International Journal of Safety and Security Engineering
Vol. 16, No. 7, July, 2026, pp. 1423-1438
Journal homepage: http://iieta.org/journals/ijsse

Evaluating Deep Belief Networks on IoMT Dataset: The Impact of Different Network Designs on Attack Detection Accuracy

Marwa Taresh Obeid, Enas Mohammad Hussein*

Computer Science Department, College of Education, Mustansiriyah University, Baghdad 10052, Iraq

Corresponding Author Email: marwath1987@uomustansiriyah.edu.iq

https://doi.org/10.18280/ijsse.160702

Received: 19 April 2026
Revised: 22 June 2026
Accepted: 29 June 2026
Available online: 31 July 2026

Keywords:
Deep Belief Network, intrusion detection system, Internet of Things healthcare, Internet of Medical Things security, machine learning, network security, regularisation, cybersecurity

ABSTRACT

The rapid growth of the Internet of Medical Things (IoMT) brings significant improvements in healthcare services but also introduces serious security challenges, as data is continuously transmitted across networks and devices operate under constrained resources. This study examines Deep Belief Network (DBN) architectures for intrusion detection in Internet of Things (IoT) healthcare systems, focusing on how architectural design choices affect detection performance. A structured experimental framework was designed, evaluating six DBN configurations that systematically vary depth, width, regularisation, and activation function. Two publicly available benchmark datasets were used: Canadian Institute for Cybersecurity- Internet of Medical Things (CIC IoMT 2024), a high-dimensional 18-class multi-protocol dataset, and Edith Cowan University-Internet of Health Things (ECU-IoHT), a compact 5-class, 10-feature dataset. Results show that the Regularised DBN incorporating Batch Normalisation and Dropout achieves the best overall performance, reaching 97.72 ± 0.04% accuracy and 0.994 ± 0.002 (2) Area Under the ROC Curve (AUC) (mean ± std, 5 independent runs) on CIC IoMT 2024, and perfect classification (1.000 ± 0.000) on the ECU-IoHT dataset. In contrast, deeper models exhibit unstable behaviour and inferior performance, particularly on minority classes, while wider models yield only marginal improvements on complex tasks. Despite strong overall results, three attack classes (Distributed Denial of Service (DDoS), Publish Flood, OS Scan, and Recon VulScan) remain poorly detected (F1 < 0.25) across all configurations, indicating persistent feature-space overlap and class imbalance challenges. The findings highlight that balanced architectural design with appropriate regularisation is critical for robust and reliable intrusion detection in IoMT environments.

1. INTRODUCTION

The Internet of Things (IoT) is a fast-changing field allowing communication between electronic sensors and devices that are connected to the internet. One important use of this technology is Internet of Medical Things (IoMT), where a connection is made between healthcare services and patients by using many sensors, like electrocardiography devices, heart rate monitors, and blood pressure tools [1].

These developments have changed patient care in a significant way with remote diagnosis and health monitoring, which will continue for a long time. However, the wide connection and constant data exchange inside IoMT systems have created serious cybersecurity problems [2, 3]. Sensitive medical data becomes the main target for attackers, which can cause dangers with data change, man-in-the-middle attacks and denial of service attacks [4, 5]

Many medical devices have limited resources and sufficient computational power and energy are not available; also, traditional security methods like complex encryption or strong firewalls cannot be supported easily. Because of this situation, an urgent need is seen for advanced intrusion detection systems (IDS) that are specially designed for the healthcare environment [6, 7].

Artificial intelligence (AI) has an important role in improving security by detecting unusual behavior. Traditional machine learning (ML) methods have been used before in IDS research, but now the network traffic in IoT has become larger and more complex, so deeper models are needed. Deep learning (DL) methods, such as deep neural networks (DNN) and recurrent neural networks (RNN), are considered better solutions because they can learn by themselves and also understand complex patterns from high-dimensional data [6, 8, 9].

Among many DL methods, the Deep Belief Network (DBN) is seen as an important way for finding network attacks. A DBN is made from many layers of random hidden variables, and it is often arranged like a stack of Restricted Boltzmann Machines (RBMs). These structures are built to learn mixed and hierarchical data forms, and they are very useful for feature learning and also for reducing dimensionality in network intrusion detection situations [1, 6].

The performance of a DBN-based IDS is strongly affected by its design, including how many hidden layers exist and how many units are inside each layer. In present studies, three hidden layers are often used by default; however, there is still

no clear agreement about the best number of neurons for getting high detection results [6, 10, 11].

This study aims to examine how different DBN architectures affect attack detection accuracy using publicly available benchmark datasets, namely CICIoMT2024 and Edith Cowan University-Internet of Health Things (ECU-IoHT), by investigating how architectural variations in depth, width, regularisation, and activation function influence detection performance against threats such as spoofing, Denial of Service (DoS), and reconnaissance.

The growth of the IoMT changed healthcare in a strong way, with real-time monitoring and data-based insight, many cybersecurity weaknesses have also been created. Because healthcare data is very sensitive and many medical devices have limited resources, traditional security methods like sophisticated firewalls are often not possible to apply, and therefore intelligent IDS systems become necessary [4, 10].

1.1 Traditional classifiers in Internet of Medical Things Security

ML methods have been widely studied for detecting attacks in IoMT network traffic. Algorithms like Support Vector Machines (SVMs), Random Forest, and K-Nearest Neighbors (KNN) show noticeable performance in many research areas. Random Forest is often seen as a very reliable method for stable threat detection and sometimes reaching nearly perfect accuracy in certain cases [12, 13].

However, an important limitation in many studies exists, since models are tested in closed-set conditions, which do not include new or unknown attack types or even small changes in known attacks. Also, shallow ML models have difficulty when dealing with high-dimensional data, which is very common in modern IoMT systems; therefore, more advanced architectures are needed [8, 14].

1.2 The Evolution toward Deep Learning and Deep Belief Networks

DL is becoming a better solution for IoMT security because it can learn features automatically and can understand complex patterns in large datasets. Models like DNN and Long Short-Term Memory networks often perform better than traditional ML models for detecting different cyber threats. Among these methods, DBNs are very important, since they are generative models built from stacked Restricted RBMs, and they use layer-by-layer training, which allows more stable performance, even in unsupervised situations [15].

By using many hidden layers, DBNs are able to take high-level abstract ideas from raw network data, and this can help more correctly classify malicious actions. However, the setup of the DBN structure, like how many layers and neurons should be used, is still not clearly agreed upon, and it remains an active research problem with no universal answer [6, 16].

The performance of these DL models strongly depends on the quality and also the variety of benchmarking datasets that are used for testing. The CICIoMT2024 dataset shows important progress because it gives a multi-protocol environment where traffic from MQTT, BLE, and Wi-Fi is included, and this reflects the mixed and complex nature of modern IoMT systems. Similarly, the ECU-IoHT dataset gives a more specific framework for studying cyberattacks such as Address Resolution Protocol (ARP) spoofing and Smurf attacks inside IoHT systems. Even though such datasets are available, many existing studies are still limited to only binary classification and do not give the needed multiclass detail, which is important for better threat reduction in healthcare networks [1, 17, 18].

Several studies demonstrated the effectiveness of DBNs for intrusion detection in IoMT and related IoT environments, consistently reporting high detection accuracy despite differences in network architectures, deployment settings and evaluation datasets [15, 19-21]. Early research established the feasibility of distributed DBN architectures for smart city IoT environments and multi-layer DBNs for wireless sensor networks, highlighting their scalability and strong detection performance [19, 20]. Subsequent work extended models to healthcare IoMT environments where DBNs achieved robust detection across diverse attack categories including Botnet, DoS/DDoS and Infiltration attacks, while further improvements were obtained through metaheuristic feature selection and hyperparameter optimization [1, 15]. Ensemble DBN architectures have also been investigated for cloud-supported IoT and SCADA systems, demonstrating that combining multiple DBNs can enhance adaptability and malware detection performance in industrial environments [21].

Beyond standalone DBN models, recent research has increasingly explored hybrid deep learning architectures to address the growing complexity of IoMT. These approaches integrate complementary learning paradigms, such as hierarchical meta learning for zeroday attack detection, recurrent neural networks for sequential traffic analysis and CNNs for extracting spatial temporal traffic features achieving detection accuracies approaching or exceeding 99% across diverse IoMT datasets [2, 8, 22]. Collectively, these findings suggest that while DBNs remain a competitive solution for intrusion detection, hybrid deep learning models can further improve feature representation and generalization, particularly in complex and heterogeneous IoMT environments.

Despite these advances, a critical pattern emerges: each study adopts a single fixed DBN or DL architecture and compares it against alternative model families such as SVM or CNN or RNN without systematically varying the internal architecture of the DBN. Key design variables like network depth, layer width, regularisation strategy and activation function are treated as constants rather than experimental factors, making it difficult to determine whether reported performance improvements stem from the choice of DL. Furthermore, several DBN studies rely on older non-IoMT-specific benchmarks such as NSL-KDD [22] and CICIDS2017 [15], which lack the multi-protocol heterogeneity and fine-grained attack taxonomy of modern IoMT datasets like CICIoMT2024 [17, 20] and ECU-IoHT [23]. Binary evaluation settings [8] further obscure class level weaknesses that are critical in healthcare security contexts where even a single missed attack category can have serious clinical consequences. These gaps motivate the present study's systematic and ablation style comparison of six DBN configurations across two complementary IoMT datasets.

1.3 Research gaps and study rationale

Despite recent advances, several important gaps remain in DBN intrusion detection for IoMT environments. First, most studies employ a single DBN architecture and compare it with other models such as SVM, CNN, or RNN while rarely investigating the DBN design space itself. Key architectural

factors, including network depth, layer width, regularisation and activation functions are seldom examined systematically, making it difficult to identify the sources of performance improvements. Second, evaluations are often conducted on outdated or non-IoMT datasets such as NSL-KDD and CICIDS2017 or focus only on binary classification, failing to capture the multiclass, imbalanced and heterogeneous nature of real IoMT traffic. Third, cross-dataset validation is largely absent as models achieving high accuracy on a single dataset are rarely tested for generalisation across datasets with different characteristics. Finally, the impact of modern regularisation techniques, particularly Batch Normalisation and Dropout remains underexplored in DBN fine-tuning for IoMT intrusion detection [1, 2, 6, 15].

This study addresses these limitations by conducting a systematic evaluation of DBN architectures for IoMT intrusion detection. Six DBN configurations are designed to isolate the effects of depth (2 to 4 RBM layers), width (128 to 512 units), regularisation (Batch Normalisation and Dropout) and activation functions. The models are evaluated on two complementary datasets: Canadian Institute for Cybersecurity- Internet of Medical Things (CIC IoMT 2024) with 18-class high-dimensional benchmark, and ECU-IoHT with a compact 5-class dataset enabling cross-dataset generalisation analysis. The best-performing model is further validated through five independent runs with results reported as mean ± standard deviation to ensure reproducibility. By controlling architectural variables while keeping all other experimental conditions fixed, this work provides a rigorous ablation-style comparison of DBN design choices and addresses a key methodological gap in IoMT intrusion detection research.

2. METHODOLOGY

The methodology evaluates DBN architectures for intrusion detection in IoT healthcare environments. A systematic experimental framework was designed to study the effects of network depth, width, regularisation, and activation function on classification performance. Six different DBN configurations are tested on the CIC IoMT 2024 and ECU-IoHT datasets. All experiments follow a two-phase training paradigm: first, unsupervised layer-wise pretraining via RBMs, followed by supervised fine-tuning of the complete network. A flow diagram of the proposed approach is shown in Figure 1.

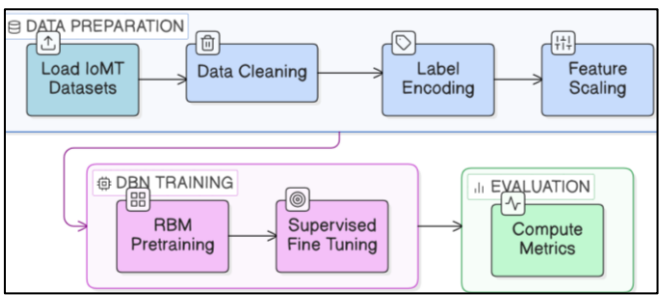

Figure 1. Flow diagram of the proposed approach

2.1 Canadian Institute for Cybersecurity Internet of Medical Things 2024 dataset

The experiments are done using the CIC IoMT 2024 dataset, which already has an official split for train and test sets. This fixed division is kept without change to ensure the natural distribution is not broken and to allow for fair comparison with other works, and also the imbalance of IoMT traffic stays realistic. No extra methods like resampling, oversampling, undersampling, or class weighting are used; the models are forced to learn directly from raw traffic distribution, which can be more difficult but also closer to real conditions.

The dataset includes those categories [20]:

- DDoS
- DoS
- Reconnaissance
- MQTT-based attacks
- Spoofing

In total, there are 18 different attack variants included. The data contains 46 features taken from header statistics, protocol metadata, and also behavior properties, such as flow duration, packet rate, TCP flag counters, and header length [20].

2.2 Edith Cowan University-Internet of Health Things dataset

The ECU-IoHT dataset was created in to solve a problem which is the lack of public healthcare datasets for detecting attacks on NIDS systems. This dataset was generated from a realistic IoHT devices, and was mainly based on the Libelium MySignals healthcare kit, which includes body sensors like temperature, blood pressure, and heart rate measurements which send biometric data to a cloud server using Wi-Fi and Bluetooth, and this enables vulnerability study to be done in more practical conditions [23].

The dataset includes benign traffic and four primary attack categories executed via established penetration testing methodologies:

- ARP Spoofing (ARP Poisoning)
- DoS
- Reconnaissance (Nmap Port Scanning)
- Smurf Attacks (DDoS variant)

Additionally, script injection attacks are included which are done by using the MITM framework, where the feature extraction was done with Argus and TShark software, where statistical flow features are taken, like source and destination bytes, packet numbers, and protocol metadata for TCP, TLS, ICMP, DNS, and ARP. The dataset has a total of 8,905 flows, and it shows a natural imbalance, which is similar to real IoHT systems. For example, Smurf attacks (77,920) are very high in number and represent significant anomaly group, on the other hand, DoS attacks (639) are quite low, and this situation gives a harder and more realistic condition for testing DL detection performance [23].

2.3 Data preprocessing

Preprocessing is kept minimal to isolate the influence of model architecture on classification performance. The following steps were applied identically to both datasets in the given order:

(1) Rows containing any null (NaN) values were removed via the dropna function. No imputation was performed.
(2) Exact duplicate rows were removed via the drop_duplicates function.
(3) Categorical attributes, including the target label column only were encoded to integers using scikit-learn's LabelEncoder. For each categorical column, the

encoder was fitted on the concatenation of training and test sets to ensure that all categories appear in the encoder's vocabulary, then transformation was applied separately to the training and test splits. This procedure is applied to the label column only and does not cause feature-level data leakage, because the target label carries no information about the input feature distribution.

(4) Feature scaling: Numerical features were standardised using a StandardScaler. The scaler was fitted exclusively on the training set and then applied to the test set, ensuring no information leakage from test data into the scaling parameters.

(5) Train/test split: For CIC IoMT 2024, the original authors' pre-defined 80/20 stratified split was used without modification. For ECU-IoHT, which lacks an official split, the data was randomly partitioned into 80% training and 20% testing using a fixed random seed (random_state = 42).

No data balancing strategies are introduced, ensuring that results reflect realistic deployment conditions. The ECU-IoHT dataset was evaluated using cross-validation of 5 folds, since it did not provide a split from the dataset author.

2.4 Deep Belief Network architecture

2.4.1 Restricted Boltzmann Machine

The fundamental building block of each DBN is the RBM, a generative stochastic model consisting of a visible layer $v \in \{0,1\}^{n_v}$ and a hidden layer $h \in \{0,1\}^{n_h}$ with no intra-layer connections. The energy function of an RBM is defined as in Eq. (1):

$$E(v,h) = -v^{\top}Wh - b^{\top}v - c^{\top}h \quad (1)$$

where, $W \in \mathbb{R}^{n_v \times n_h}$ is the weight matrix, $b \in \mathbb{R}^{n_v}$ is the visible bias vector, and $c \in \mathbb{R}^{n_h}$ is the hidden bias vector. The conditional distributions are given by Eq (2) and Eq. (3):

$$P(h_j = 1 \mid \mathrm{v}) = \sigma\left(\sum_i W_{ij}\, v_i + c_j\right) \quad (2)$$

$$P(v_i = 1 \mid \mathrm{h}) = \sigma\left(\sum_j W_{ij}\, h_j + b_i\right) \quad (3)$$

where, $\sigma(\cdot)$ denotes the sigmoid activation function. Weights are initialised from a Gaussian distribution $\mathcal{N}(0, 0.01^2)$ and biases are initialised to zero.

2.4.2 Two-phase training paradigm

All DBN variants follow a consistent two-phase training procedure:

Phase 1: Unsupervised Pretraining**.** Each RBM layer is trained greedily using Contrastive Divergence with one step of Gibbs sampling (CD-1). The input to the first RBM is the preprocessed feature vector; subsequent RBMs receive the hidden activations of the preceding layer. Training is performed using Stochastic Gradient Descent (SGD) with a learning rate of $\eta_{pre} = 0.01$ for 10 epochs per RBM. The batch size during pretraining is 2048. This unsupervised phase initialises the network weights to capture the underlying data distribution, providing a favourable starting point for discriminative fine-tuning.

Phase 2: Supervised Fine-Tuning**.** After pretraining, the RBM weights are transferred to a feedforward neural network with an appended softmax classification layer. The full network is trained end-to-end using the Adam optimiser with an initial learning rate of $\eta_{ft} = 0.01$. A Cosine Annealing learning rate schedule is applied over 100 epochs, with a minimum learning rate of $\eta_{min} = 0.0001$. The loss function is Cross-Entropy Loss. Batch size during fine-tuning is 2048. Data is shuffled via random permutation at the start of each epoch.

2.5 Experimental design

Six experiments are designed to systematically isolate the effect of a single architectural variable, with Experiment 2 (Standard DBN) serving as the reference baseline. The hidden-unit dimensions (128, 256, 512) follow the widely adopted convention of powers-of-two scaling, which is theoretically justified by the binary nature of RBM hidden units (each unit models a Bernoulli random variable) and practically motivated by efficient GPU memory alignment. The Standard DBN configuration (128–128–64) mirrors the three-layer architecture frequently cited as a de facto baseline in DBN-based IDS literature (Sohn, 2021). The Shallow DBN drops one layer to isolate depth; the Deep DBN extends to four layers with a progressively narrowing bottleneck (256 → 128 → 64 → 32) to evaluate hierarchical compression; the Wide DBN quadruples the first-layer width to 512 to test representational capacity independently of depth. Table 1 summarises all six configurations.

Table 1. Summary of Deep Belief Network (DBN) experiment configurations

Name	Architecture	Key Variable	Layers
Shallow DBN	input → 128 → 64 → output	Depth (shallow)	2
Standard DBN	input → 128 → 128 → 64 → output	Baseline	3
Deep DBN	input → 256 → 128 → 64 → 32 → output	Depth (deep)	4
Wide DBN	input → 512 → 256 → 128 → output	Width	3
Regularised DBN	input → 128(BN + Drop) → 128(BN + Drop) → 64(BN + Drop) → output	Regularisation	3
ReLU DBN	input → 128(ReLU) → 128(ReLU) → 64(ReLU) → output	Activation	3

Note: All experiments use CD-1 pretraining (10 epochs, SGD, $lr = 0.01$) and Adam fine-tuning (100 epochs, $lr = 0.01$, cosine annealing).

2.5.1 Experiment 1: Shallow Deep Belief Network

The Shallow DBN is formed by two RBM layers (128 → 64 hidden units), then a softmax classifier is added after it. This setup is used as a minimal depth baseline, in order to check if a shallow representation can be enough for separating the 18 traffic classes in IoT Healthcare intrusion detection.

2.5.2 Experiment 2: Standard Deep Belief Network (baseline)

The Standard DBN is built with three RBM layers (128 → 128 → 64 hidden units), and it is used as the main reference

model, against which all other configurations are compared. The design structure is similar to common DBN designs found in network intrusion detection studies, and it ensures a kind of balance between representation ability and computational cost.

2.5.3 Experiment 3: Deep Deep Belief Network

The network is extended into four RBM layers (256 → 128 → 64 → 32 hidden units) where the features are gradually compressed step by step. This experiment investigates the deeper feature hierarchy's ability to capture discriminative patterns in IoTH traffic, and if improve the classification of rare or overlapping attack types.

2.5.4 Experiment 4: Wide Deep Belief Network

The Wide DBN keeps three RBM layers, but their width is increased strongly (512 → 256 → 128 hidden units), so the total number of parameters becomes about four times more than the Standard DBN. This experiment is designed to test whether wider hidden layers can give more capacity for modeling complex decision boundaries between the 18 traffic classes.

2.5.5 Experiment 5: Regularised Deep Belief Network

The Regularised DBN augments the Standard DBN topology (128 → 128 → 64) with Batch Normalisation and Dropout ($p = 0.3$) after each hidden layer during supervised fine-tuning. Batch Normalisation stabilises the distribution of layer activations, while Dropout acts as an implicit ensemble method that discourages co-adaptation of neurons. The forward pass applies sigmoid activation, followed by Batch Normalisation, and then Dropout at each hidden layer, as shown in Eq. (4):

$$h^{(l)} = Dropout\left(BN\left(\sigma\left(W^{(l)}h^{(l-1)} + b^{(l)}\right)\right), p = 0.3\right) \quad (4)$$

This experiment evaluates whether regularisation during fine-tuning improves generalisation on unseen IoT Healthcare traffic.

2.5.6 Experiment 6: ReLU Deep Belief Network

The Rectified Linear Unit (ReLU) DBN replaces sigmoid activations with ReLU activations in the supervised fine-tuning phase, while retaining sigmoid activations during unsupervised pretraining (as required by the RBM's probabilistic formulation). The architecture remains identical to the Standard DBN (128 → 128 → 64). This experiment investigates whether modern activation functions yield superior gradient flow and faster convergence compared to the traditional sigmoid, which has historically been used in DBN architectures.

All experiments are evaluated using the following metrics to provide a comprehensive assessment of classification performance:

(1) Accuracy: The proportion of correctly classified instances across all classes given by Eq. (5):

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \quad (5)$$

(2) Area Under the ROC Curve (AUC): Computed using the One-vs-Rest (OVR) strategy, the AUC measures the model's ability to distinguish each class from all others. A macro-averaged AUC and per-class AUC values are reported by Eq. (6):

$$AUC_{macro} = \frac{1}{C}\sum_{c=1}^{C} AUC_c \quad (6)$$

where, $C = 18$ is the number of classes.

(3) Precision, Recall, and F1-Score: Per-class and macro-averaged precision, recall, and F1-score are computed from the classification report by Eqs. (7)–(9):

$$Precision_c = \frac{TP_c}{TP_c + FP_c} \quad (7)$$

$$Recall_c = \frac{TP_c}{TP_c + FN_c} \quad (8)$$

$$F1_c = \frac{2 \cdot Precision_c \cdot Recall_c}{Precision_c + Recall_c} \quad (9)$$

2.6 Implementation details

Models are implemented using the PyTorch framework and executed on a GPU. Scikit-learn is used for preprocessing using the StandardScaler and LabelEncoder modules. Reproducibility is ensured by setting a fixed random seed of 42 before any data splitting or model initialisation. Key implementation parameters are summarized in Table 2.

Table 2. Hyperparameters for all experiments

Parameter	Value
Pretraining algorithm	CD-1 (Contrastive Divergence, k = 1 Gibbs step)
Pretraining optimiser	SGD, lr = 0.01
Pretraining epochs per Restricted Boltzmann Machine (RBM)	10
Pretraining batch size	2,048
Fine-tuning optimiser	Adam, lr = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$
Fine-tuning scheduler	CosineAnnealingLR (T_max = 100, $\eta_{min} = 1 \times 10^{-5}$)
Fine-tuning epochs	100
Fine-tuning batch size	2,048
Dropout rate (Regularised Deep Belief Network (DBN))	0.15
Loss function	Cross-Entropy Loss
Weight initialisation (RBM)	W ~ $\mathcal{N}$ (0, 0.01), biases = 0
Feature scaling	StandardScaler (zero mean, unit variance)
Random seed (data split & base init)	42
Repeated-run seeds (Section 4.2.11)	1, 2, 3, 4, 5

3. RESULTS AND DISCUSSION

3.1 Experimental overview

We conducted a comprehensive evaluation of DBN architectures for network intrusion detection across two distinct IoT healthcare traffic datasets. The two benchmarks differ in scale, dimensionality and classification complexity enabling cross-dataset assessment of DBN architectural

choices. Table 3 shows a comparative overview between the datasets.

The identical experimental protocol is applied to both datasets. All DBN variants undergo 10 epochs of unsupervised RBM pretraining via Contrastive Divergence (CD-1) with a learning rate. $\eta = 0.01$, followed by 100 epochs of supervised fine-tuning with the Adam optimiser under a cosine annealing schedule (initial $\eta = 0.01$, $\eta_{\min} = 10^{-4}$) and mini-batch size 2,048. Input features are standardised using StandardScaler. Six architectural variants are evaluated on both datasets.

A key distinction between the two benchmarks is the severity of the classification challenge: the CIC IoMT 2024 dataset presents 18 traffic classes across a 45-dimensional feature space with nearly 900 K test instances, while ECU-IoHT comprises 5 attack categories over a compact 10-feature space with 22 K test instances. As will be shown, this difference in intrinsic problem difficulty leads to qualitatively different behaviours across DBN architectures, providing complementary insight into the conditions under which each architectural choice is beneficial.

Table 3. Comparative overview of experimental datasets and evaluation conditions

Property	CIC Internet of Medical Things (IoMT) 2024	Edith Cowan University-Internet of Health Things (ECU-IoHT)
Total instances	~4,461,330	111,207
Training instances	~3,569,064	88,966
Test instances	892,266	22,241
Number of classes	18	5
Input features	45	10
Split strategy	Stratified 80/20	Cross validation 5 folds

3.2 Results on the Canadian Institute for Cybersecurity Internet of Medical Things 2024 dataset

3.2.1 Overview of classification performance

Table 4 presents the main evaluation metrics for all six DBN configurations on the CIC IoMT 2024 test set. For Experiment 1 and Experiment 2, an independent second training run was additionally conducted in order to quantify sensitivity to random initialization; these extra results are discussed within the respective experiment sections.

Table 4. Summary of classification performance across all six Deep Belief Network (DBN) experiments on Canadian Institute for Cybersecurity- Internet of Medical Thing (CIC IoMT 2024)

Architecture	Acc.	Weighted F1	Test Loss
Shallow	0.9521	0.9490	0.3513
Standard	0.8714	0.8576	0.5270
Deep	0.9403	0.9348	0.3194
Wide	0.9469	0.9417	0.2755
Regularised	0.9718	0.9687	0.2071
ReLU	0.9715	0.9686	0.5853

The Regularised DBN (Experiment 5) is showing the highest overall performance, with a test accuracy is 97.18%, and macro-AUC is 0.9970. This result is considered stable in the evaluation stage. The ReLU DBN (Experiment 6) is almost the same in accuracy, about 97.15%, but AUC becomes lower at 0.9850; this means the probability output is less well calibrated, even if the final class decision is nearly identical, so some uncertainty quality is reduced.

The Standard DBN baseline (Experiment 2, main run) is recording the lowest test accuracy, 87.14%, and test loss is 0.5270, which is consistent with a training process that likely falls into not good local minimum in that particular run. However, another independent run of the same configuration is reaching 97.66% accuracy, and this suggests strong randomness in training outcome when regularisation is not applied, and performance is unstable across runs.

3.2.2 Experiment 1: Shallow Deep Belief Network

The Shallow DBN comprises two RBM pretraining layers (128 and 64 hidden units), the most compact architecture in the study. It achieves a test accuracy of 95.21% and a macro-AUC of 0.9887 in its primary run, as shown in Table 5.

A second independent training run produced substantially improved results: test accuracy of 97.87% and macro-AUC of 0.9952 (weighted F1 = 0.9760). The accuracy improvement of 2.66 percentage points between runs highlights the sensitivity of DBN training to random weight initialisation and the stochastic nature of CD-1 pretraining with a 45-feature input. The improved run shows markedly better precision and recall for Classes 'DDoS ICMP', 'DDoS SYN', 'DDoS TCP', and 'benign'.

Table 5. Shallow Deep Belief Network (DBN) per-class scores on Canadian Institute for Cybersecurity- Internet of Medical Thing (CIC IoMT 2024)

Class	Precision	Recall	F1-Score
Address Resolution Protocol (ARP) Spoofing	0.363	0.589	0.449
Distributed Denial of Service (DDoS) ICMP	0.869	0.983	0.922
DDoS Publish Flood	0.993	0.139	0.244
DDoS SYN	0.960	0.980	0.970
DDoS TCP	0.835	0.954	0.890
DDoS UDP	0.967	0.986	0.976
DoS Connect Flood	0.996	1.000	0.998
DoS ICMP	0.952	0.638	0.764
DoS Publish Flood	0.539	1.000	0.701
DoS SYN	0.990	0.961	0.975
DoS TCP	0.991	0.975	0.983
DoS UDP	0.960	0.914	0.937
Malformed Data	0.894	0.653	0.755
OS Scan	0.623	0.114	0.193
Ping Sweep	0.900	0.426	0.578
Port Scan	0.850	0.970	0.906
Recon VulScan	0.316	0.170	0.221
benign	0.943	0.930	0.936
Weighted avg	0.956	0.952	0.949

3.2.3 Experiment 2: Standard Deep Belief Network /baseline

The Standard DBN, which adds a third 128-unit hidden layer to the Shallow DBN, serves as the architectural reference baseline. In its primary run, it achieves 87.14% test accuracy and a macro-AUC of 0.9872, the lowest accuracy across all six experiments. As shown in Table 6.

3.2.4 Experiment 3: Deep Deep Belief Network

The Deep DBN employs four RBM pretraining layers with a progressively narrowing architecture (256→128→64→32), culminating in a 32-unit final representation. It achieves 94.03% test accuracy and a macro-AUC of 0.9924, as can be seen in Table 7.

Relative to the Standard DBN primary run, the Deep DBN

model shows an increase in test accuracy by 6.89% and also gives a slightly higher AUC by +0.0052; this is seen in Table 7.

Table 6. Experiment 2: Standard Deep Belief Network (DBN) per-class precision, recall, and F1-score on Canadian Institute for Cybersecurity- Internet of Medical Thing (CIC IoMT 2024)

Class	Precision	Recall	F1-Score
Address Resolution Protocol (ARP) Spoofing	0.327	0.493	0.393
Distributed Denial of Service (DDoS) ICMP	0.724	0.948	0.821
DDoS Publish Flood	0.986	0.126	0.223
DDoS SYN	0.924	0.970	0.946
DDoS TCP	0.680	0.687	0.684
DDoS UDP	0.837	0.981	0.903
DoS Connect Flood	0.994	1.000	0.997
DoS ICMP	0.561	0.150	0.236
DoS Publish Flood	0.536	1.000	0.698
DoS SYN	0.980	0.925	0.952
DoS TCP	0.937	0.948	0.942
DoS UDP	0.910	0.500	0.646
Malformed Data	0.888	0.667	0.762
OS Scan	0.588	0.119	0.198
Ping Sweep	0.811	0.456	0.583
Port Scan	0.831	0.968	0.894
Recon VulScan	0.325	0.166	0.220
benign	0.940	0.920	0.930
Weighted avg	0.879	0.871	0.858

Table 7. Experiment 3: Deep Deep Belief Network (DBN) per-class precision, recall, and F1-score on Canadian Institute for Cybersecurity- Internet of Medical Thing (CIC IoMT 2024)

Class	Precision	Recall	F1-Score
Address Resolution Protocol (ARP) Spoofing	0.256	0.536	0.347
Distributed Denial of Service (DDoS) ICMP	0.719	0.988	0.832
DDoS Publish Flood	0.995	0.127	0.225
DDoS SYN	0.923	0.971	0.946
DDoS TCP	0.765	0.892	0.824
DDoS UDP	0.966	0.988	0.977
DoS Connect Flood	0.998	1.000	0.999
DoS ICMP	0.801	0.087	0.158
DoS Publish Flood	0.536	1.000	0.698
DoS SYN	0.981	0.925	0.952
DoS TCP	0.978	0.959	0.968
DoS UDP	0.965	0.911	0.937
Malformed Data	0.891	0.641	0.746
OS Scan	0.642	0.124	0.208
Ping Sweep	0.822	0.574	0.676
Port Scan	0.860	0.972	0.913
Recon VulScan	0.393	0.169	0.236
benign	0.940	0.914	0.927
Weighted avg	0.946	0.940	0.935

Class 'DoS UDP' has a great improvement, recall becomes 0.911 compared to 0.500 in Standard primary; however, Class 'DoS ICMP' recall is still very low at 0.087 and this is a critical problem. The gap between train and test accuracy is 2.63 percentage points (train: 96.66%, test: 94.03%), which can suggest moderate overfitting situation. The 32-unit bottleneck may cause compression risk in representation, and the still-low recall for Class 'DoS ICMP' and also Class 'OS Scan' (0.124) shows difficulty in keeping minority-class discriminative patterns through four sequential stochastic Bernoulli sampling stages.

3.2.5 Experiment 4: Wide Deep Belief Network

The Wide DBN quadruples the capacity of the first hidden layer (512 units versus 128 in the Standard DBN), maintaining a three-layer RBM stack. It achieves 94.69% test accuracy and macro-AUC of 0.9935.

Class 'DoS Connect Flood' achieves a perfect F1-score of 1.000, and Class 'DoS UDP' recovers substantially (F1 = 0.977, compared to 0.646 in the Standard DBN primary run) as shown in Table 8. However, the Wide DBN is unable to overcome the persistent weaknesses observed in Classes 'DDoS Publish Flood', 'DoS ICMP', 'OS Scan', and 'Recon VulScan'.

Table 8. Experiment 4: Wide Deep Belief Network (DBN) scores on Canadian Institute for Cybersecurity- Internet of Medical Thing (CIC IoMT 2024)

Class	Precision	Recall	F1-Score
Address Resolution Protocol (ARP) Spoofing	0.259	0.530	0.348
Distributed Denial of Service (DDoS) ICMP	0.721	0.977	0.830
DDoS Publish Flood	0.956	0.123	0.218
DDoS SYN	0.904	0.971	0.936
DDoS TCP	0.679	0.670	0.674
DDoS UDP	0.991	0.991	0.991
DoS Connect Flood	1.000	1.000	1.000
DoS ICMP	0.683	0.102	0.177
DoS Publish Flood	0.536	1.000	0.698
DoS SYN	0.981	0.906	0.942
DoS TCP	0.934	0.950	0.942
DoS UDP	0.977	0.979	0.977
Malformed Data	0.905	0.635	0.746
OS Scan	0.577	0.113	0.188
Ping Sweep	0.878	0.509	0.644
Port Scan	0.850	0.970	0.906
Recon VulScan	0.331	0.132	0.188
benign	0.935	0.910	0.922
Weighted avg	0.951	0.947	0.942

3.2.6 Experiment 5: Regularised Deep Belief Network

The Regularised DBN augments the Standard DBN topology with BN and Dropout applied after each hidden layer during supervised fine-tuning; the RBM pretraining phase is unchanged. It achieves the best overall performance across all primary runs: test accuracy of 97.18%, and macro-AUC of 0.9970, as shown in Table 9. To address run-to-run variability, this configuration was additionally evaluated across 5 independent training runs (seeds 1–5) using the same fixed train/test split. The repeated-run summary is: accuracy = 97.72 ± 0.04%, macro-AUC = 0.9940 ± 0.0023, and weighted F1 = 0.9742 ± 0.0004 (mean ± sample std). The per-class F1 mean ± std across the 5 runs is reported in Table 9.

Relative to the Standard DBN run, the Regularised DBN delivers marked per-class improvements: Class 1 F1 +0.142 (0.963 vs. 0.821), Class 'DDoS TCP' F1 +0.235 (0.919 vs. 0.684), Class 'DoS ICMP' F1 +0.664 (0.900 vs. 0.236, the most significant per-class improvement in the study), and Class 'DoS UDP' F1 +0.345 (0.991 vs. 0.646), confirming that BN and Dropout together steer the optimiser towards better-

generalising solutions. All 18 classes exceed a per-class AUC of 0.980, indicating reliable discrimination in the probabilistic domain. Class 'DDoS Publish Flood' remains the most persistent challenge (recall = 0.138, F1 = 0.241), consistent across all models, suggesting that its feature distribution substantially overlaps with neighbouring classes in the 45-dimensional input space. Classes 'OS Scan' and 'Recon VulScan' also retain low F1-scores (0.163 and 0.209).

Table 9. Experiment 5: Regularized Deep Belief Network (DBN) performance on Canadian Institute for Cybersecurity-Internet of Medical Thing (CIC IoMT 2024) (single primary run)

Class	Precision	Recall	F1-Score
Address Resolution Protocol (ARP) Spoofing	0.396	0.508	0.445
Distributed Denial of Service (DDoS) ICMP	0.948	0.978	0.963
DDoS Publish Flood	0.933	0.138	0.241
DDoS SYN	0.971	0.979	0.975
DDoS TCP	0.882	0.959	0.919
DDoS UDP	0.997	0.995	0.996
DoS Connect Flood	0.985	1.000	0.992
DoS ICMP	0.946	0.859	0.900
DoS Publish Flood	0.540	0.999	0.701
DoS SYN	0.990	0.967	0.978
DoS TCP	0.991	0.987	0.989
DoS UDP	0.988	0.994	0.991
Malformed Data	0.914	0.592	0.719
OS Scan	0.795	0.091	0.163
Ping Sweep	0.942	0.385	0.546
Port Scan	0.848	0.973	0.906
Recon VulScan	0.416	0.140	0.209
benign	0.931	0.945	0.938
Weighted avg	0.975	0.972	0.969

Figure 2 shows the per-class F1-score means and standard deviations across the 5 independent runs. The 5-run results confirm that the Regularised DBN performance is stable where the standard deviation of accuracy is only 0.045 percentage points across runs. Most classes exhibit tight F1 distributions (Table 10), though 'Ping Sweep' (std = 0.067) and 'Recon VulScan' (std = 0.032) show the highest relative variability, attributable to their small support sizes (169 and 973 instances, respectively).

Table 10. Per-class F1-score mean ± std for the Regularised Deep Belief Network (DBN) over 5 independent runs (seeds 1-5) on Canadian Institute for Cybersecurity- Internet of Medical Thing (CIC IoMT 2024)

Class	F1 Mean	F1 Std
Address Resolution Protocol (ARP) Spoofing	0.3511	0.0201
Distributed Denial of Service (DDoS) ICMP	0.9919	0.0012
DDoS Publish Flood	0.2212	0.0183
DDoS SYN	0.9870	0.0010
DDoS TCP	0.9509	0.0050
DDoS UDP	0.9984	0.0001
DoS Connect Flood	0.9937	0.0012
DoS ICMP	0.9751	0.0027
DoS Publish Flood	0.6972	0.0041
DoS SYN	0.9900	0.0008
DoS TCP	0.9968	0.0006
DoS UDP	0.9975	0.0004
Malformed Data	0.5899	0.0042
OS Scan	0.1490	0.0146
Ping Sweep	0.6502	0.0666
Port Scan	0.9168	0.0021
Recon VulScan	0.0731	0.0324
benign	0.9286	0.0045

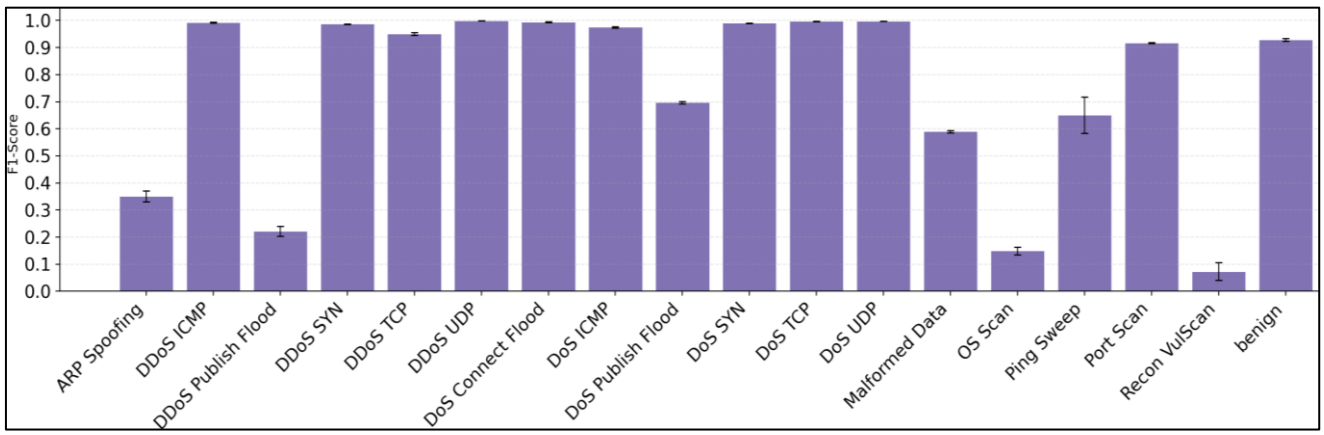

Figure 2. F1 mean ± std bars with error caps for 5-run validation

3.2.7 Experiment 6: ReLU Deep Belief Network

The ReLU DBN replaces sigmoid activations in the supervised fine-tuning phase with Rectified Linear Units while retaining sigmoid activations during RBM pretraining achieving test accuracy of 97.15% and macro-AUC of 0.9850, as shown in Table 11.

Figure 3 shows the per-class precision, recall, and F1-score for the ReLU DBN on CIC IoMT 2024. The ReLU DBN matches the Regularized model closely in test accuracy (97.15% vs. 97.18%) and weighted F1-score (0.9686 vs. 0.9687), despite near-identical accuracy, reflecting the known tendency of ReLU networks trained with Adam to assign sharp, overconfident probability mass to dominant classes without correspondingly calibrating the full softmax distribution. This is evidenced by the AUC of 0.9850, the lowest in the results, and the lowest per-class AUC for Class 'DDoS Publish Flood' of 0.821, indicating that ReLU activations, while beneficial for decision-boundary sharpness, degrade probabilistic discrimination for inherently difficult classes.

3.2.8 Comparative analysis

(1) Effect of network depth

On the contrary, the general expectation that deeper architectures learn more discriminative hierarchical representations, the Shallow DBN outperforms the Standard DBN by 8.07% in accuracy in the primary run comparison as can be seen in Table 12. The Deep DBN partially improves at 94.03%, exceeding the Standard DBN by 6.89%, but falls 1.18% points below the Shallow DBN. This non-monotonic depth-performance relationship is partly attributable to the training instability of the Standard DBN primary run; its second run achieves 97.66%, rendering the Deep DBN inferior by a further 3.63 percentage points. The AUC improvement of the Deep DBN over the Shallow DBN (+0.0037) is modest and does not translate to higher accuracy. These results indicate that for the CIC IoMT 2024 dataset, additional RBM depth beyond two layers does not yield consistent accuracy gains while incurring proportional training time increases.

(2) Effect of network width

The Wide DBN improves upon the Standard DBN by 7.55 percentage points in accuracy, as shown in Table 13, demonstrating that greater layer capacity assists convergence on the 45-feature, 18-class input. However, this improvement comes at a 2.14× training time cost, and relative to the Standard DBN accuracy (97.66%), the Wide DBN is inferior by 2.97 percentage points, indicating that wider layers do not overcome the fundamental convergence instability of unregularised DBN training. The Regularised DBN achieves superior accuracy (97.18%), confirming that regularisation is both more effective and more computationally efficient than brute-force width expansion.

Table 11. Experiment 6: ReLU Deep Belief Network (DBN) results on Canadian Institute for Cybersecurity- Internet of Medical Thing (CIC IoMT 2024)

Class	Precision	Recall	F1-Score
Address Resolution Protocol (ARP) Spoofing	0.292	0.470	0.360
Distributed Denial of Service (DDoS) ICMP	0.936	0.974	0.955
DDoS Publish Flood	0.971	0.145	0.252
DDoS SYN	0.980	0.980	0.980
DDoS TCP	0.878	0.967	0.920
DDoS UDP	0.996	0.995	0.996
DoS Connect Flood	0.983	1.000	0.991
DoS ICMP	0.948	0.819	0.879
DoS Publish Flood	0.542	0.999	0.703
DoS SYN	0.993	0.973	0.983
DoS TCP	0.996	0.991	0.994
DoS UDP	0.984	0.995	0.989
Malformed Data	0.873	0.627	0.730
OS Scan	0.901	0.077	0.142
Ping Sweep	0.780	0.462	0.580
Port Scan	0.854	0.975	0.910
Recon VulScan	0.342	0.151	0.210
benign	0.930	0.928	0.929
Weighted avg	0.976	0.971	0.969

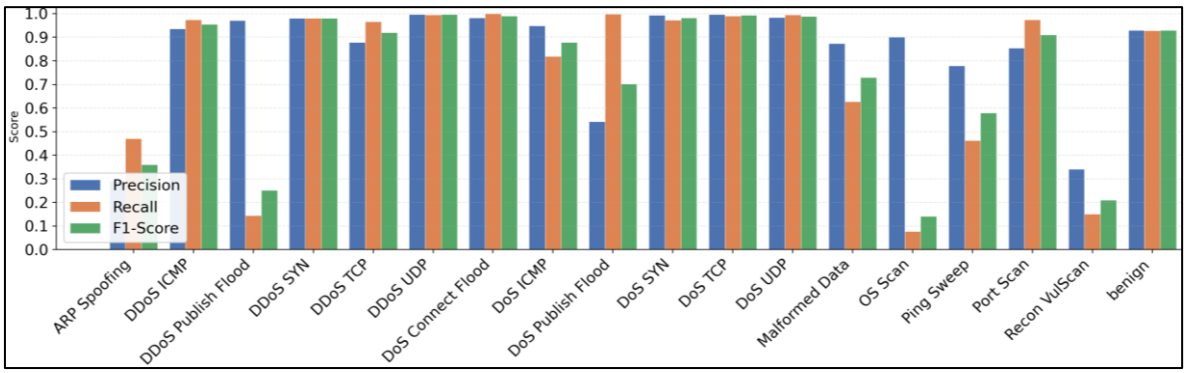

Figure 3. Per-class precision/recall/F1 for ReLU Deep Belief Network (DBN)

Table 12. Classification performance as a function of Deep Belief Network (DBN) depth on Canadian Institute for Cybersecurity- Internet of Medical Thing (CIC IoMT 2024)

Architecture	RBM Layers	Test Accuracy	AUC
Shallow	2	0.9521	0.9887
Standard	3	0.8714	0.9872
Deep	4	0.9403	0.9924

Note: Restricted Boltzmann Machine = RBM; Area Under the ROC Curve = AUC.

Table 13. Classification performance as a function of first-layer width on CIC Internet of Medical Things (IoMT) 2024

Architecture	First-Layer Width	Test Acc.	AUC
Standard	128	0.8714	0.9872
Wide	512	0.9469	0.9935

(3) Effect of regularisation

As shown in Table 14, the Regularised DBN substantially outperforms the unregularised Standard DBN. Against the Standard DBN, the Regularised DBN delivers a 10.04 percentage point accuracy improvement, confirming that regularisation provides the convergence stability that the unregularised model achieves only unreliably.The macro-AUC of 0.9970 is the highest recorded in the study, attributable to Batch Normalisation's beneficial effect on class probability calibration.

Table 14. Effect of Batch Normalisation and dropout ($p = 0.3$) on Deep Belief Network (DBN) performance on Canadian Institute for Cybersecurity- Internet of Medical Thing (CIC IoMT 2024)

Configuration	Test Accuracy	AUC
Standard	0.8714	0.9872
Regularised (BN + Drop)	0.9718	0.9970

(4) Effect of activation function

The ReLU DBN achieves 97.15% accuracy compared to 87.14% in the Standard DBN (Table 15) a difference of 10.01

percentage points, reflecting confident class probability assignments characteristic of unregularised ReLU networks under Adam optimisation. ReLU is thus a cost-free accuracy improvement but sacrifices probabilistic discrimination quality relative to the regularised sigmoid model.

Table 15. Effect of activation function on Deep Belief Network (DBN) performance on Canadian Institute for Cybersecurity- Internet of Medical Thing (CIC IoMT 2024)

Activation	Test Accuracy	Area Under the ROC Curve (AUC)
Sigmoid (Standard)	0.8714	0.9872
ReLU	0.9715	0.9850

3.2.9 Training dynamics and generalisation

Train-test accuracy gaps are uniformly modest across all six experiments (1.75–4.39 percentage points), indicating that no configuration is severely overfit in the classification sense. The Regularised DBN achieves the smallest accuracy gap (0.0183), confirming that BN and Dropout effectively constrain memorisation. Across 5 independent runs, the Regularised DBN's train-test accuracy gap averages 0.0183 with negligible variation, demonstrating that BN and Dropout not only improve single-run generalisation but also eliminate run-to-run divergence. Table 16 summarises these convergence statistics for all six experiments.

3.2.10 Class-level performance and persistent challenges

Three classes exhibit structural difficulty that no DBN variant resolves, as shown in Figure 4. Class 'DDoS Publish Flood' returns F1-scores of 0.218–0.252 with uniformly high precision (0.933–0.995) but extremely low recall (0.123–0.145), indicating near-complete misclassification of true instances into other categories; this pattern is consistent with substantial feature-space overlap between Class 'DDoS Publish Flood' and neighbours. Class 'OS Scan' achieves F1-scores of 0.142–0.208 with a recall below 0.124 in most experiments, despite a non-negligible support of 2,941. Class 'Recon VulScan' (973 test instances, the second-rarest class) reaches F1-scores of only 0.188–0.236, with both precision and recall in the low-to-mid range across all configurations.

Table 16. Training convergence statistics for all six experiments on Canadian Institute for Cybersecurity- Internet of Medical Thing (CIC IoMT 2024) (primary runs)

Exp.	Test Accuracy	Train Accuracy	Acc. Gap
Shallow	0.9521	0.9777	0.0256
Standard	0.8714	0.9153	0.0439
Deep	0.9403	0.9666	0.0263
Wide	0.9469	0.9717	0.0248
Regularised	0.9718	0.9901	0.0183
ReLU	0.9715	0.9890	0.0175

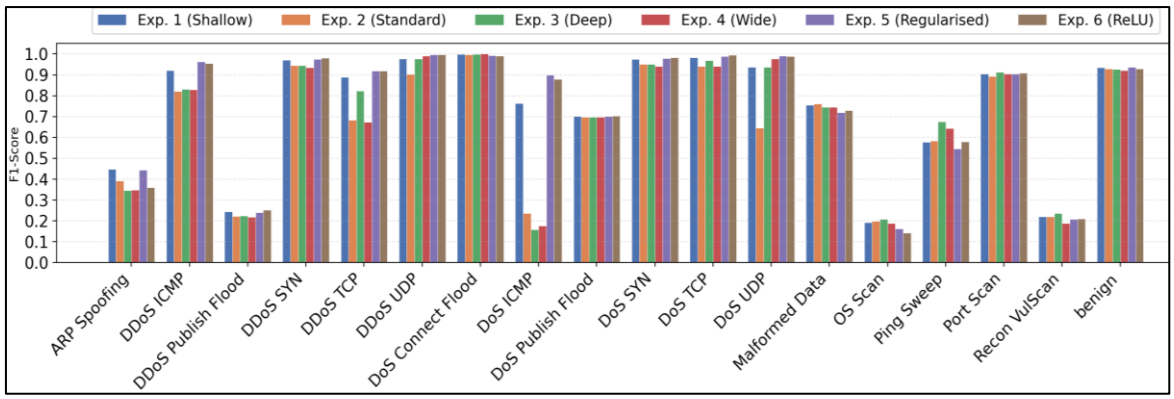

Figure 4. Per class F1-score for all experiments on CIC Internet of Medical Things (IoMT) 2024 as a bar chart representation

Table 17. Per-class F1-score across all six Canadian Institute for Cybersecurity- Internet of Medical Thing (CIC IoMT 2024) experiments (primary runs)

Class	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6
Address Resolution Protocol (ARP) Spoofing	0.449	0.393	0.347	0.348	0.445	0.360
Distributed Denial of Service (DDoS) ICMP	0.922	0.821	0.832	0.830	0.963	0.955
DDoS Publish Flood	0.244	0.223	0.225	0.218	0.241	0.252
DDoS SYN	0.970	0.946	0.946	0.936	0.975	0.980
DDoS TCP	0.890	0.684	0.824	0.674	0.919	0.920
DDoS UDP	0.976	0.903	0.977	0.991	0.996	0.996
DoS Connect Flood	0.998	0.997	0.999	1.000	0.992	0.991
DoS ICMP	0.764	0.236	0.158	0.177	0.900	0.879
DoS Publish Flood	0.701	0.698	0.698	0.698	0.701	0.703
DoS SYN	0.975	0.952	0.952	0.942	0.978	0.983
DoS TCP	0.983	0.942	0.968	0.942	0.989	0.994
DoS UDP	0.937	0.646	0.937	0.977	0.991	0.989
Malformed Data	0.755	0.762	0.746	0.746	0.719	0.730
OS Scan	0.193	0.198	0.208	0.188	0.163	0.142
Ping Sweep	0.578	0.583	0.676	0.644	0.546	0.580
Port Scan	0.906	0.894	0.913	0.906	0.906	0.910
Recon VulScan	0.221	0.220	0.236	0.188	0.209	0.210
benign	0.936	0.930	0.927	0.922	0.938	0.929

In contrast, Class 'DoS Publish Flood' attains a stable F1 near 0.70 across all experiments by achieving a recall ≈ of 1.000 at the cost of a precision ≈ 0.54, a pattern invariant to architectural choice. Class 'DoS Connect Flood' performs near-perfectly (F1 ≥ 0.991) in all six experiments, indicating complete separability from all other classes. Detailed results for all the runs are presented in Table 17.

3.2.11 Repeated-run stability of the regularised DBN

To quantify training stability and address the concern that DBN performance is sensitive to random initialization, the Regularised DBN (the best-performing configuration) was trained 5 times with independent random seeds (1–5) on the fixed CIC IoMT 2024 train/test split. Table 18 summarises the results.

The low standard deviations of 0.045 percentage points for accuracy and 0.23 points for AUC confirm that the Regularised DBN is effectively deterministic in outcome despite the stochastic nature of CD-1 pretraining and random weight initialization. The range (max-min) of accuracy across 5 runs is only 0.11 percentage points in stark contrast to the unregularised Standard DBN, whose two runs differed by 8.52 points. This stability is attributable to Batch Normalisation reducing internal covariate shift which narrows the effective loss landscape, and Dropout preventing co-adapted feature detectors that are brittle to initialization changes.

Table 18. Repeated-run statistics for the regularised Deep Belief Network (DBN) on Canadian Institute for Cybersecurity- Internet of Medical Thing (CIC IoMT 2024) (5 independent runs, fixed 80/20 split)

Metric	Mean	Std	Min	Max
Accuracy	0.9772	0.0004	0.9764	0.9775
Macro-Area Under the ROC Curve (AUC)	0.9940	0.0023	0.9911	0.9968
Weighted F1	0.9742	0.0004	0.9736	0.9747
Macro F1	0.7477	0.0042	0.7431	0.7518

3.3 Results on the Edith Cowan University-Internet of Health Things dataset

3.3.1 Overview of classification performance

Table 19 presents the top-level metrics for all six DBN configurations on the ECU-IoHT test fold.

Table 19. Summary of classification performance across all six Deep Belief Network (DBN) experiments on Edith Cowan University- Internet of Health Things (ECU-IoHT)

Model	Acc.	Weighted F1	Test Loss
Shallow	1.0000	1.0000	2.929×10^{-4}
Standard	0.9887	0.9886	5.824×10^{-2}
Deep	0.9738	0.9631	8.790×10^{-2}
Wide	1.0000	1.0000	8.612×10^{-5}
Regularised	1.0000	1.0000	3.750×10^{-6}
ReLU	0.9999	0.9999	1.223×10^{-3}

Three configurations, Shallow (Exp. 1), Wide (Exp. 4), and Regularised (Exp. 5), achieve perfect test accuracy (100.00%) and macro-AUC of 1.0000. The Regularised DBN attains the lowest test loss in the study (3.750×10^{-6}), reflecting maximally calibrated softmax outputs. The Deep DBN (Exp. 3) records the lowest overall accuracy (97.38%) and macro-AUC (0.9625), attributable to catastrophic collapse of minority-class predictions.

3.3.2 Experiment 1: Shallow DBN

Despite comprising only two RBM pretraining layers, the Shallow DBN achieves perfect classification performance.

Per-class AUC: All five classes ≥ 0.9999; macro-AUC = 1.0000. Test loss (2.929×10^{-4}) and train loss (2.882×10^{-4}) are nearly identical, yielding a train–test loss gap of 4.7×10^{-6}. Convergence is rapid: from an initial test loss of approximately 0.39 at epoch 1, the model descends below. 10^{-3} by approximately epoch 50. These results demonstrate that for the ECU-IoHT 10-feature space, a two-layer DBN is sufficient to learn a perfectly separable embedding of the five attack classes.

3.3.3 Experiment 2-Standard DBN/baseline

The Standard DBN, the architectural baseline, introduces a third 128-unit hidden layer. Unlike ECU-IoHT Shallow DBN this additional capacity is not beneficial where test accuracy falls to 98.87% as detailed in Table 20.

Per-class AUC: ARP Spoofing = 1.0000, DoS Attack = 1.0000, Nmap Port Scan = 1.0000, No Attack = 0.9523, Smurf Attack = 0.9622; macro-AUC = 0.9829.

Errors are entirely concentrated in the No Attack-Smurf Attack interface. No Attack instances are misclassified as Smurf Attack, indicating that the Standard DBN's latent representations do not perfectly resolve the statistical boundary between these two classes.

Table 20. Experiment 2: Standard Deep Belief Network (DBN) per-class precision, recall, and F1-score on Edith Cowan University- Internet of Health Things (ECU-IoHT)

Class	Precision	Recall	F1-Score
Address Resolution Protocol (ARP) Spoofing	1.0000	1.0000	1.0000
DoS Attack	1.0000	1.0000	1.0000
Nmap Port Scan	1.0000	1.0000	1.0000
No Attack	1.0000	0.9460	0.9723
Smurf Attack	0.9841	1.0000	0.9920
Weighted avg	0.9888	0.9887	0.9886

3.3.4 Experiment 3: Deep DBN

The Deep DBN employs four RBM layers. The 32-unit bottleneck imposes a 10:1 compression on the 10-feature input, producing catastrophic failure for the two smallest minority classes, as shown in Table 21.

All ARP Spoofing and all DoS Attack instances are misclassified as Nmap Port Scan, yielding precision and recall of exactly 0.000 for classes 0 and 1. The macro F1-score of 0.5653 is the lowest recorded across both datasets and all experiments. This collapse is characteristic of representational exhaustion in deep bottleneck architectures: the 32-unit final layer cannot maintain the subtle statistical distinctions between low-frequency minority classes and the adjacent majority class under heavily imbalanced CD-1 pretraining. While per-class AUC values retain some residual signal (ARP Spoofing: 0.9329; DoS Attack: 0.9053), the supervised fine-tuning stage fails to exploit this signal, given the extreme compression. From an intrusion detection perspective, the complete failure to detect ARP Spoofing and DoS Attacks

renders this architecture unsuitable for healthcare IoT deployment.

Table 21. Experiment 3: Deep Deep Belief Network (DBN) per-class metrics on Edith Cowan University- Internet of Health Things (ECU-IoHT)

Class	Precision	Recall	F1-Score
ARP Spoofing	0.0000	0.0000	0.0000
DoS Attack	0.0000	0.0000	0.0000
Nmap Port Scan	0.7051	1.0000	0.8271
No Attack	1.0000	0.9987	0.9994
Smurf Attack	1.0000	1.0000	1.0000
Weighted avg	0.9556	0.9738	0.9631

Note: Per-class AUC: ARP Spoofing = 0.9329, DoS Attack = 0.9053, Nmap Port Scan = 0.9741, No Attack = 1.0000, Smurf Attack = 1.0000; macro-AUC = 0.9625. Address Resolution Protocol = ARP.

3.3.5 Experiment 4: Wide DBN

The Wide DBN achieves perfect classification on ECU-IoHT, with the most rapid convergence trajectory in the study.

Per-class AUC: All five classes = 1.0000; macro-AUC = 1.0000. The perfect Wide DBN result confirms that the Standard DBN's errors arise from insufficient representational capacity at a given depth level, rather than from a fundamental limitation of the sigmoid DBN architecture on this dataset.

3.3.6 Experiment 5: Regularized DBN

The Regularized design achieves perfect classification with the lowest test loss in the entire study across both datasets. To verify stability this configuration was further evaluated using 5 repeated 10-fold cross-validation runs (seeds 1–5). Across all 50 folds, the model achieved perfect accuracy AUC and weighted F1 of (1.0 ± 0.0) with an average test loss of 0.0060 ± 0.0009.

Per-class AUC: All classes ≥ 0.9999 (1.0000-ε); macro-AUC = 1.0000. Batch Normalization is the principal driver of this accelerated convergence by eliminating internal covariate shift across the sigmoid layers. The zero variance across all 50 test folds confirms that the Regularised DBN produces perfectly reproducible decisions on the ECU-IoHT dataset regardless of random seed.

3.3.7 Experiment 6: ReLU DBN

The ReLU DBN achieves near-perfect performance with only a few misclassifications, as detailed in Table 22.

Table 22. Experiment 6: ReLU Deep Belief Network (DBN) per-class metrics on Edith Cowan University-Internet of Health Things (ECU-IoHT)

Class	Precision	Recall	F1-Score	Support
Address Resolution Protocol (ARP) Spoofing	1.0000	1.0000	1.0000	444
DoS Attack	1.0000	0.9925	0.9962	133
Nmap Port Scan	0.9993	0.9993	0.9993	1,394
No Attack	0.9998	1.0000	0.9999	4,670
Smurf Attack	1.0000	1.0000	1.0000	15,600
Weighted avg	0.9999	0.9999	0.9999	22,241

Per-class AUC: ARP Spoofing = 1.0000, DoS Attack = 0.99999, Nmap Port Scan = 0.99946, No Attack = 0.99999, Smurf Attack = 1.0000; macro-AUC = 0.9999.

One DoS Attack instance is misclassified as Nmap Port Scan, and one Nmap Port Scan instance as No Attack; all other classes are classified without error.

3.3.8 Comparative analysis

(1) Effect of network depth on ECU-IoHT

Performance degrades monotonically with depth, as shown in Table 23, where the Shallow DBN achieves perfect accuracy, the Standard DBN produces 252 misclassifications, and the Deep DBN produces 577 misclassifications concentrated entirely in the minority classes. Two compounding factors drive this degradation. First, greedy CD-1 pretraining through multiple stochastic Bernoulli sampling stages progressively distorts the low-dimensional input signal, discarding the fine-grained structure critical for minority-class separation. Second, the 32-unit terminal layer imposes a 10:1 compression ratio on the 10-feature input, providing insufficient representational space to preserve statistical distinctions between the rare-class patterns and the dominant Nmap Port Scan category. For shallow, low-dimensional IoT traffic features, representational efficiency favors a shallow two-layer network over deeper abstraction.

Table 23. Classification performance on Edith Cowan University-Internet of Health Things (ECU-IoHT) in terms of Deep Belief Network (DBN) depth

Architecture	RBM Layers	Test Accuracy	AUC
Shallow	2	1.0000	1.0000
Standard	3	0.9887	0.9829
Deep	4	0.9738	0.9625

Note: Restricted Boltzmann Machine = RBM; Area Under the ROC Curve = AUC.

(2) Effect of network width on ECU-IoHT

Table 24 shows the effect of increasing first-layer width on ECU-IoHT performance. Increasing the first-layer width from 128 to 512 units eliminates all misclassifications. The fourfold expansion provides exponentially greater binary pattern capacity, ensuring that the CD-1 algorithm converges to well-separated class-discriminative features within 10 pretraining epochs.

Table 24. Classification performance as a function of first-layer width on Edith Cowan University-Internet of Health Things (ECU-IoHT)

Architecture	First-Layer Width	Test Accuracy	AUC
Standard	128	0.9887	0.9829
Wide	512	1.0000	1.0000

Note: Area Under the ROC Curve = AUC.

(3) Effect of Regularization on ECU-IoHT

Table 25 reports the effect of applying Batch Normalisation and Dropout on the performance of ECU-IoHT. The transition to the Regularized DBN improves the accuracy, which was seen in the simultaneous improvement in training fit and generalization, reflecting the synergistic effect of Batch Normalization, which accelerates gradient flow by eliminating internal covariate shift, and Dropout, which forces distributed, robust decision boundaries. The resultant probabilistic calibration is orders of magnitude superior to any other configuration in the study.

(4) Effect of activation function on ECU-IoHT

Table 26 compares the effect of sigmoid and ReLU activations on ECU-IoHT performance. ReLU activation

reduces misclassifications by 99.2% and decreases test loss by a factor of 47.6, at negligible computational cost. This gain is attributable to ReLU's full gradient magnitude for positive activations, enabling more effective error back-propagation through sigmoid-pretrained weight matrices. However, the Regularised DBN continues to achieve lower test loss, underscoring the value of Batch Normalisation for probabilistic calibration.

Table 25. Effect of Batch Normalization and dropout ($p = 0.3$) on performance on Edith Cowan University-Internet of Health Things (ECU-IoHT)

Configuration	Test Accuracy	AUC
Standard (no regularization)	0.9887	0.9829
Regularized (BN + Dropout 0.3)	1.0000	1.0000

Note: Area Under the ROC Curve = AUC.

Table 26. Effect of activation function on the performance of Edith Cowan University-Internet of Health Things (ECU-IoHT)

Activation	Test Accuracy	AUC
Sigmoid (Standard)	0.9887	0.9829
ReLU	0.9999	0.9999

Note: Area Under the ROC Curve = AUC.

3.3.9 Training dynamics on Edith Cowan University-Internet of Health Things

Table 27 presents the training convergence statistics for all six ECU-IoHT experiments. In sharp contrast to CIC IoMT 2024, where train-test accuracy gaps range from 1.75 to 4.39 percentage points, ECU-IoHT experiments exhibit near-zero accuracy gaps for all configurations except the Standard and Deep DBNs.

Table 27. Training convergence statistics for all six Edith Cowan University-Internet of Health Things (ECU-IoHT) experiments

Model	Test Accuracy	Train Accuracy	Acc. Gap
Shallow	1.0000	1.0000	0.0000
Standard	0.9887	0.9888	−0.0001
Deep	0.9738	0.9726	+0.0012
Wide	1.0000	1.0000	0.0000
Regularized	1.0000	1.0000	0.0000
ReLU	0.9999	0.9999	0.0000

3.4 Cross-dataset analysis

3.4.1 Dataset complexity and its effect on achievable accuracy ceilings

The most salient distinction between the two experimental series is the qualitative difference in achievable accuracy. On the ECU-IoHT-a 5-class, 10-feature, three of six DBN configurations attain perfect test accuracy (100.00%), and the remaining four exceed 97.38%. On CIC IoMT 2024, an 18-class, 45-feature benchmark, no configuration achieves perfect accuracy; the best primary-run result is 97.18%, and three persistent class-level challenges (Classes 2, 13, 16) remain unresolved by any architecture tested.

This disparity reflects two fundamental differences in problem difficulty. The first one is that the larger class count (18 vs. 5) demands substantially finer discrimination in the latent feature space, particularly for rare classes where the training signal is proportionally weaker under imbalanced CD-1 pretraining. While the Second is where the higher input dimensionality (45 vs. 10 features) introduces a richer feature-interaction landscape that benefits from the greedy layer-wise pretraining procedure of the DBN.

3.4.2 Consistency of the regularized DBN across datasets

The Regularized DBN (BN + Dropout) is the single configuration that achieves the best or joint-best performance across every primary metric on both datasets. On CIC IoMT 2024 it ranks first in test accuracy (97.72 ± 0.04% on 5 runs) and AUC (0.9940 ± 0.0023). On ECU-IoHT it achieves perfect classification (1.00 ± 0.0 across 5 of 10-fold CV runs) with the lowest test loss (3.750×10^{-6} in the primary run; 0.006 ± 0.0009 across repeated CV). This consistent superiority confirms that regularization via BN and Dropout is the most robust and dataset agnostic architectural strategy for DBN intrusion detection in IoT healthcare environments. BN primary contribution is convergence acceleration and calibration stability while Dropout's contribution is the prevention of co-adaptive feature dependencies which is beneficial irrespective of input dimensionality or class count.

3.4.3 Relationship between dataset complexity and model depth

The Deep DBN differing results of 94% accuracy on CIC IoMT 2024 but 0% F1 on ECU-IoHT, which comes from how DBN handles input size, class count and layer compression. Two key factors explain this:

Compression impact: On CIC IoMT, a 32-unit terminal layer mildly compresses a large input preserving key patterns. On ECU-IoHT multiple layers aggressively shrink a small 10-feature input (2:1 final compression) losing subtle details that distinguish rare classes (e.g., ARP Spoofing) from dominant ones (Nmap Port Scan).

In the Deep DBN, each RBM layer performs CD-1 pretraining using stochastic Bernoulli steps. On the low-dimensional ECU-IoHT dataset with small class margins, each step adds cumulative error that gradually erodes decision boundaries, especially for minority classes. By the fourth layer, noise drowns out class signals. In contrast, the higher-dimensional CIC IoMT 2024 dataset provides redundancy preserving enough discriminative information through the bottleneck.

The 18-class CIC IoMT 2024 task benefits from deeper layers that help separate overlapping classes. But ECU-IoHT's 5-class problem is nearly linearly separable with just two layers (the Shallow DBN achieves 100% accuracy, Table 28). Adding more layers introduces unnecessary complexity and confuses fine-tuning, especially given severe class imbalance (ARP Spoofing: 236 instances, DoS Attack: 64 instances).

Table 28. Cross-dataset performance for the best and worst Deep Belief Network (DBN) configurations on each dataset

Metric	CIC IoMT 2024 Best	CIC IoMT 2024 Worst	ECU-IoHT Best	ECU-IoHT Worst
Acc.	0.9718 (Reg.)	0.8714 (Std.)	1.0000 (Sha. / Wide / Reg.)	0.9738 (Deep)
Area Under the ROC Curve (AUC)	0.9970 (Reg.)	0.9850 (ReLU)	1.0000 (multiple)	0.9625 (Deep)

Note: Canadian Institute for Cybersecurity- Internet of Medical Things = CIC IoMT 2024; Edith Cowan University-Internet of Health Things = ECU-IoHT.

These findings suggest a dataset-dependent depth rule for DBN-based IoMT intrusion detection: use a shallow two-layer network when input features are few and classes are hard to separate; deeper networks help only when features are rich (e.g., 30+ features with 10+ classes). This aligns with prior work noting that optimal depth always depends on the dataset.

Width has opposite effects on different datasets. On ECU-IoHT (low-dimensional), a wide DBN achieves perfect accuracy, outperforming the standard version. On CIC IoMT 2024 (high-dimensional, multi-class), the wide DBN gets 94.69% accuracy. Expansion helps simple, low-dimensional problems but fails to scale to complex, high-dimensional ones, where regularization offers more benefit than simply adding capacity.

3.4.4 Training stochasticity and reproducibility

The two datasets exhibit fundamentally different training stability profiles. On ECU-IoHT run-to-run variability is low while the Standard DBN produces 252 misclassifications in its primary run; the architecture is consistent in its error pattern by always misclassifying No Attack as Smurf Attack. All six architectures that converge successfully do so to near-identical solutions with near-zero accuracy gaps. On CIC IoMT 2024, the Standard DBN primary and secondary runs differ by 8.52 percentage points in accuracy and the Shallow DBN two runs differ by 2.66 percentage points, a substantial variance for a deterministic architecture under fixed hyperparameters. This instability is a direct consequence of the larger and more complex problem landscape with 18 classes and 45 features; the CD-1 energy landscape contains more local minima, and the greedy layer-wise weight initialization is more sensitive to the random seed that determines the initial Boltzmann machine parameter values.

The practical implication is that on complex multi-class IoT traffic datasets, the unregularized DBN architectures require either multiple training runs with selection of the best outcome or the systematic application of Batch Normalization and Dropout to reduce sensitivity to initialization. Quantitatively, the Regularized DBNs' accuracy standard deviation across 5 runs is only 0.045 percentage points on CIC IoMT 2024 and 0.000 on ECU-IoHT compared with an 8.52-percentage-point spread observed across two runs of the unregularised Standard DBN. The Regularized DBN thus eliminates this instability on both datasets making it the architecturally preferred choice for deployment.

3.4.5 ReLU activation-accuracy vs. calibration tradeoff

The ReLU DBN demonstrated a consistent cross-dataset pattern where it achieved near-optimal accuracy at negligible computational overhead over the sigmoid baseline, but sacrifices probabilistic calibration as measured by AUC. On ECU-IoHT, the ReLU model reduces errors by 99.2% and on CIC IoMT 2024, it nearly matches the Regularized DBN in accuracy within 0.03 percentage points, but records an AUC lower by 0.0120. For applications requiring calibrated probability outputs, such as risk stratification in healthcare IoT monitoring where continuous risk scores are clinically actionable, the regularized sigmoid architecture is preferred. For applications requiring only hard classification labels under tight computational budgets, the ReLU DBN provides an effective and efficient alternative.

3.5 Comparison with existing literature

The results are broadly consistent with prior research employing DBNs and other DL approaches for IoMT intrusion detection, while also revealing important differences in evaluation conditions and behaviors.

Most existing studies report very high accuracies in the range of 98%–99.9%. For instance, Jayanthi et al [1] achieved 98.71% using an optimized DBN with metaheuristic tuning, while Manimurugan et al. [15] reported up to 99.37% accuracy on the CICIDS 2017 dataset. Similarly, Otoum et al. [20] and Huda et al. [21]. reported near-perfect accuracies exceeding 99.7%. These results are comparable to the performance achieved in this work on the ECU-IoHT dataset where multiple DBN configurations reached 100% accuracy.

However, a key distinction lies in dataset complexity and evaluation realism. Many prior works rely on older or simplified benchmarks (e.g., NSL-KDD or CICIDS 2017) or binary classification settings, which are inherently less challenging. This study evaluates models on the more complex CIC IoMT 2024 dataset with18 classes, where the best achieved accuracy reached 97.18%. This gap highlights the increased difficulty of modern IoMT intrusion detection tasks and suggests that previously reported near-perfect results may not generalize to more realistic scenarios.

Recent non-DBN works support this observation. Uddin et al. [2] reported 99.77% accuracy on CICIoMT2024 using hierarchical meta-learning, while Mohammadi et al. [22] achieved 99% with a CNN model. Additionally, Algethami and Alshamrani [8] reported perfect performance on ECU-IoHT, consistent with the results observed in this work for low-dimensional datasets.

Overall, the comparison indicates that while DBNs remain competitive, particularly when properly regularized, their performance is sensitive to dataset complexity and architectural design.

3.6 Practical applicability and security engineering considerations

A security-focused view adds key issues for real-world DBN-based IDS in IoMT are false alarms, latency, drift resilience and feasibility.

False alarms matter in healthcare, where each alert requires clinician review and too many false ones cause fatigue. The regularised DBN hits high precision (≥0.90) in 14 of 18 CIC IoMT 2024 classes, but is weaker on 'DoS Publish Flood' (~0.54) and 'ARP Spoofing' (~0.40) creating extra triage load. On ECU-IoHT, its flawless precision removes this problem. Any real rollout needs either class-specific thresholds or verification for low-precision attacks.

The DBN's offline two-phase training is computationally expensive but acceptable for periodic updates. Inference is just a single forward pass through a few dense layers achieving sub-millisecond latency on modern edge hardware.

Concept drift and temporal robustness. IoMT traffic changes over time due to new devices, attacks and shifting patterns. Current DBN tests use static data missing this drift. A two-phase approach helps where unsupervised pretraining can update with new unlabeled data and supervised fine-tuning can adjust with small labeled batches.

Security engineering summary. The Regularised DBN suits IoMT with 97.72% accuracy, low false positives, fast inference and an adaptable architecture. However, three attack classes have high false positives, needing extra mitigation like human review.

4. CONCLUSION

This study provided a systematic investigation of the influence of DBN architecture design on IDS performance in an IoTH environment. Experiments were done on CIC IoMT 2024 and ECU-IoHT datasets and results showed that DBN performance is sensitive to architecture choices especially the regularization; the depth and the width are seen to be important.

Findings confirm that regularization is a critical element for achieving stable and robust performance. The Regularized DBN with Batch Normalization and Dropout showed better performance compared to all other models across both datasets. On CIC IoMT 2024, it reached 97.72 ± 0.04% accuracy and 0.994 ± 0.0023 AUC (mean ± std, 5 runs), while on ECU-IoHT it achieved perfect classification (1.0 ± 0. across 5 × 10-fold CV). Training stability was quantitatively validated where the Regularised DBN's run-to-run accuracy spread was only 0.11 percentage points, compared to an 8.52-point spread for the unregularised Standard DBN. This confirms the strong importance of controlling overfitting and stabilising learning behavior in DBN-based IDS.

In contrast, increasing depth of network did not contribute to consistent improvement. Deeper architectures in some cases provided a small benefit in high-dimensional data and also created instability. This problem was more obvious in ECU-IoHT dataset, where Deep DBN failed to detect some important minority attack classes. Increasing width of network improved performance in low-dimensional scenarios, but it was not efficient for more complex datasets with higher dimensions. The Shallow DBN showed that a simpler structure can still achieve competitive or even better results when feature space is compact and well arranged.

Additionally, ReLU activation functions improved classification accuracy; probabilistic calibration was worse compared to sigmoid-based regularized models. This shows an important trade-off between classification strength and probability reliability.

Despite the strong overall outcomes, some limitations exist. Persistent misclassification of some classes, especially Classes 'DDoS Publish Flood', 'OS Scan', and 'Recon VulScan' in CIC IoMT 2024, indicates a feature overlap problem and difficulty caused by imbalanced distributions of data. Furthermore, the study focused only on DBN architectures, and no comparison was made with other modern DL models such as CNNs. The lack of data balancing methods, although intentional, may reduce performance in highly skewed class situations. Finally, repeated-run statistics (mean ± std over 5 seeds) are currently available only for the Regularised DBN configuration; extending this analysis to all six architectures would further strengthen confidence in the reported comparisons.

This study analyses DBN designs for IoT healthcare intrusion detection. Findings show well regularised, moderately complex models balance accuracy and generalisation best. However, three attack classes (DDoS Publish Flood, OS Scan, VulScan) underperform consistently, with F1-scores below 0.25.

While the Regularised DBN provides a strong baseline, deploying it securely in healthcare would need extra steps like boosting data for weak classes or combining it with a secondary detector, or using a two-stage system with simple checks to filter traffic first.

ACKNOWLEDGMENT

The authors would like to express gratitude to Mustansiriyah University – Baghdad, Iraq, for the support of this work.

REFERENCES

[1] Jayanthi, S., Suhasini, S., Sharmili, N., et al. (2025). A deep dive into artificial intelligence with enhanced optimization-based security breach detection in the Internet of Health Things-enabled smart city environment. Scientific Reports, 15(1): 22909. https://doi.org/10.1038/s41598-025-05850-z

[2] Uddin, M.A., Chu, N.H., Rafeh, R. (2025). A hierarchical IDS for zero-day attack detection in Internet of medical things networks. arXiv preprint, arXiv:2508.10346. https://doi.org/10.48550/arXiv.2508.10346

[3] Manoharan, A., Thathan, M. (2024). Enhanced IoMT security framework using group teaching optimized auto-encoder for intrusion detection. Scientific Reports, 14(1): 30360. https://doi.org/10.1038/s41598-024-80581-1

[4] Kumari, M., Gaikwad, M., Chavan, S.A. (2025). A secure IoT-edge architecture with data-driven AI techniques for early detection of cyber threats in healthcare. Discover Internet of Things, 5(1): 54. https://doi.org/10.1007/s43926-025-00147-z

[5] Rbah, Y., Mahfoudi, M., Fattah, M., et al. (2024). Hybrid software-defined network-based deep learning framework for enhancing Internet of medical things cybersecurity. IAES International Journal of Artificial Intelligence (IJ-AI), 13(3): 3599. https://doi.org/10.11591/ijai.v13.i3.pp3599-3610

[6] Sohn, I. (2021). Deep belief network based intrusion detection techniques: A survey. Expert Systems with Applications, 167: 114170. https://doi.org/10.1016/j.eswa.2020.114170

[7] Vijayakumar, K.P., Pradeep, K., Balasundaram, A., Prusty, M.R. (2023). Enhanced cyber attack detection process for internet of health things (IoHT) devices using deep neural network. Processes, 11(4): 1072. https://doi.org/10.3390/pr11041072

[8] Algethami, S.A., Alshamrani, S.S. (2024). A deep learning-based framework for strengthening cybersecurity in internet of health things (IoHT) environments. Applied Sciences, 14(11): 4729. https://doi.org/10.3390/app14114729

[9] Hasan, F.M., Mahmood, S.A., Hussien Saeed, E.M. (2025). Driver activities detection system based on optimized ReSNet101 deep learning model and transfer learning. International Journal of Intelligent Engineering & Systems, 18(3). https://inass.org/wp-content/uploads/2024/12/2025043020-2.pdf.

[10] Anuva, S.T., Iqbal, S., Zulkernine, M. (2025). LIDIT: Low-latency intrusion detection in IoMT devices using TinyML. In GLOBECOM 2025-2025 IEEE Global Communications Conference, Taipei, Taiwan, pp. 6093-6098. https://doi.org/10.1109/GLOBECOM59602.2025.11432258

[11] Mohseni, N.A., Saeed, E.M.H. (2025). Hybrid feature representation of face images via mesh landmark

encoding and lightweight CNNs. Advanced Engineering, Technology and Applications on Power Systems, 1764: 619-632. https://doi.org/10.1007/978-3-032-13921-4_44

[12] Alharith, R., Ahmed, H., Ibrahim, A.O., Saleh, M.A., Saule, A., Saltanat, A. (2025). Anomaly detection in IoT healthcare security using machine learning methods. In 2025 IEEE 5th International Conference on Smart Information Systems and Technologies (SIST), Astana, Kazakhstan, pp. 1-6. https://doi.org/10.1109/SIST61657.2025.11139151.

[13] Alarcón García-Saavedra, A. (2024). Evaluation of ML models for network attack detection in IoMT environments. Projecte Final de Màster Oficial. https://hdl.handle.net/2117/423206.

[14] Kavkas, N.C., Yildiz, K. (2025). Enhancing IoMT security with deep learning based approach for medical IoT threat detection. In 2025 13th International Symposium on Digital Forensics and Security (ISDFS), Boston, MA, USA, pp. 1-5. https://doi.org/10.1109/ISDFS65363.2025.11012062

[15] Manimurugan, S., Al-Mutairi, S., Aborokbah, M.M., Chilamkurti, N., Ganesan, S., Patan, R. (2020). Effective attack detection in internet of medical things smart environment using a deep belief neural network. IEEE Access, 8: 77396-77404. https://doi.org/10.1109/ACCESS.2020.2986013

[16] Balakrishnan, N., Rajendran, A., Pelusi, D., Ponnusamy, V. (2021). Deep belief network enhanced intrusion detection system to prevent security breach in the Internet of Things. Internet of Things, 14: 100112. https://doi.org/10.1016/j.iot.2019.100112

[17] Dadkhah, S., Neto, E.C.P., Ferreira, R., Molokwu, R.C., Sadeghi, S., Ghorbani, A. (2024). Ciciomt2024: Attack vectors in healthcare devices-a multi-protocol dataset for assessing iomt device security. Internet of Things, https://doi.org/10.20944/preprints202402.0898.v1

[18] Kumar, M., Kim, S. (2024). Securing the internet of health things: Embedded federated learning-driven long short-term memory for cyberattack detection. Electronics, 13(17): 3461. https://doi.org/10.3390/electronics13173461

[19] Diro, A.A., Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for Internet of Things. Future Generation Computer Systems, 82: 761-768. https://doi.org/10.1016/j.future.2017.08.030

[20] Otoum, Y. (2022). AI-based intrusion detection systems to secure Internet of Things (IoT). Doctoral dissertation. University of Ottawa. http://doi.org/10.20381/ruor-28290

[21] Huda, S., Miah, S., Yearwood, J., Alyahya, S., Al-Dossari, H., Doss, R. (2018). Securing the operations in SCADA-IoT platform-based industrial control system using ensemble of deep belief networks. Applied Soft Computing, 71: 66-77. https://doi.org/10.1016/j.asoc.2018.06.006

[22] Mohammadi, A., Ghahramani, H., Asghari, S.A., Aminian, M. (2024). Securing healthcare with deep learning: A CNN-based model for medical IoT threat detection. In 2024 19th Iranian Conference on Intelligent Systems (ICIS), Sirjan, Iran, pp. 168-173. https://doi.org/10.1109/ICIS64839.2024.10887510

[23] Ahmed, M., Byreddy, S., Nutakki, A., Sikos, L.F., Haskell-Dowland, P. (2021). ECU-IoHT: A dataset for analyzing cyberattacks in Internet of Health Things. Ad Hoc Networks, 122: 102621. https://doi.org/10.1016/j.adhoc.2021.102621