An Optimized Deep Learning Model for Detecting Sick Building Syndrome in Healthcare Environments

Hayder Qasim Flayyih[1], Jumana Waleed[1], Amer M. Ibrahim[2] and Mohammed Y. Shakor[3]
[1]*Department of Computer Science, College of Science, University of Diyala, 32001 Baqubah, Iraq*
[2]*Department of Civil Engineering, College of Engineering, University of Diyala, 32001 Baqubah, Iraq*
[3]*Department of Information Technology, College of Computer and Information Technology, University of Garmian, 46021 Kalar, Iraq*
{scicomphd222303, jumana.waleed}@sciences.uodiyala.edu.iq, amer.ibrahim@eng.uodiyala.edu.iq, mohammed.yousif@garmian.edu.krd

Keywords: SBS & IAQ, Deep Learning Model Optimization, Neural Architecture Search, Progressive Learning, Singular Value Decomposition.

Abstract: Early detection of Sick Building Syndrome (SBS) in healthcare environments is vital to safeguard occupant health, especially in regions like Iraq where structured indoor air quality (IAQ) monitoring systems are largely unavailable. This study introduces a novel deep learning framework for real-time SBS prediction, developed using the first publicly available IAQ dataset collected from Baqubah Teaching Hospital, Diyala, Iraq. The dataset spans eight months and includes ten environmental parameters such as CO_2, TVOC, $PM_{2.5}$, PM_{10}, CO, O_3, temperature, humidity, air quality index, and light intensity. The proposed framework employs a hybrid 1D-CNN-BiLSTM model designed to capture both local spatial correlations and bidirectional temporal dependencies in IAQ signals. To enhance its suitability for real-world deployment, particularly on resource-limited devices, the model undergoes a series of multi-level optimization techniques that significantly improve efficiency while maintaining high predictive accuracy. Experimental evaluations against multiple deep learning baselines demonstrate that the optimized model achieves a strong balance between accuracy and computational performance. These results highlight the framework's practicality for continuous, real-time monitoring of SBS risk factors. Beyond its technical contribution, this research provides the first IAQ dataset from Iraq and establishes a foundation for intelligent environmental health management in healthcare buildings across developing regions.

1 INTRODUCTION

Indoor Air Quality (IAQ) is a critical determinant of human health, particularly in healthcare environments where patients and medical staff are continuously exposed to confined indoor conditions. Poor IAQ has been strongly linked to SBS, a condition characterized by respiratory irritation, fatigue, cognitive impairment, and headaches [1]. According to the World Health Organization (WHO), inadequate IAQ contributes to over 3.2 million premature deaths each year, underscoring the global urgency for effective IAQ monitoring and prediction systems [2].

In Iraq, the situation is especially concerning. Aging hospital infrastructure, inadequate ventilation systems, and rapid urbanization have intensified SBS risks in healthcare buildings. However, no structured IAQ datasets or automated monitoring systems are currently available, and existing approaches rely mainly on manual inspections and occupant surveys. These methods are inherently subjective, prone to inaccuracy, and lack real-time responsiveness, emphasizing the urgent need for intelligent, automated solutions [3].

Recent advances in Artificial Intelligence (AI), particularly in Deep Learning (DL), have enabled powerful approaches for modeling complex environmental data. DL architectures such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and hybrid CNN-LSTM frameworks have shown promising results in capturing nonlinear and temporal dependencies in IAQ patterns [4]. Nonetheless, most existing studies focus on pollutant forecasting or general IAQ prediction without addressing real-time

applicability in resource-constrained healthcare environments. Additionally, the absence of effective model optimization strategies often leads to computationally intensive frameworks that limit deployment feasibility [5].

To address these challenges, the present study makes several key contributions that are both methodological and practical.

- Dataset development. The first structured IAQ dataset from Iraq, collected at Baqubah Teaching Hospital, Diyala.
- Hybrid DL framework. A novel 1D-CNN-BiLSTM model that integrates spatial and temporal learning for SBS detection.
- Optimization techniques. Multi-level model optimization to improve efficiency and reduce inference time without compromising accuracy.
- Comparative analysis. Benchmarking against several baseline DL architectures to demonstrate superior performance and computational efficiency.
- Practical validation. Evaluation for real-time SBS monitoring in hospitals, emphasizing deployability in low-resource healthcare settings.

The novelty of this research lies in introducing the first publicly available IAQ dataset from Iraq and in demonstrating that an optimized hybrid 1D-CNN-BiLSTM model can achieve an effective balance between predictive accuracy and computational efficiency-making it suitable for real-time hospital deployment.

This study aims to develop and optimize a lightweight hybrid deep learning model for real-time SBS detection in Iraqi hospitals.

2 RELATED WORKS

2.1 Deep Learning Approaches for Indoor Air Quality and SBS Prediction

Recent research has increasingly adopted deep learning (DL) approaches for modeling Indoor Air Quality (IAQ) and classifying conditions related to Sick Building Syndrome (SBS). Poor IAQ has been closely linked to SBS across a variety of indoor environments, including workplaces, hospitals, and residential spaces [6].

Shi et al. (2018) [7] developed an enhanced Backpropagation (BP) neural network to classify indoor temperature and humidity in industrial environments using real environmental data from Chongqing, China. The model achieved strong predictive performance, with R^2 values of 0.9897 for temperature and 0.9778 for humidity. Similarly, Zhao et al. (2018) [8] utilized a Recurrent Neural Network (RNN) for air quality prediction in three industrial U.S. cities. Their findings demonstrated that RNNs outperformed traditional Support Vector Machine (SVM) and Random Forest (RF) models, reinforcing the value of DL methods in time-series air quality forecasting.

Elmaz et al. (2021) [9] advanced this research by developing a CNN-LSTM hybrid model that captured both spatial and temporal dependencies in IAQ data. Their results showed that CNN-LSTM consistently outperformed standalone MLP and LSTM models across different forecasting intervals (1-120 minutes). Bao et al. (2022) [10] further enhanced this approach through the FL-CNN-LSTM model, integrating fuzzy logic with CNN-LSTM for IAQ forecasting using data from indoor $PM_{2.5}$ sensors in Shanghai. The model achieved higher accuracy and interpretability compared with traditional CNN-LSTM baselines.

Building upon hybrid frameworks, Fernandes and Gonçalves (2023) [11] proposed a BiLSTM model for short- and long-term IAQ prediction. Their model achieved an MAE of 2.892 and an RMSE of 8.703 for one-minute predictions, confirming the strength of bidirectional learning in environmental forecasting. Zhu et al. (2024) [12] introduced a CNN-BiLSTM model enhanced with Adaptive Particle Swarm Optimization (APSO) for air quality prediction. The optimization process improved model accuracy and robustness, achieving an MAE of 29.19 and an RMSE of 38.93. Wang et al. (2024) [13] developed a lightweight multiscale dilated CNN framework optimized for IoT edge and mobile devices, reducing parameters by 86.7% and floating-point operations by 88.5%, while maintaining a Top-1 accuracy of 94.2%.

An overview of these studies, including their datasets, architectures, and performance outcomes, is summarized in Table 1.

While previous research has established the effectiveness of CNN-LSTM hybrids, fuzzy logic integration, and optimization-based models for IAQ forecasting, several limitations remain. Most studies have been conducted in industrialized countries, focusing primarily on general air quality prediction rather than SBS classification in healthcare environments. Moreover, many models are computationally intensive, limiting their use in real-time or resource-constrained deployments.

Table 1: Comparative summary of previous DL-based approaches for IAQ and SBS prediction.

Authors (Year)	Model	Dataset	Advantages	Limitations
Shi et al. (2018) [7]	Improved BP Neural Network	Chongqing industrial data	Accurate temp. & humidity prediction	Weak for long-term RH forecasting
Zhao et al. (2018) [8]	RNN	US EPA AQ data	Strong sequential modeling	No deployment focus
Elmaz et al. (2021) [9]	CNN-LSTM	Building Z, Antwerp	High IAQ forecast accuracy	Long-term error accumulation
Bao et al. (2022) [10]	FL-CNN-LSTM	Shanghai PM2.5 dataset	Integrates fuzzy logic, interpretable	High computational cost
Fernandes & Gonçalves (2023) [11]	BiLSTM	GAMS IAQ dataset	Strong short-term forecasting	Limited optimization
Zhu et al. (2024) [12]	APSO-CNN-BiLSTM	Xi'an AQI dataset	Dynamic hyperparameter tuning	Prediction only, no classification
Wang et al. (2024) [13]	Multiscale Dilated CNN	GAOs-2 (image-based AQ)	High accuracy, lightweight	Limited generalization
This Work	Optimized 1D-CNN-BiLSTM	Iraqi hospital IAQ dataset	High accuracy, real-time deployable	First SBS dataset in Iraq

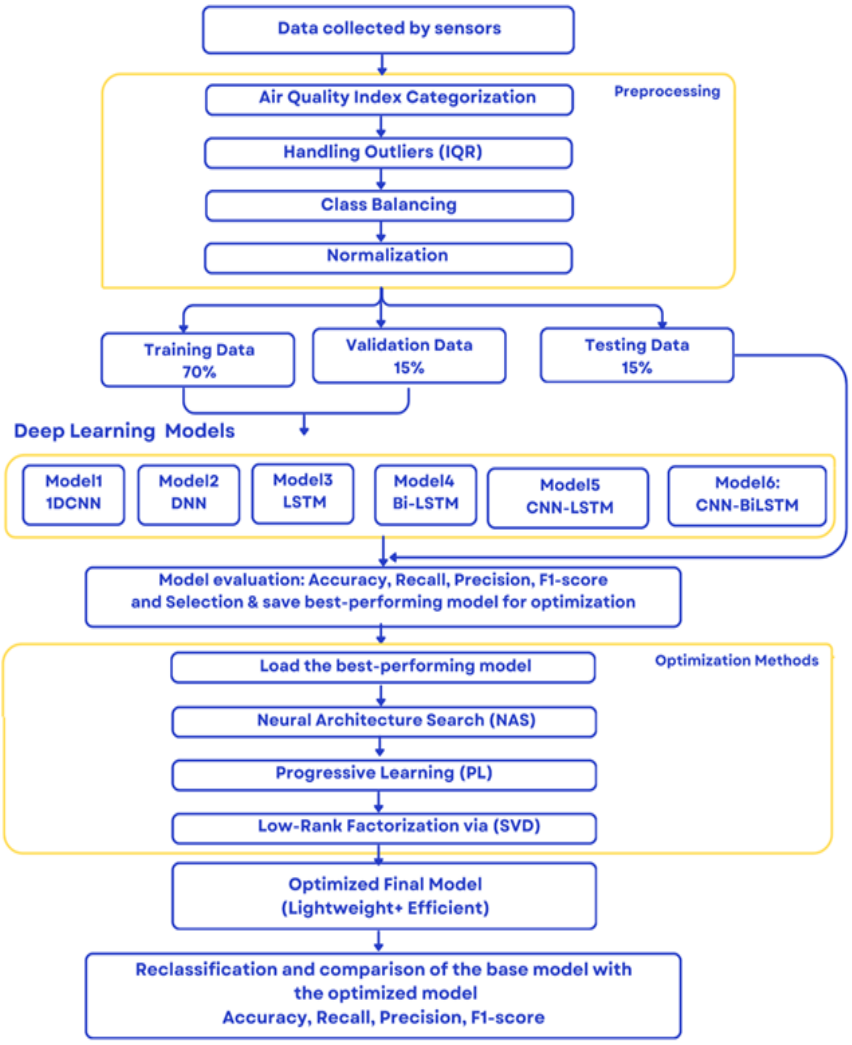


Figure 1: Workflow of the proposed SBS detection system.

2.2 Research Gap and Problem Statement

Despite growing interest in intelligent IAQ prediction, significant gaps persist:

- Most studies focus on forecasting pollutant concentrations rather than classifying SBS conditions in hospitals.
- Prior work rarely addresses model optimization for real-time inference on low-power devices.
- Research in the Middle East, particularly Iraq, remains scarce despite the region's growing IAQ challenges.

In Iraq, rapid construction growth, poor ventilation, and outdated infrastructure have heightened SBS risks, especially in hospitals. Existing IAQ data are limited, inconsistent, and often exceed WHO standards [14]. To date, no study has introduced a structured IAQ or SBS dataset from Iraq or developed optimized DL models suitable for real-time hospital deployment [15].

This research therefore fills a critical gap by providing the first structured IAQ dataset from Iraq and proposing an optimized, lightweight hybrid deep learning framework for real-time SBS detection in healthcare environments.

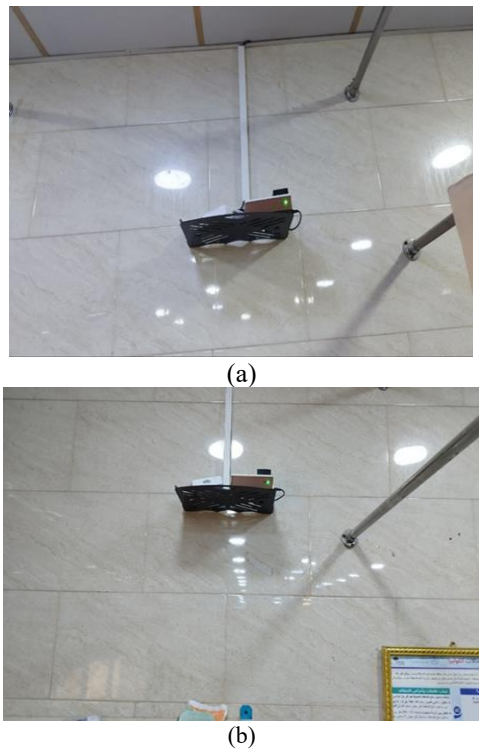
3 PROPOSED METHODOLOGIES

3.1 Workflow Overview

The proposed framework follows a five-stage pipeline designed to transform raw sensor data into actionable SBS alerts. As illustrated in Figure 1, the workflow begins with continuous data collection from a custom IoT device deployed in hospital wards. The raw data undergo rigorous pre-processing to ensure quality and balance. Next, multiple DL models are trained and benchmarked, with the best-performing architecture selected for further refinement. A multi-stage optimization process is then applied to reduce computational overhead without compromising accuracy. Finally, the optimized model is integrated into a real-time classification system capable of generating health advisories based on IAQ levels.

3.2 Dataset

The dataset was collected from the Emergency Department (male and female wards) of Baqubah Teaching Hospital between May 2024 and January 2025, during a period of high patient occupancy. A custom IoT-based monitoring system was deployed, integrating multiple calibrated sensors to continuously measure ten indoor environmental parameters, see Figure 2.



(a)

(b)

Figure 2: Deployment of the IoT-based monitoring system in hospital wards: a) device placement in the men's ward; b) device placement in the women's ward.

Specifically, the system incorporated the following sensors: DSM501A for particulate matter ($PM_{2.5}$ and PM_{10}), MQ-7 for carbon monoxide (CO), CCS811 and an infrared CO_2 sensor for carbon dioxide (CO_2) and total volatile organic compounds (TVOC), DHT22 for temperature and relative humidity, MQ131 for ozone (O_3), MQ135 for overall air quality (AQ), and a light-dependent resistor (LDR) for light intensity. The Arduino Mega, built on the ATmega2560 microcontroller with 86 programmable I/O pins and 256 KB of flash memory, was employed to manage data acquisition and processing. This configuration ensured efficient integration of multiple sensors and reliable real-time IAQ monitoring, see Figure 3.

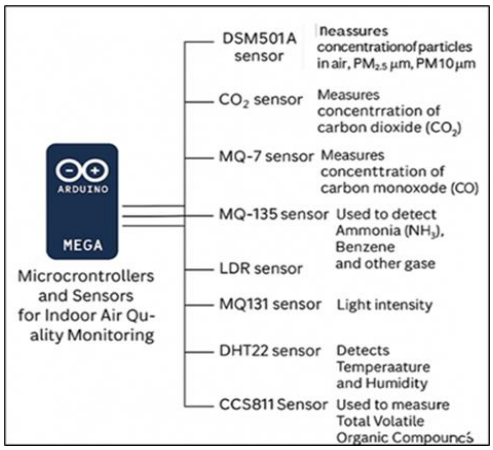


Figure 3: Sensor interfacing diagram with the MEGA 2560 PRO microcontroller [16].

The Air Quality Index (AQI) was calculated based on pollutant concentrations according to EPA standards. AQI values range from 0 to 500 and are divided into six categories: Good (0-50), Moderate (51-100), Unhealthy for Sensitive Groups (101-150), Unhealthy (151-200), Very Unhealthy (201-300), and Hazardous (301-500), as illustrated in Figure 4. The system generated a total of 523,524 time-stamped records at a frequency of one sample per minute. However, the dataset exhibited severe class imbalance, with the Moderate class dominating and both Good and Hazardous classes being significantly underrepresented.

Air Quality Index Levels of Health Concern	Numerical Value	Meaning
Good	0 to 50	Air quality is considered satisfactory, and air pollution poses little or no risk
Moderate	51 to 100	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.
Unhealthy for Sensitive Groups	101 to 150	Members of sensitive groups may experience health effects. The general public is not likely to be affected.
Unhealthy	151 to 200	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects.
Very Unhealthy	201 to 300	Health warnings of emergency conditions. The entire population is more likely to be affected.
Hazardous	301 to 500	Health alert: everyone may experience more serious health effects

Figure 4: AQI levels [17].

To address this issue, a balancing strategy was applied using oversampling of minority classes, resulting in a dataset expansion to 1,014,440 samples, with 202,888 records per class. This ensured equitable representation across categories while preserving data authenticity.

3.3 Data Pre-Processing

To ensure data quality and model reliability, a comprehensive pre-processing pipeline was applied, as detailed below:

- Missing Value Handling. No missing values were detected in the raw dataset due to the robustness of the IoT acquisition system and the deployment of an uninterruptible power supply (UPS). Thus, imputation was not required.
- Outlier Detection and Treatment. Outliers, which can distort data distribution and reduce DL model performance, were detected using the Interquartile Range (IQR) method across environmental variables such as CO_2, TVOC, PM10, PM2.5, and O_3. Values outside the range $[Q_1 - 1.5\times IQR, Q_3 + 1.5\times IQR]$ were considered outliers and replaced with the median of the respective feature. As shown in the accompanying boxplot Figure 5, this approach preserves temporal continuity and statistical integrity while minimizing the influence of extreme values, resulting in a more stable data distribution that enhances model accuracy and reliability.
- Class Balancing. The original AQ dataset was highly imbalanced, with the “moderate” class dominating (202,888 samples), while critical categories such as “good” and “hazardous” were severely underrepresented. Such skewed distributions bias DL models toward majority classes, reducing performance on rare but important conditions. To address this, a random duplication-based oversampling strategy was applied, replicating minority class samples until all five categories contained 202,888 records, as shown in Figure 6. This process expanded the dataset from 523,524 to 1,014,440 samples, ensuring equal class representation. While this approach preserved data authenticity and improved model convergence and classification accuracy, it relied solely on duplication rather than synthetic generation, which is an acknowledged limitation.

- Normalization. Data normalization is a crucial pre-processing step in preparing datasets for ML/DL models and AI applications. It transforms input features into a common scale, typically between 0 and 1, which enhances model convergence and accuracy [18].

$$Norm(x) = \frac{x - x_{min}}{x_{max} - x_{min}}.$$

This ensured uniform feature scales, accelerating model convergence and improving gradient stability.

- Data Splitting. Using stratified sampling, the dataset was divided into training (70%, 710,108 samples), validation (15%, 152,166 samples), and testing (15%, 152,166 samples). This ensured proportional class representation across subsets.

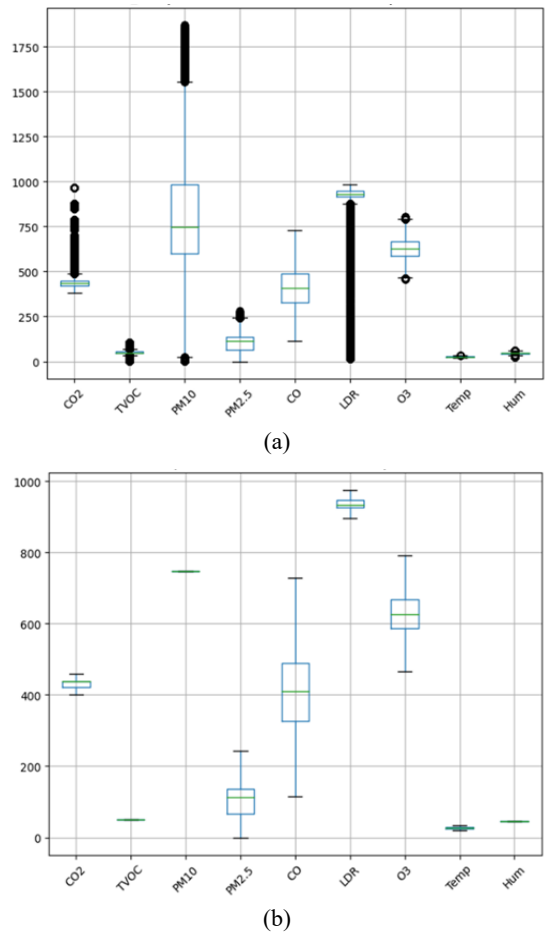


Figure 5: Boxplot of features a) before, b) after outlier replacement.

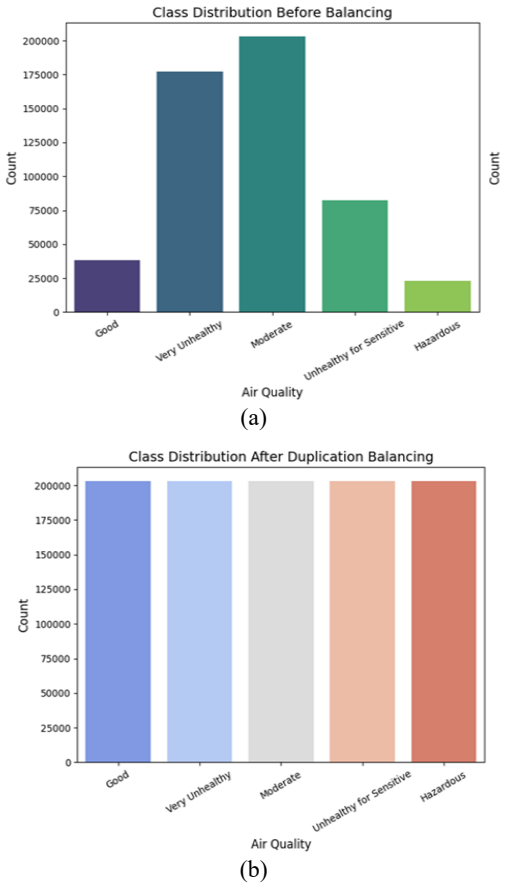


Figure 6: Effect of random duplication on dataset class balancing: a) imbalanced class distribution before oversampling; b) balanced class distribution after oversampling.

The pre-processing pipeline was validated through diagnostic visualizations (e.g., boxplots before and after outlier handling, and histograms pre/post class balancing), confirming high-quality input for subsequent model training and fair evaluation.

3.4 Baseline Model Architecture

The baseline model, compared against five other architectures (DNN, 1D-CNN, LSTM, BiLSTM, and 1D-CNN-LSTM), adopted a hybrid 1D-CNN-BiLSTM architecture to integrate feature extraction and temporal modelling. In this design, Conv1D and MaxPooling layers were employed to capture local spatial patterns, while two stacked BiLSTM layers

modelled sequential and bidirectional temporal dependencies. To stabilize learning and enhance generalization, Batch Normalization was incorporated, followed by fully connected layers with ReLU activation and a final Softmax layer for classification [19], [20].

This configuration achieved a notable accuracy of 94.8%, outperforming all the other baseline models in capturing both spatial and temporal dependencies. However, the model incurred a high inference time of 20.8 seconds, rendering it computationally expensive and unsuitable for real-time applications.

3.5 Optimization Techniques

To enhance model efficiency and enable deployment on edge devices with constrained memory and processing power, three complementary optimization strategies were employed. First, NAS is an automated DL methodology designed to discover high-performing neural network architectures without manual trial-and-error. Instead of manually defining layer types, sizes, or the number of units, NAS treats network design as a search problem, optimizing both architecture and hyperparameters to maximize model performance. In this project, NAS was applied to tune the number of dense layer units in the 1D-CNN-BiLSTM model, adjusting units in a structured manner (±16 based on layer index), resulting in improved convergence, reduced overfitting, and enhanced overall performance.

NAS frameworks typically consist of three main components:

- Search Space. Defines all possible architectures, including layer types, connectivity, activation functions, and kernel sizes.
- Search Strategy. Guides exploration of the search space, often using reinforcement learning, evolutionary algorithms, or random search.
- Performance Estimation. Evaluates candidate architectures efficiently, usually on a validation subset, reducing computational costs.

Mathematically, NAS can be formulated as an optimization problem:

$$A^* = arg^{max}_{A \in A} Performance(A, \theta).$$

Where A∗ is the optimal architecture, A is the set of candidate architectures, θ are the learnable parameters, and Performance (A,θ) is the performance metric [21].

Second, PL is a curriculum-based training strategy in deep learning that gradually increases the complexity of tasks during training. Instead of presenting the model with the full complexity of the data from the start, PL begins with simpler tasks and progressively introduces more challenging patterns. This phased approach helps stabilize training, improve convergence, and reduce the risk of overfitting or catastrophic forgetting, particularly when fine-tuning pre-trained models.

In this study, PL was applied by gradually unfreezing and fine-tuning the dense layers of the 1D-CNN-BiLSTM model. Early layers were initially frozen, allowing the model to build foundational representations, while deeper layers were progressively adapted. This structured learning process enhanced generalization and promoted smoother gradient propagation throughout the network.

Mathematically, the PL loss at training step t can be expressed as [22]:

$$L(t)=Lsimple+\lambda t \cdot Lcomplex.$$

Where Lsimple is the loss for simpler tasks, Lcomplex is the loss for complex tasks, and λt is a time-dependent weighting factor that increases over training. This formulation reflects the adaptive emphasis on more complex tasks as the model becomes proficient, leading to improved performance on challenging or previously unseen data.

Finally, SVD is a matrix compression technique widely used in deep learning to reduce model size and computational requirements while preserving critical information. In this study, LRF was applied to the dense weight matrices of the 1D-CNN-BiLSTM model using Singular Value Decomposition as weight matrix W∈ $\mathbb{E}$m×n, which can be decomposed using SVD [23], [24]:

$$W \approx U\Sigma V^T.$$

This approach significantly reduced the number of parameters, training time, and inference latency, making the model more suitable for real-time deployment in resource-constrained environments. Truncated reconstruction also acts as an implicit regularizer, improving generalization and minimizing overfitting. The effectiveness of LRF was validated through confusion matrices, classification reports, and learning curves, confirming that model performance remained stable despite compression.

Mathematically, the quality of the low-rank approximation is assessed using the Frobenius norm:

$$||A\text{-}AK||F = \sqrt{\sum_{i=k+1}^{min\,(m,n)} \sigma_i^2}.$$

Where A denotes the original matrix, AK denotes rank-k approximation, and σi denotes singular values beyond the top-k components. A smaller reconstruction error implies a more accurate approximation with minimal information loss.

This optimization pipeline yielded a lightweight, fast, and accurate model, striking a balance between predictive performance and computational feasibility, a critical requirement for real-time applications.

4 EXPERIMENTAL RESULTS

To establish a reliable benchmark, six deep learning (DL) models were evaluated under identical training conditions using the IAQ Baquba Hospital dataset. Stratified splits were applied to ensure proportional class representation, and all models were trained with the same optimizer, batch size, and epochs. The tested models included DNN, 1D-CNN, LSTM, BiLSTM, 1D-CNN-LSTM, and the baseline hybrid 1D-CNN-BiLSTM. Performance was assessed using accuracy, F1-score, inference time, and parameter count.

As summarized in Table 2, the 1D-CNN-BiLSTM achieved the highest accuracy (94.8%), confirming the benefit of combining convolutional feature extraction with bidirectional temporal modelling. However, this came at the cost of a long inference time (20.8 s) and a relatively large number of parameters, which limited its applicability in real-time healthcare environments. In contrast, the standalone 1D-CNN achieved the lowest accuracy (89.2%), reflecting its inability to fully capture temporal dependencies. LSTM and BiLSTM models demonstrated competitive performance (~92%) but lacked the local spatial extraction capabilities of CNN layers.

The evaluation results are summarized in Figure 7, starting with the performance of the 1D-CNN-BiLSTM model. The confusion matrices and classification reports confirm its robustness, correctly classifying over 98% of cases in critical categories such as Good and Hazardous, with only minor confusion between adjacent classes (e.g., Very Unhealthy vs. Hazardous). The classification metrics further highlight balanced precision, recall, and F1-scores across all five classes, resulting in 94% overall accuracy.

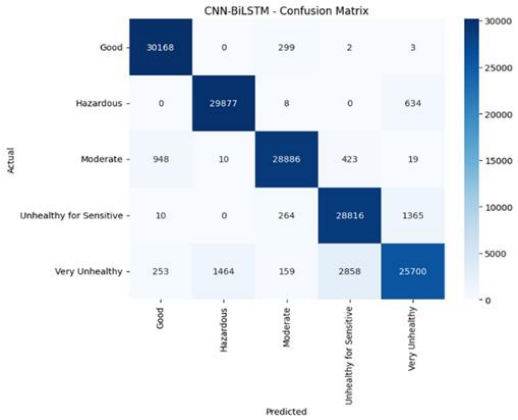
In Figure 8, the training and validation curves of the 1D-CNN-BiLSTM model further support these outcomes, with accuracy steadily increasing to 94.27% while loss values declined consistently. This indicates stable convergence and strong learning effectiveness without signs of overfitting, confirming the ability of the hybrid model to jointly capture spatial and temporal dependencies.

In comparison, the 1D-CNN baseline achieved weaker results. The confusion matrices and classification reports show an overall accuracy of 89%, with noticeable performance drops in the Very Unhealthy category, where recall was significantly lower. While the model demonstrated reasonable classification in other categories, its lack of temporal modelling limited precision and recall balance, as shown in Figure 9.

Similarly, in Figure 10, the training and validation curves of the 1D-CNN model exhibited stable learning but reached only 89.16% validation accuracy, with higher loss values compared to the hybrid model. These findings reflect its reduced capacity to generalize, underscoring the advantages of incorporating BiLSTM layers for improved SBS detection.

Table 2: Performance comparison of baseline DL models for IAQ classification.

Model	Accuracy	Inference Time(s)	Training Time(s)	Trainable Parameters
1DCNN-BiLSTM	0.943	20.80	2388.49	445445
LSTM	0.926	8.69	801.076	75141
BiLSTM	0.923	10.91	710.259	149893
1DCNN-LSTM	0.916	16.1457	1494.8	183749
DNN	0.908	7.481	404.336	11781
1DCNN	0.892	7.705	420.896	25477

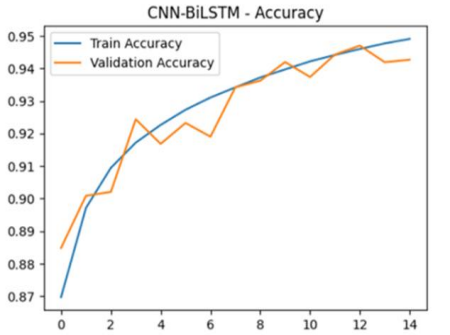


(a)

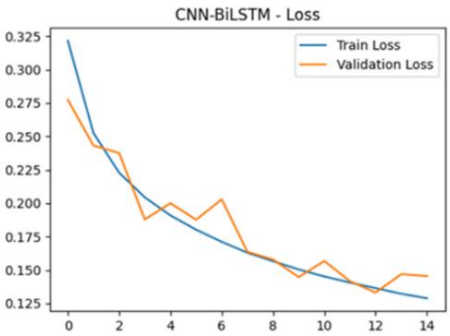
	precision	recall	f1-score	support
Good	0.96	0.99	0.98	30472
Hazardous	0.95	0.98	0.97	30519
Moderate	0.98	0.95	0.96	30286
Unhealthy for Sensitive	0.90	0.95	0.92	30455
Very Unhealthy	0.93	0.84	0.88	30434
accuracy			0.94	152166
macro avg	0.94	0.94	0.94	152166
weighted avg	0.94	0.94	0.94	152166

(b)

Figure 7: Performance evaluation of the 1D-CNN-BiLSTM model: a) confusion matrix; b) classification report.

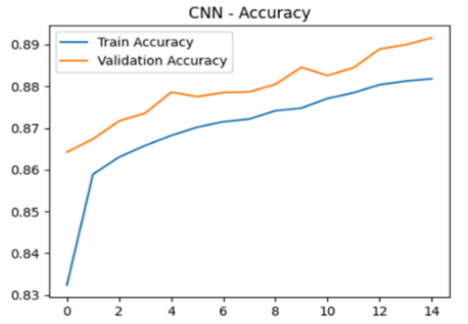


(a)

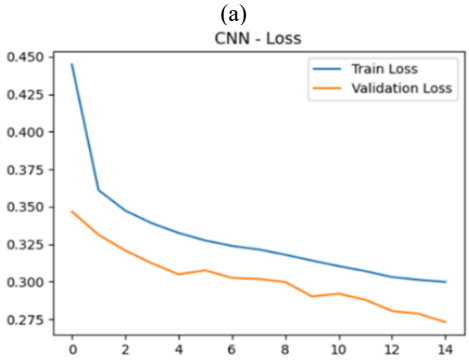


(b)

Figure 8: Training performance of the 1D-CNN-BiLSTM model: a) training accuracy curve; b) validation accuracy and loss curves.

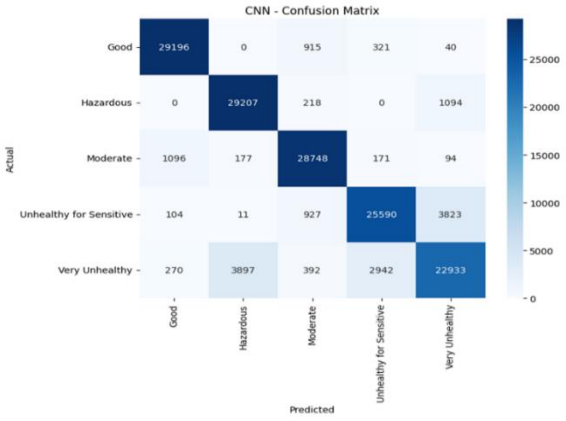


(a)



(b)

Figure 9: Training performance of the 1D-CNN model: a) training accuracy curve; b) validation accuracy and loss curves.



(a)

	precision	recall	f1-score	support
Good	0.95	0.96	0.96	30472
Hazardous	0.88	0.96	0.92	30519
Moderate	0.92	0.95	0.94	30286
Unhealthy for Sensitive	0.88	0.84	0.86	30455
Very Unhealthy	0.82	0.75	0.79	30434
accuracy			0.89	152166
macro avg	0.89	0.89	0.89	152166
weighted avg	0.89	0.89	0.89	152166

(b)

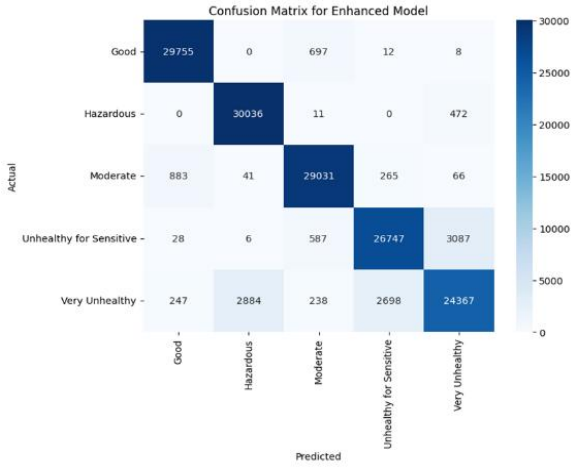
Figure 10: Performance evaluation of the 1D-CNN model: a) confusion matrix; b) classification report.

4.2 Optimized Model Performance

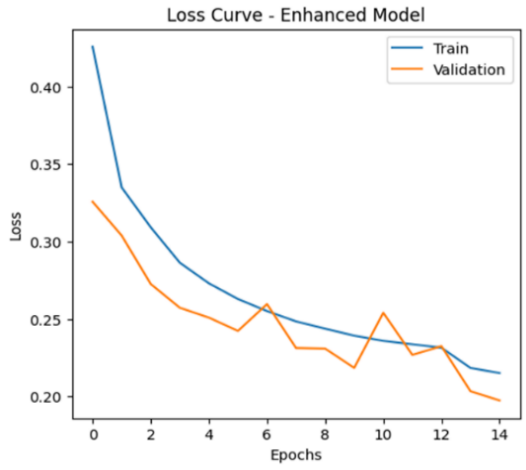
Although the baseline 1D-CNN-BiLSTM achieved superior accuracy, its inference time of 20.8 seconds rendered it unsuitable for real-time hospital deployment. To overcome this limitation, three optimization techniques, NAS, PL, and SVD, were applied.

As shown in Table 3, the optimized hybrid model achieved a strong trade-off:

- Accuracy: 92.0% (only a 2.8% drop).
- Inference time: 0.66 seconds (≈31× faster).
- Model size: 1.72 MB (≈67% smaller).
- Trainable parameters: reduced slightly while preserving representational capacity.



(a)

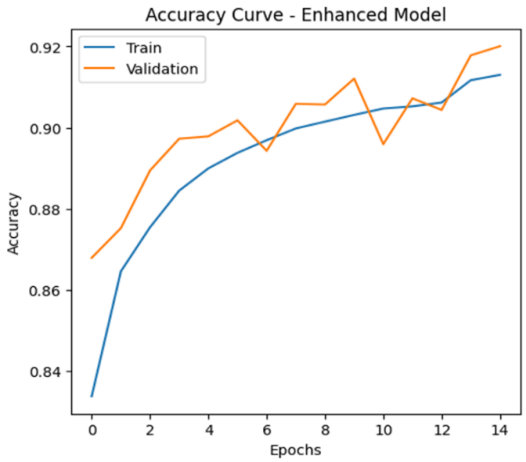


(b)

Figure 11: Performance evaluation of the enhanced model: a) confusion matrix; b) classification report.

Table 3: Baseline vs optimized 1D-CNN-BiLSTM performance.

Model	Accuracy (%)	Inference Time (s)	Model Size (MB)
Original 1D-CNN-BiLSTM	94.8	20.8	5.17
Optimized (NAS+PL +SVD)	92.0	0.66	1.72



(a)

	precision	recall	f1-score	support
Good	0.96	0.98	0.97	30472
Hazardous	0.91	0.98	0.95	30519
Moderate	0.95	0.96	0.95	30286
Unhealthy for Sensitive	0.90	0.88	0.89	30455
Very Unhealthy	0.87	0.80	0.83	30434
accuracy			0.92	152166
macro avg	0.92	0.92	0.92	152166
weighted avg	0.92	0.92	0.92	152166

(b)

Figure 12. Training performance of the enhanced model: a) training accuracy curve; b) validation accuracy and loss curves.

The confusion matrix and classification report in Figure 11 demonstrate the enhanced model's strong and balanced classification capability. Most predictions were concentrated along the diagonal, with 29,755 samples correctly identified as Good and 30,036 as Hazardous. The model achieved an overall accuracy of 92%, with macro- and weighted-average precision, recall, and F1-scores all equal to 0.92. Performance was particularly strong in the Good and Moderate classes (F1 = 0.97 and 0.95), while the Hazardous class also achieved high accuracy (precision = 0.91, recall = 0.98, F1 = 0.95). However, performance was slightly weaker in more challenging

categories, such as Unhealthy for Sensitive (F1 = 0.89) and Very Unhealthy (F1 = 0.83), mainly due to confusion with adjacent categories.

The training and validation curves further confirm the robustness of the enhanced model, as illustrated in Figure 12. Accuracy steadily improved, with training rising from 83% to over 91% and validation accuracy peaking above 92%. Similarly, loss values declined consistently, with training loss decreasing from 0.42 to 0.21 and validation loss falling below 0.20. The close alignment between training and validation trends indicates strong generalization and minimal overfitting. These results highlight the effectiveness of the optimization strategies (NAS, PL, and SVD) in ensuring a stable learning process and enhancing the efficiency of the 1D-CNN-BiLSTM model.

These results underscore that the optimization pipeline successfully produced a lightweight, fast, and accurate model, bridging the gap between high predictive performance and real-time deployability in resource-constrained healthcare environments.

5 DISCUSSION

The findings confirm that hybrid deep learning architectures particularly the 1D-CNN-BiLSTM-offer superior capability for modeling complex multivariate IAQ time-series data in the context of Sick Building Syndrome (SBS) detection. The model effectively combines convolutional layers for identifying short-term pollutant variations with BiLSTM layers that capture long-term temporal dependencies.

Although the baseline hybrid achieved strong predictive accuracy, its long inference time limited real-time applicability. To address this, a multi-level optimization strategy was implemented, integrating techniques such as Neural Architecture Search, Progressive Learning, and Low-Rank Factorization. These optimizations significantly improved computational efficiency while maintaining high predictive performance, demonstrating the framework's suitability for continuous IAQ monitoring in healthcare environments.

From an applied perspective, the optimized framework can be embedded into hospital Internet-of-Things (IoT) ecosystems for automated ventilation control and real-time air quality alerts. This integration aligns with public health strategies focused on reducing exposure to indoor pollutants and preventing SBS-related symptoms among medical staff and patients.

This research provides three key contributions:

- A Novel Dataset. The release of the first structured IAQ dataset from Iraq, collected at Baqubah Teaching Hospital and publicly released on Kaggle, addressing a geographical research gap and enabling SBS research in underrepresented regions.
- Architectural Innovation. A hybrid design that effectively integrates local feature extraction with long-range temporal modelling, consistently outperforming unimodal baselines.
- Practical Deployability. A lightweight, optimized model capable of delivering sub-second inference and clinically reliable alerts on low-cost devices such as Raspberry Pi.

Despite these achievements, two limitations remain. First, the dataset originates from a single healthcare facility, which may limit generalizability across different building types or climates. Broader multi-site validation is required to confirm robustness. Second, while random oversampling effectively mitigated class imbalance, it does not fully enhance representation of rare SBS conditions such as hazardous air states. Future research should investigate advanced balancing methods (e.g., SMOTE, ADASYN) and adaptive generative techniques to improve generalization.

Future work will also expand data collection across multiple hospitals and explore privacy-preserving deployment approaches such as Federated Learning to support scalable and ethical implementation.

In conclusion, this study introduces the first deployable AI framework for SBS detection in Iraqi hospitals, bridging the gap between predictive accuracy and computational efficiency. The proposed system provides a practical foundation for intelligent, real-time IAQ monitoring and offers a scalable model for use in low- resource healthcare settings.

6 CONCLUSIONS

This study presented an enhanced hybrid 1D-CNN-BiLSTM framework designed for real-time SBS using IAQ data collected from Baqubah Teaching Hospital in Iraq. The baseline model achieved an accuracy of 94.8%, whereas the optimized version maintained a competitive accuracy of 92% while delivering major computational improvements, including a 96.8% reduction in inference time and a 66.7% decrease in model size. These advancements significantly strengthen the model's suitability for

deployment on edge-based IoT devices within resource-constrained healthcare environments, where rapid decision-making and efficient energy consumption are critical.

Beyond the algorithmic optimization, this research contributes to addressing a notable gap in IAQ-related studies within the Middle East, where high-resolution datasets and AI-driven SBS detection frameworks remain limited. The proposed system provides a scalable and practical solution for environmental health monitoring and represents a step toward the development of intelligent, data-driven IAQ management infrastructures aimed at fostering healthier indoor conditions in hospitals and other clinical facilities.

Looking ahead, future extensions of this research will focus on enhancing the generalizability and robustness of the model by expanding data collection across multiple hospitals, geographical locations, and varying building layouts. More advanced oversampling and class-balancing strategies-such as SMOTE, ADASYN, or generative augmentation-will be investigated to improve performance under imbalanced IAQ conditions. Additionally, integrating lightweight neural architectures (e.g., MobileNet, Efficient Net-Lite) and privacy-preserving paradigms such as Federated Learning will further support multi-site training without compromising patient or facility confidentiality.

Moreover, embedding the system directly into hospital HVAC networks and cloud-edge IoT platforms would enable proactive, continuous, and automated IAQ regulation. Finally, transitioning from pure classification to long-term IAQ forecasting and predictive analytics will facilitate early intervention strategies for mitigating SBS risks, thereby strengthening preventive healthcare and operational resilience in medical environments.

REFERENCES

[1] S. Subramaniam et al., "Artificial intelligence technologies for forecasting air pollution and human health: A narrative review," Sustainability, vol. 14, no. 9951, pp. 1-10, 2022, [Online]. Available: https://doi.org/10.3390/su14169951.

[2] World Health Organization, "Household air pollution and health," WHO Fact Sheet, 2024, [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health.

[3] İ. Arikan, Ö. F. Tekin, and O. Erbaş, "Relationship between sick building syndrome and indoor air quality among hospital staff," Medicina del Lavoro, vol. 109, no. 6, pp. 435-443, 2018.

[4] T. D. Piyadasa and K. Gunawardana, "A review on oversampling techniques for solving the data imbalance problem in classification," Int. J. Adv. ICT Emerg. Reg., vol. 16, no. 1, pp. 1-10, 2023.

[5] H. Q. Flayyih, J. Waleed, and A. M. Ibrahim, "IoT-based AI methods for indoor air quality monitoring systems: A systematic review," Int. J. Comput. Digit. Syst., vol. 16, no. 1, pp. 813-826, 2024.

[6] H. Q. Flayyih, J. Waleed, and A. M. Ibrahim, "Indoor air quality prediction in sick building using machine and deep learning: Comparative analysis," Diyala J. Eng. Sci., vol. 18, no. 1, pp. 203-218, 2025.

[7] X. Shi, W. Lu, Y. Zhao, and P. Qin, "Prediction of indoor temperature and relative humidity based on cloud database using an improved BP neural network in Chongqing," IEEE Access, vol. 6, pp. 30559-30569, 2018, [Online]. Available: https://doi.org/10.1109/ACCESS.2018.2842962.

[8] X. Zhao and R. Zhang, "A deep recurrent neural network for air quality classification," J. Inf. Hiding Multimedia Signal Process., vol. 9, no. 2, pp. 346-351, 2018.

[9] F. Elmaz, R. Eyckerman, W. Casteels, S. Latré, and P. Hellinckx, "CNN-LSTM architecture for predictive indoor temperature modeling," Build. Environ., vol. 206, art. 108327, pp. 1-10, 2021, [Online]. Available: https://doi.org/10.1016/j.buildenv.2021.108327.

[10] R. Bao, W. Jiang, and Y. Zhou, "FL-CNN-LSTM: Indoor air quality prediction using fuzzy logic and CNN-LSTM model," Environ. Eng. Manag. J., vol. 22, no. 9, pp. 1873-1892, 2023.

[11] N. Fernandes and J. Gonçalves, "Multivariate and multi-output indoor air quality prediction using bidirectional LSTM," J. Build. Eng., vol. 18, pp. 123-134, 2023, [Online]. Available: https://doi.org/10.1016/j.jobe.2022.106676.

[12] X. Zhu, F. Zou, and S. Li, "Enhancing air quality prediction with an adaptive PSO-optimized CNN-BiLSTM model," Appl. Sci., vol. 14, no. 13, p. 5787, 2024, [Online]. Available: https://doi.org/10.3390/app14135787.

[13] Z. Wang, Y. Yang, and F. Wu, "A lightweight air quality monitoring method based on multiscale dilated convolutional neural network," IEEE Trans. Ind. Informat., vol. 20, no. 1, pp. 1-10, 2024, [Online]. Available: https://doi.org/10.1109/TII.2023.3325632.

[14] I. J. Al Rikabi, M. A. Abuelnour, T. K. Abed, A. H. M. J. Al Obaidy, and A. A. M. Omara, "Investigation of indoor air quality in several offices in Baghdad: Case study," in Proc. 8th Int. Eng. Conf. Adv. Comput. Civil Eng. Towards Eng. Innov. Sustain. (IEC-2022), Erbil, Iraq, Jul. 2022, pp. 78-82.

[15] H. Q. Flayyih, J. Waleed, and A. M. Ibrahim, "Estimating indoor air quality in sick buildings using machine learning and deep learning: A comparative analysis," Int. J. Intell. Eng. Syst., vol. 17, no. 5, pp. 198-211, 2024, [Online]. Available: https://doi.org/10.22266/ijies2024.1031.18.

[16] H.-I. Liu et al., "Lightweight deep learning for resource-constrained environments: A survey," ACM Comput. Surv., vol. 1, no. 1, pp. 1-40, 2022, [Online]. Available: https://doi.org/10.1145/3514223.

[17] H. P. L. de Medeiros and G. Girão, "An IoT-based air quality monitoring platform," in Proc. IEEE Int. Smart Cities Conf. (ISC2), Piscataway, NJ, USA, 2020, pp. 1-6, [Online]. Available: https://doi.org/10.1109/ISC2.2020.9239021.

[18] S. S. Alaoui, B. Aksasse, and Y. Farhaoui, "Air pollution prediction through Internet of Things technology and big data analytics," Int. J. Comput. Intell. Stud., vol. 8, no. 2-3, pp. 177-191, 2019.

[19] H. Nizam-Ozogur and Z. Orman, "A heuristic-based hybrid sampling method using a combination of SMOTE and ENN for imbalanced health data," Expert Syst., vol. 41, no. 1, art. e13596, 2024, [Online]. Available: https://doi.org/10.1111/exsy.13596.

[20] J. Waleed, S. Albawi, H. Q. Flayyih, and A. Alkhayyat, "An effective and accurate CNN model for detecting tomato leaf diseases," in Proc. 4th Int. Iraqi Conf. Eng. Technol. Their Appl. (IICETA-2021), Al-Najaf, Iraq, 2021, pp. 33-40, [Online]. Available: https://doi.org/10.1109/IICETA51927.2021.9502288.

[21] P. Ren et al., "A comprehensive survey of neural architecture search: Challenges and solutions," ACM Comput. Surv., vol. 54, no. 4, art. 76, 2021, [Online]. Available: https://doi.org/10.1145/3447588.

[22] H. M. Fayek, L. Cavedon, and H. R. Wu, "Progressive learning: A deep learning framework for continual learning," Neural Netw., vol. 128, pp. 345-357, 2020, [Online]. Available: https://doi.org/10.1016/j.neunet.2020.05.001.

[23] L. Wu and Q. Lin, "Improved Parker's method for topographic models using Chebyshev series and low-rank approximation," Geophys. J. Int., vol. 209, no. 2, pp. 1296-1325, 2017, [Online]. Available: https://doi.org/10.1093/gji/ggx064.

[24] S. S. Mohammed and J. M. Al-Tuwaijari, "Skin disease classification system based on metaheuristic algorithms," AIP Conf. Proc., vol. 2475, no. 1, art. 070008, pp. 1-8, Mar. 2023, [Online]. Available: https://doi.org/10.1063/5.0102907.