized in Min-Max Scaling for the process of data normalization in order to increase performance in IoT-based healthcare analytics.

3.3. Feature Extraction Using Discrete Wavelet Transform (DWT)

Following min–max normalization for data pre-processing in SM-L-CNN, feature extraction is carried out via the DWT. This transformation decomposes one-dimensional physiological time-series signals into multiple sub-band time–frequency components, effectively capturing the non-stationary characteristics of healthcare data. In DWT, the signal is analyzed using scaled and shifted versions of a mother wavelet function, as defined in Equation (2).

$$\psi_{i,l}(s) = \frac{1}{\sqrt{2^i}} \psi \frac{s - k2^i}{2^i} \quad (2)$$

Scaled and shifted wavelet function $\psi_{i,l}(s)$ used to perform DWT-based feature extraction for analyzing physiological time-series healthcare signals of IoT sensors at multiple resolutions. The mother wavelet function $\psi \frac{s-k2^i}{2^i}$ is the (prototype function) used for decomposing healthcare sensor signals into their time-frequency representations. i control the frequency of healthcare data. lis a translation (shift) capturing patterns from a wearable sensor.s is a time variable to collect the signal from a wearable sensor, and k represents the time shift for the position of the health signal.$\frac{1}{\sqrt{2^i}}$is a normalization factor that works as a healthcare signal decomposition. The wavelet coefficients are the result of projecting the signal x(s) onto the wavelet basis, as represented by the following Equation (3).

$$\gamma_{i,l} = \int_{\infty}^{\infty} x(s) \frac{1}{\sqrt{2^i}} \psi \frac{(s - k2^i)}{2^i} dt \quad (3)$$

$\gamma_{i,l}$Wavelet coefficient of extracted feature from healthcare IoT signals. $\int_{\infty}^{\infty} x(s)$is a normalized time series input of a healthcare signal. s is a continuous time variable indicating healthcare data. ψ Mother wavelet function used for feature extraction of healthcare signals for efficient edge-based analysis. i Scale factor that controls the resolution level of signal decomposition in the SM-L-CNN model.l is a translation capturing patterns from a wearable sensor in time intervals.k Translation factor indicating the amount of time shift for monitored health signals.$\frac{1}{\sqrt{2^i}}$ Normalization factor used to maintain consistency of signals. dt Infinitesimally small time step is used in the integration process of the input signal from the healthcare IoT system. Implementation of the DWT operation is performed using two filter banks and down-sampling, which can be mathematically expressed as Equation (4).

$$z_{low}[m] = (w * h)[m] \downarrow 2, z_{high}[m] = (w * g)[m] \downarrow 2 \quad (4)$$

Low-frequency coefficients $z_{low}[m]$, which correspond to the slow variations found in health signals generated with the SM-L-CNN model. High-frequency detail coefficients $z_{high}[m]$, fast variation, and abnormal signals in a wearable sensor. Input health signal w that undergoes normalization and pre-processing.h a low-pass filter is used to extract trends from healthcare signals using the discrete wavelet transformation method. g a high-pass filter is used to extract sharp variation components from health signals. $(w * h)$ Convolution process is carried out between the input health signal and the filter mask to generate wavelet sub-band components using the SM-L-CNN Model. $\downarrow 2$ sampling by two to lower the health signal's resolution without sacrificing medical data.

3.4. Patient Monitoring using Sterna Migration-Optimized Lightweight Convolutional Neural Network (SM-L-CNN) in smart healthcare IoT systems

The proposed SM-L-CNN framework utilizes an LCNN to perform efficient feature extraction and classification of healthcare IoT data for patient monitoring. The SMO algorithm adaptively tunes hyperparameters, including weights, learning rate, and resource allocation within bounded limits, reducing latency, accuracy, energy consumption, and computational cost while ensuring accurate and resource-efficient edge deployment.

3.4.1 Lightweight Convolutional Neural Network (LCNN)

The Lightweight Convolutional Neural Network (LCNN) performs feature extraction and classification, learning hierarchical patterns with reduced complexity. Mathematically, LCNN processes features through sequential convolutional operations, as defined in Equation (5).

$$\mathrm{G_j} = \sigma\left(\mathrm{X_j} * \mathrm{G_j} - 1 + \mathrm{a_j}\right) \quad (5)$$

$\mathrm{G_j}$ represents the output feature map of the $\mathrm{j^{th}}$ convolutional layer in a lightweight CNN, capturing healthcare patterns from healthcare IoT signals. σ denotes the ReLU activation function, which improved the learning performance in Healthcare IoT Systems. $\mathrm{X_j}$ refers to the convolutional filter weights for effective feature extraction from a complex healthcare dataset.$*$ indicates the convolution operation for extracting meaningful input from a healthcare sensor. $\mathrm{G_j} - 1$ signifies the input feature map from the prior convolutional layer, derived from normalized healthcare IoT data. $\mathrm{a_j}$ is the bias term of the $\mathrm{j^{th}}$ layer, modified during training to enhance model fitting and classification accuracy. Following feature extraction, the LCNN transfers learned features into probability distributions over health states to conduct classification, expressed in Equations (6 and 7).

$$\mathrm{o_j} = \mathrm{Softmax}\left(\mathrm{y_j}\right) \quad (6)$$

In edge-based smart healthcare IoT classification, $\mathrm{o_j}$ denotes the predicted probability for the $\mathrm{j^{th}}$ healthcare class (e.g., normal or abnormal). The Softmax function converts raw logits into normalized probabilities for multi-class healthcare decisions, while $\mathrm{y_j}$ signifies the logit value from the final dense layer of the lightweight CNN after feature extraction from IoT sensor data.

$$\mathrm{z} = argmax\left(o_j\right) \quad (7)$$

z indicates the final predicted healthcare condition in the smart healthcare IoT system. $argmax$ is the decision function that ensures an edge layer for timely healthcare intervention.

3.5. Hyperparameter tuning with Sterna Migration-Optimization (SMO)

Following feature extraction using the DWT, the proposed SM-L-CNN framework employs the SMO algorithm to perform effective hyperparameter tuning and feature optimization. This optimization stage plays a crucial role in enhancing the performance of the LCNN deployed at the edge layer. The design of the SMO algorithm makes sure that all the parameters of the network, like filters, learning rate, and resource management, are optimized for efficient use of the computational resources available in such smart IoT healthcare systems. The optimization of the algorithm ensures low latency, computation cost, and energy consumption without sacrificing accuracy. The decision space in the SMO technique uses upper and lower limits, expressed in Equation (8).

$$L = (L_1, L_2, \dots, L_c), U = (U_1, U_2, \dots, U_c) \quad (8)$$

L and $U$are lower and upper bound vector defining the lowest permissible values of optimization parameters. L_c and $U_c$are the lower and upper bounds of the c^{th} decision variable defining the lowest constraint. c dimensionality of the optimization problem. To simulate IoT healthcare signals' movement of SM, various cluster centers are created inside the search space using Equation (9).

$$T_i = L + \gamma_i(U - L), \gamma_i \sim U(0.2,0.8), i = 1,2, \dots, n \quad (9)$$

T_i First cluster center associated with a possible solution while optimizing the parameters of LCNN. $L$and U lower and upper bounds vector of minimum permissible limits in healthcare signals. γ_i Scale factor randomly generated to calculate the improved diversity in the healthcare system. $U(0.2,0.8)$ Uniform distribution range considered to generate random cluster center values, where i represents the cluster center number, which varies among sub-populations. $n$is the total number of clusters considered for generating multiple solutions. Each cluster generates solutions through a fan-shaped dispersion strategy defined as Equation (10).

$$O_{i,j}^0 = T_i + \rho_{i,j}(U - L) \circ \in_{i,j} \quad (10)$$

$O_{i,j}^0$ Initial solution showing the initial population in cluster i. $\rho_{i,j}$ is the distribution coefficient of the healthcare signal controlling the spread of the individuals from the cluster center. $\circ$ It is hadamard operator, $\epsilon_{i,j}$ direction vector providing variation for generating candidates. j is the individual number showing the individual candidate in cluster i. To ensure feasibility, boundary correction is applied as Equation (11).

$$O_{i,j,c}^0 = min\{max(O_{i,j,c}^0, L_c), U_c\} \tag{11}$$

$O_{i,j,c}^0$ represents corrected parameter values for subject j at cluster i in dimension c. $min\{max\ (\cdot)\}$ is improved model performance of the Healthcare system, and also improves classification accuracy, reducing latency in smart healthcare edge environments.

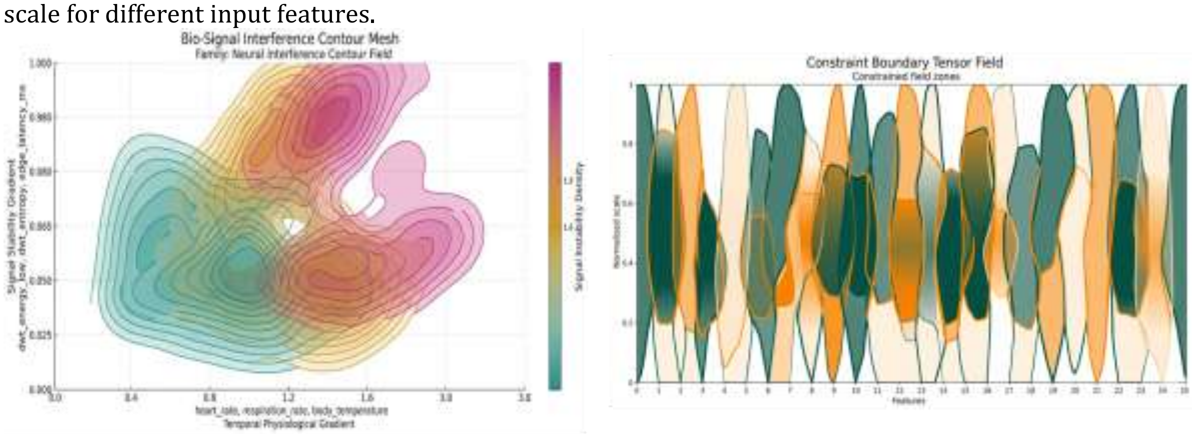
Hyperparameter configuration of the proposed SM-L-CNN model: The input size is set to 128 × 128 × 3. The batch size varies between 16 and 64, and the number of training epochs ranges from 50 to 150. The learning rate is selected within 0.0001 to 0.001 using the Adam optimizer. The network consists of 3 to 6 convolutional layers with 32 to 128 filters and a kernel size of 3 × 3. A dropout rate between 0.2 and 0.5 is applied to prevent overfitting, while dense layer units range from 64 to 256 with L2 regularization (0.0001). For optimization, the SMO algorithm uses a population size of 20 to 50, iterations between 30 and 100, and a migration factor ranging from 0.1 to 0.5.

4. Result

The experimental evaluation of the SM-L-CNN framework shows significant enhancements in edge intelligence performance for smart healthcare IoT environments. The experimental setup includes an Intel Core i7 processor, NVIDIA RTX 3060 GPU, 16–32 GB RAM, and 512 GB SSD with a 64-bit system. The software environment uses Windows 10, Python 3.8+, TensorFlow/PyTorch, Scikit-learn, NumPy, Pandas, Matplotlib, Seaborn, and Jupyter Notebook or VS Code.

4.1 Exploratory Analysis

Figure 2(a) presents the suggested mapping technique, which has successfully captured the peaks of neural interference densities at about 1.8. The technique retains high resolution on the temporal gradient and ensures localization of the bio-signal energies. Visualize the density and overlapping of neural interference signalsto locate areas of high signal instability. Figure 2(b) shows stability in handling 25 different features. The normalized scale is maintained at a steady range between 0.2 and 0.8 to prevent any boundary leakage during the transformation process. The distribution of the features and boundaries is a tensor field in a normalized scale for different input features.



(a) (b)

Figure 2: Data Visualization (a) Bio-Signal Interference, and (b) Constraint Boundary Tensor Field

4.2. Evaluation Metrics

The performance of the proposed SM-L-CNN framework is evaluated using classification metrics in an instantaneous patient monitoring context, where patient conditions are categorized as Delayed, Stable, Critical, and At_Risk. Accuracy measures the overall correctness of predictions from IoT-based physiological data. Precision focuses on measuring how well the algorithm is able to correctly classify each class, ensuring minimal erroneous predictions especially with Critical and At_Risk cases. On the other hand, recall measures how well the real patient condition can be identified, focusing especially on high risk patient conditions for proper interventions. This balance between precision and recall is ensured by using the F1-Score metric which helps ensure consistent performance even with class imbalance.

4.3. Performance Evaluation

The efficiency of the SM-L-CNN model is compared with existing machine learning and IoT-based algorithms, such as RF and DT [19], and the AI-enabled IoT-CPS model [20], using evaluation metrics including Accuracy, Precision, Recall, and F1-Score. The SM-L-CNN model achieves the highest accuracy of 0.953, clearly outperforming DT (0.80), RF (0.90), and IoT-CPS (0.851) [19], [20]. In terms of precision, the proposed model attains 0.925, exceeding DT (0.65), RF (0.74), and IoT-CPS (0.863). Similarly, its recall score of 0.958 indicates its excellent capacity to recognize relevant instances and is better than that of all baselines. The F1-Score of 0.914 even shows the balanced nature of its performance, which is much better than DT's (0.65), RF's (0.77), and IoT-CPS's (0.864). Even though RF is better than DT and IoT-CPS is better than both DT and RF, still, our model is superior to them. This can be seen in Table 1 and Figure 3 below.

Table 1: Comparison analyses of Existing and Proposed Method

Metrics	DT [19]	RF [19]	AI-Enabled IoT-CPS [20]	SM-L-CNN [Proposed]
Accuracy	0.80	0.90	0.851 (85.1)	0.953
Precision	0.65	0.74	0.863 (86.3)	0.925
Recall	0.66	0.80	0.865 (86.5)	0.958
F1-Score	0.65	0.77	0.864 (86.4)	0.914

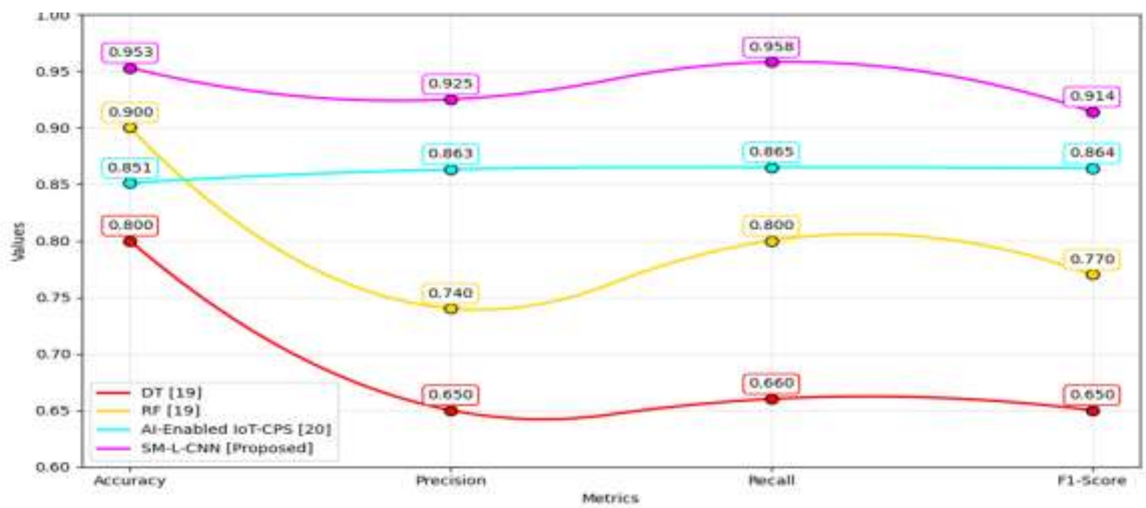


Figure 3: Performance Comparison Metrics of the existing and proposed models

4.4. Discussion

The DT model [19], in particular, is relatively inefficient, as the algorithm is incapable of adequately detecting complex non-linear dependencies between physiological signals, hence unable to be used in edge intelligence. RF model [19] demonstrates better prediction potential due to the usage of ensembles; nevertheless, it seems that it possesses insufficient features for efficient processing of high-dimensional data in Internet of Things healthcare-related applications. Moreover, it lacks the necessary optimizations that would allow using it in edge intelligence applications. While AI-enabled IoT-CPS model [20] shows improved predictability due to the combination of IoT technology and AI in health care application areas, it has difficulties in achieving scalability, excessive computational cost, and insufficient adaptability to an edge computing setting. To address this problem, the proposed SM-L-CNN approach exploits DWT feature extraction and Sterna migration-based lightweight CNN in order to improve feature representation and prediction capability without increasing computational complexity. Due to such a framework, it will be possible to achieve adequate optimization of parameters in conjunction with energy saving at the edge layer.

5. Conclusion

In line with the set goals, the research has been successful in developing a very accurate edge intelligence architecture for smart healthcare IoT through SM-L-CNN. The combination of lightweight DL methods and the Sterna Migration optimization algorithm contributes positively to system performance through latency reduction, improved accuracy, and reduced computational complexity. The application of Min-Max normalization and Discrete Wavelet Transform (DWT) makes it possible for efficient preprocessing and feature extraction in physiological data. Experiments indicate high performance, scoring accuracy of 0.953, precision of 0.925, recall of 0.958, and F1-score of 0.914. These results prove the success of the developed architecture. Overall, the SM-L-CNN architecture provides an energy-efficient and intelligent solution for patient health monitoring and classification. Nonetheless, its reliance on edge devices with modest computing abilities limits its performance when power consumption is tightly controlled, especially when dealing with multimodal physiological data. In addition, the Sterna Migration algorithm introduces extra complexity and security risks. Thus, further studies should consider improvements on multimodal data fusion and adaptation strategies, such as federated learning and online learning.

REFERENCES

1. Ala, A., Simic, V., Pamucar, D. and Bacanin, N., 2024. Enhancing patient information performance in Internet of things-based smart healthcare system: Hybrid artificial intelligence and optimization approaches. Engineering Applications of Artificial Intelligence, 131, p.107889. https://doi.org/10.1016/j.engappai.2024.107889
2. Rajagopal, D. and Subramanian, P.K.T., 2025. AI augmented edge and fog computing for Internet of Health Things (IoHT). PeerJ Computer Science, 11, p.e2431.https://doi.org/10.7717/peerj-cs.2431
3. Shumba, A.T., Montanaro, T., Sergi, I., Fachechi, L., De Vittorio, M. and Patrono, L., 2022. Leveraging IoT-aware technologies and AI techniques for real-time critical healthcare applications. Sensors, 22(19), p.7675.https://doi.org/10.3390/s22197675
4. Jabeen, T., Jabeen, I., Ashraf, H., Jhanjhi, N.Z., Yassine, A. and Hossain, M.S., 2023. An intelligent healthcare system using IoT in wireless sensor network. Sensors, 23(11), p.5055. https://doi.org/10.3390/s23115055
5. Yadegari, F. and Asosheh, A., 2025. A unified IoT architectural model for smart hospitals: enhancing interoperability, security, and efficiency through clinical information systems (CIS). Journal of Big Data, 12(1), p.149.https://doi.org/10.1186/s40537-025-01197-4
6. Wang, K., Kong, S., Chen, X. and Zhao, M., 2024. Edge computing empowered smart healthcare: monitoring and diagnosis with deep learning methods. Journal of Grid Computing, 22(1), p.30.https://doi.org/10.1007/s10723-023-09726-2
7. Rafique, W., 2025. Ml-raspf: A machine learning-based rate-adaptive framework for dynamic resource allocation in smart healthcare iot. Algorithms, 18(6), p.325.https://doi.org/10.3390/a18060325
8. Umer, M., Sadiq, S., Karamti, H., Karamti, W., Majeed, R. and Nappi, M., 2022. IoT based smart monitoring of patients' with acute heart failure. Sensors, 22(7), p.2431.https://doi.org/10.3390/s22072431
9. Tripathy, S.S., Imoize, A.L., Rath, M., Tripathy, N., Bebortta, S., Lee, C.C., Chen, T.Y., Ojo, S., Isabona, J. and Pani, S.K., 2022. A novel edge-computing-based framework for an intelligent smart healthcare system in smart cities. Sustainability, 15(1), p.735. https://doi.org/10.3390/su15010735
10. Cruz Castaneda, W.A. and Bertemes Filho, P., 2024. Improvement of an Edge-IoT Architecture Driven by Artificial Intelligence for Smart-Health Chronic Disease Management. Sensors, 24(24), p.7965. https://doi.org/10.3390/s24247965
11. Bhatia, M., 2024. An AI-enabled secure framework for enhanced elder healthcare. Engineering Applications of Artificial Intelligence, 131, p.107831. https://doi.org/10.1016/j.engappai.2023.107831
12. Tang, Z., Tang, Z., Liu, Y., Tang, Z. and Liao, Y., 2024. Smart healthcare systems: A new IoT-Fog based disease diagnosis framework for smart healthcare projects. Ain Shams Engineering Journal, 15(10), p.102941. https://doi.org/10.1016/j.asej.2024.102941
13. Gao, X., He, P., Zhou, Y. and Qin, X., 2024. A smart healthcare system for remote areas based on the edge–cloud Continuum. Electronics, 13(21), p.4152. https://doi.org/10.3390/electronics13214152
14. Akter, M., Moustafa, N. and Turnbull, B., 2024. SPEI-FL: Serverless privacy edge intelligence-enabled federated learning in smart healthcare systems. Cognitive Computation, 16(5), pp.2626-2641. https://doi.org/10.1007/s12559-024-10310-3

15. Hameurlaine, M., Moussaoui, A. and Bensalah, M., 2025. Real-Time Smart Healthcare System Based on Edge-Internet of Things and Deep Neural Networks for Heart Disease Prediction. Informatica, 49(1). https://doi.org/10.31449/inf.v49i1.6875
16. Alnaim, A.K. and Alwakeel, A.M., 2023. Machine-learning-based IoT–edge computing healthcare solutions. Electronics, 12(4), p.1027. https://doi.org/10.3390/electronics12041027
17. Nancy, A.A., Ravindran, D., Raj Vincent, P.D., Srinivasan, K. and Gutierrez Reina, D., 2022. Iot-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning. Electronics, 11(15), p.2292. https://doi.org/10.3390/electronics11152292
18. Elaziz, M.A., Dahou, A., Mabrouk, A., Ibrahim, R.A. and Aseeri, A.O., 2023. Medical image classifications for 6G IoT-enabled smart health systems. Diagnostics, 13(5), p.834. https://doi.org/10.3390/diagnostics13050834
19. Maqbool, S., Bajwa, I.S., Maqbool, S., Ramzan, S. and Chishty, M.J., 2023. A smart sensing technologies-based intelligent healthcare system for diabetes patients. Sensors, 23(23), p.9558. https://doi.org/10.3390/s23239558
20. Ramasamy, L.K., Khan, F., Shah, M., Prasad, B.V.V.S., Iwendi, C. and Biamba, C., 2022. Secure smart wearable computing through artificial intelligence-enabled internet of things and cyber-physical systems for health monitoring. Sensors, 22(3), p.1076. https://doi.org/10.3390/s22031076