



Please cite the Published Version

Abbas, Moneeb, Kuo, Wen-Chung, Mahmood, Khalid , Akram, Waseem, Mehmood, Sajid and Bashir, Ali Kashif  (2025) Conv-MTD: A CNN Based Multi-Label Medical Tubes Detection and Classification Model to Facilitate Resource-constrained Point-of-care Devices. IEEE Journal of Biomedical and Health Informatics. pp. 1-12. ISSN 2168-2208

DOI: <https://doi.org/10.1109/jbhi.2025.3543245>

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/638639/>

Usage rights:  In Copyright

Additional Information: This is an accepted manuscript of an article which appeared in IEEE Journal of Biomedical and Health Informatics

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

Conv-MTD: A CNN Based Multi-Label Medical Tubes Detection and Classification Model to facilitate resource-constrained point-of-care devices

Moneeb Abbas, Wen-Chung Kuo, Khalid Mahmood *Senior Member, IEEE*, Waseem Akram, Sajid Mehmood, and Ali Kashif Bashir *Senior Member, IEEE*

Abstract—Computer-aided detection through deep learning is becoming a prevalent approach across various fields, including the detection of anomalies in medical procedures. One such medical procedure involves the placement of medical tubes to provide nutrition or other medical interventions in critically ill patients. Medical tube placement can be highly complex and prone to subjective errors. Malposition of medical tubes is often observed and associated with significant morbidity and mortality. In addition, continuous verification using manual procedures such as capnography, pH testing, auscultation, and visual inspection through chest X-ray (CXR) imaging is required. In this paper, we propose a Conv-MTD, a medical tube detection (MTD) model that detects the placement of medical tubes using CXR images, assisting radiologists with precise identification and categorizing the tubes into normal, abnormal, and borderline placement. Conv-MTD leverages the EfficientNet-B7 architecture as its backbone, enhanced with an auxiliary head in the intermediate layers to mitigate vanishing gradient issues common in deep neural networks. The Conv-MTD is further optimized using post-training 16-bit floating-point (FP16) quantization, which significantly reduces memory consumption by 50% and 2x improvement in inference speed without compromising accuracy. This optimization allows Conv-MTD to achieve efficient performance without requiring high-end computational resources, making it suitable for deployment on point-of-care devices. Conv-MTD provided the best performance, with an average area under the receiver operating characteristic curve (AUC) of 0.95 using the open-source RANZCR CLIP dataset. The proposed Conv-MTD has the potential to operate on

resource-constrained point-of-care devices due to its use of FP16 computation, enabling low-cost and automated assessments in various healthcare settings.

Index Terms—Medical Tube Detection, Classification, Chest X-ray Analysis, Deep Learning in Radiology, Computer-Assisted Diagnosis

I. INTRODUCTION

THE current state-of-the-art convolutional neural networks (CNNs) can learn complex features from images, facilitating transformative applications in medical image analysis. Numerous CNNs are currently used in medical image analysis, enabling automated detection of abnormalities across various medical modalities [1]–[5]. One such application of these advancements is the precise identification and monitoring of medical tube placements. This proposed work examined four distinct medical tubes, namely endotracheal tube (ETT), nasogastric tube (NGT), Swan-Ganz catheter (SGC), and central venous catheter (CVC). The annotated multi-labelled Chest X-rays (CXRs) of these medical tubes are depicted in Fig 1.

The primary usage of ETT is to provide artificial ventilation to critically ill patients, and ETT malpositioning can result in the collapse of the left lung and overinflation of the right lung. Similarly, the main purpose of NGT is to provide nutrition by enteral feeding to patients unable to consume food or liquids orally, and NGT malpositioning could lead to aspiration pneumonia [6]–[9]. The same is true of CVC and Swan-Ganz catheters. Both devices are important to perform several medical procedures. Incorrect placement of CVCs can cause pneumothorax, hemothorax, and arterial puncture, while misplacement of SGC can result in vascular injury, thrombosis, air embolism, and extravasation [10]–[12]. These associated risks highlight the importance of accurate tube placement and identification to enhance patient care.

Currently, radiologists follow strict medical protocols with the utmost importance during tube placement, utilizing manual verification procedures such as capnography, pH testing, auscultation, and visual inspection through CXRs [13]–[21]. However, the aforementioned verification methods can be time-consuming, and human error remains a potential risk factor, particularly when hospitals are at full capacity. Given

Moneeb Abbas, Waseem Akram and Sajid Mehmood is with the Graduate School of Engineering Science and Technology, National Yunlin University of Science and Technology, Douliu 64002, Taiwan (e-mail: moneebabbasofficial@gmail.com, wasi.ahmad4@gmail.com, sajid21757@gmail.com).

Wen-Chung Kuo is with the Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Douliu 64002, Yunlin, Taiwan (e-mail: simonkuo@yuntech.edu.tw).

Khalid Mahmood is with the Graduate School of Intelligent Data Science, National Yunlin University of Science and Technology, Douliu 64002, Taiwan (e-mail: khalidm.research@gmail.com).

Ali Kashif Bashir is with the Department of Computing and Mathematics, Manchester Metropolitan University, United Kingdom and Centre for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, 140401, Punjab, India (emails: dr.alikashif.b@ieee.org).

these challenges, there is a pressing need for fast and reliable computer-assisted interpretation to overcome the limitations of the existing healthcare system.

Recently, deep neural networks such as EfficientNet, ResNet50, MobileNet, and Inception V3 have proven effective for automated medical image analysis [22]–[29]. Considering these successes, researchers have turned their attention to applying deep learning for automated medical tube detection using CXRs [30]–[45]. However, most of studies have focused on detecting a single type of medical tube, such as [30], [36]–[41], [43], [44], despite the fact that patients could have multiple tubes during ventilation, each serving a unique and critical function. Several studies try to overcome this limitation by proposing multi-label tube detection models such as [33], [35], [45]. To the best of our knowledge, none of the prior studies have addressed the added challenge of implementing such models on resource-constrained point-of-care devices. This complexity underscores the need for more sophisticated models capable of accurately detecting the position of multiple tubes in real-time while being computationally efficient.

In this study, EfficientNet-B7 is chosen over MobileNet, which is also well-suited for resource-constrained environments, due to its superior performance in accuracy and feature extraction, particularly in complex multi-label classification tasks. EfficientNet-B7 achieves a Top-1 accuracy of 84.4% on the ImageNet dataset, significantly outperforming MobileNetV2 Top-1 accuracy of 71.8% and MobileNetV3 Top-1 accuracy of 75.2% [46]. ResNet-50 achieves a Top-1 accuracy of 76.2%, while its deeper variant, ResNet-101, achieves 77.3%. Similarly, DenseNet-121 achieves 74.9% Top-1 accuracy, and the deeper DenseNet-201 achieves 77.3%. Despite EfficientNet-B7’s higher accuracy, it operates with approximately 37 billion FLOPs, more than MobileNetV3 Large at approximately 0.6 billion FLOPs, but provides a much better balance between computational cost and model scalability for handling large datasets. In addition, the architectural modifications in Conv-MTD, including auxiliary head and optimization techniques such as focal loss, enable the model to effectively handle class imbalance, a common challenge in medical datasets. Focal loss down-weights the loss contributions of easily classified samples while focusing more on hard to classify samples, leading to improved sensitivity and overall performance.

The primary objective of this study is to develop a deep learning model that can accurately detect and classify the placement of multi-label medical tubes in CXR images, with a focus on optimizing the model for deployment on resource-constrained devices.

Major contributions of this research are:

- The proposed Conv-MTD framework automates the detection and verification of medical tubes (ETT, NGT, SGC, and CVC) in multi-labelled CXRs, aiming to reduce reliance on manual methods that can be time-consuming and error-prone.
- The proposed model addresses two common issues in training neural networks: class imbalance and the vanishing gradient problem. It resolves this by giving more weight to hard-classified examples using focal loss and

an auxiliary head to incorporate additional loss functions to help stabilize training.

- The proposed model significantly reduces overall memory consumption and inference latency, providing a lightweight, scalable, and real-time solution to support radiologists across varied healthcare setting.

TABLE I: List of Abbreviations

Abbreviation	Meaning
CXR	Chest X-Rays
DNN	Deep Neural Network
CNN	Convolutional Neural Network
MTD	Medical Tube Detection
AUC	Area Under the Curve
ROC	Receiver Operator Characteristics
FP16	Floating Point 16
ETT	Endotracheal Tube
CVC	Central Venous Catheter
NGT	Nasogastric Tube
SGC	Swan Ganz Catheter

The list of commonly used abbreviations in this research is explained in Table I.

II. PROPOSED METHODOLOGY

The proposed Conv-MTD model is designed to automate the detection and classification of medical tube placements in CXRs by utilizing the EfficientNet-B7 architecture as its backbone. It is enhanced with an auxiliary head to support additional loss function and quantization of the trained weights to support resource-constrained devices. This section details the dataset and Conv-MTD framework, highlighting the architectural design and optimization strategies implemented to improve detection accuracy, address class imbalance, and ensure compatibility with resource-limited devices.

A. Dataset

In this study, we utilized the publicly available RANZCR CLIP dataset [47], which contains 30,083 labelled CXRs images specifically aimed at the detection and classification of medical tube placements. This dataset addresses a critical need for rapid identification of malpositioned catheters and lines, which, if undetected, can lead to severe complications or even fatal outcomes. The dataset includes three main categories of tube placements.

Normal: This category includes tubes that are positioned correctly and do not require any adjustments. **Borderline:** This category consists of tubes that, while generally functioning in their current positions, ideally need minor repositioning to ensure optimal patient safety and efficacy. **Abnormal:** This category encompasses critically misaligned tubes requiring immediate repositioning to avoid complications. The dataset was created with a consistent labelling protocol to ensure high-quality, reliable annotations across images, addressing potential issues of

TABLE II: Distribution of Labels

Label	Number of Images
ETT - Abnormal	500
ETT - Borderline	2000
ETT - Normal	7500
NGT - Abnormal	200
NGT - Borderline	250
NGT - Incompletely Imaged	300
NGT - Normal	4000
CVC - Abnormal	3500
CVC - Borderline	8000
CVC - Normal	21000
Swan Ganz Catheter Present	1500

labelling variance. The dataset class distribution is shown in Table II.

The dataset presents significant class imbalance challenges, particularly evident in the disparity between normal and abnormal cases. For example, CVC-Normal cases include 21,000 images, which substantially outnumber CVC-Abnormal cases with 3,500 images, while NGT-Abnormal cases represent only 200 images. Analysis of the class distribution revealed significant imbalances, with ratios of up to 1:105 between the least and most represented classes. To remedy this disparity, we enhanced the data set with a 3x augmentation along with focal loss, which dynamically adjusts the loss contribution of the majority class samples. In addition, we implemented a strict data segregation protocol to prevent any potential overlap between training and validation sets. The dataset is first split into training and validation sets with an 80:20 ratio. We ensured complete patient-level separation, meaning that multiple images from the same patient were kept within the same split to prevent data leakage. This separation is verified using the unique patient identifiers provided in the RANZCR CLIP dataset metadata. For the k-fold cross-validation process, we maintained this patient-level segregation across all folds, ensuring that our model performance metrics reflect genuine generalization capability rather than memorization of patient-specific features.

B. Data Preparation

To prepare the data for input into the model, each image is resized to 600x600 pixels to match the input dimensions required by the EfficientNet-B7 backbone.

1) *Data Augmentation*: CNNs are often susceptible to scaling and orientation issues during training. This study used random brightness and horizontal and vertical flipping techniques to mitigate these issues to simulate real-world variation in imaging conditions. Augmentation improves the model’s performance in handling scale and orientation issues common to clinical imaging [48]–[51]. The following transformations are applied to increase the robustness of the training data. where I represents an original CXR image.

Brightness Adjustment: Given an intensity level I , a brightness factor β is randomly sampled from a range $[\beta_{\min}, \beta_{\max}]$. The transformed image I' with adjusted brightness is computed as:

$$I' = I \times \beta, \quad \beta \in [0.8, 1.2] \quad (1)$$

This random adjustment helps the model adapt to varying lighting conditions commonly faced in CXR imaging.

Horizontal and Vertical Flips: Given that orientation inconsistencies can affect model performance, horizontal $F_H(I)$ and vertical $F_V(I)$ flips are applied probabilistically. The transformations are defined as:

$$I' = \begin{cases} F_H(I) & \text{with probability } p_H, \\ F_V(I) & \text{with probability } p_V. \end{cases} \quad (2)$$

where p_H and p_V are probabilities assigned to each flip, set at 0.5 for random application. These transformations improve generalization by simulating positional variations of tubes and lines within CXR images. The resulting images are presented in Fig. 2.

I’ll revise the section to better explain how focal loss improves model performance with imbalanced datasets. Here’s the professional revision with new content in blue:

2) *Handling Imbalance Data*: The RANZCR CLIP dataset is highly imbalanced. The imbalance dataset refers to a problem in which more examples are from one class than another. Imbalanced classes could render the model highly biased in favor of the more dominant class. To remedy the class imbalance problem, Focal loss [52]–[54] is utilized. Unlike traditional loss functions that treat all misclassifications equally, focal loss dynamically adjusts the loss contribution of each sample based on classification difficulty. Focal loss aims to emphasize hard classifier examples while reducing the weight of good classifier examples. Thus, the weight values of the incorrectly predicted sample are much higher as compared to the correctly predicted samples.

Focal Loss is defined as an extension of Cross-Entropy Loss, with an additional modulating term focusing on misclassified examples. For binary classification, the Cross-Entropy (CE) loss is given by:

$$CE(p, y) = \begin{cases} -\log(p) & \text{if } y = 1, \\ -\log(1 - p) & \text{otherwise,} \end{cases} \quad (3)$$

where p represents the model’s predicted probability for the actual class y .

In imbalanced datasets, standard cross-entropy loss can be problematic as the majority class samples dominate the gradient updates during training. Focal loss addresses this by automatically down-weighting the contribution of easy examples typically from the majority class and focusing on hard examples often from the minority class. To address the imbalance, Focal Loss (FL) introduces a modulating factor $(1 - p_t)^\gamma$ to the Cross-Entropy Loss, where p_t is the probability of the true class, defined as:

$$p_t = \begin{cases} p & \text{if } y = 1, \\ 1 - p & \text{otherwise.} \end{cases} \quad (4)$$

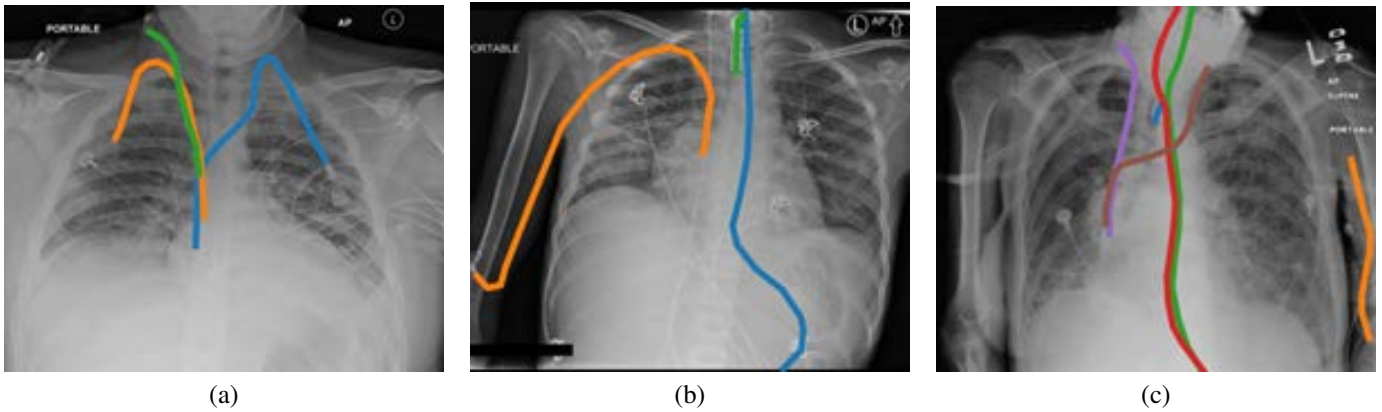


Fig. 1: (a) CXR image of a patient with CVC normal, abnormal, and borderline. (b) CXR image of a patient with NGT normal, CVC normal, and ETT normal. (c) CXR image of a patient with ETT Normal, CVC Borderline, NGT incomplete image, NGT incomplete image, CVC borderline, CVC borderline.

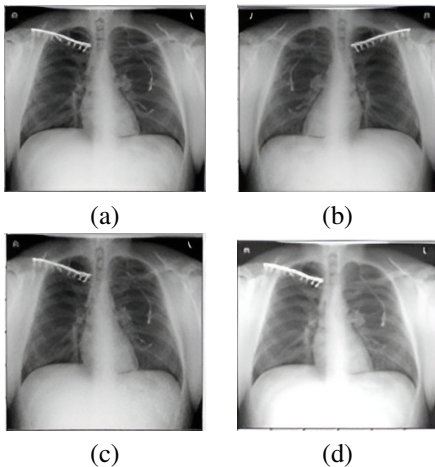


Fig. 2: Comparison of transformations applied to the original CXR image. (a) shows the original CXR image. (b) and (c) show the same image after random left-right and up-down flips, respectively. (d) shows the image after applying a random brightness transformation.

This leads to the formulation of Focal Loss as:

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t), \quad (5)$$

γ is a focusing parameter that adjusts the rate at which easy examples are down-weighted. It is typically set to values between 0.5 and 2 to fine-tune this adjustment. Through empirical testing on our medical tube detection task, we found that $\gamma = 2$ provides optimal performance by effectively reducing the loss contribution of well-classified majority class samples while maintaining sufficient gradient signal for learning. In Equation 5, when the value of the modulating term $F_\gamma = (1 - p_t)^\gamma$ increases, it minimizes the loss contribution of correctly predicted samples and increases the weight of incorrectly predicted samples. This adaptive weighting mechanism proves particularly effective for our medical tube detection task, where normal tube placements significantly out-

number abnormal cases, as it ensures the model maintains high sensitivity to rare but critical abnormal placements while preserving overall classification accuracy. Hence, the model becomes more balanced and accurate, leading to improved performance and results.

C. Model Architecture

The proposed framework utilizes a pre-trained EfficientNet-B7 as the base model. A pre-trained model offers several advantages over training from scratch, such as faster convergence due to prior optimization on large-scale datasets and improved accuracy from learned features. Fig. 3 shows the proposed methodology enhanced with an auxiliary head, which provides additional gradient flow, aiding effective backpropagation and contributing to refined feature extraction. This refinement enhances the model's ability to capture nuanced differences among tube types. The network is augmented with a global average pooling layer at the bottom in order to reduce feature map dimensions. This preserves essential information while streamlining the transition to classification. Following this, a dense layer with a sigmoid activation function is used for final classification.

1) *EfficientNet Model:* In this research, we used the pre-trained EfficientNet-B7 for weight initialization. One of the main benefits of using EfficientNet-B7 is that it uniformly scales up all the dimensions resolution, depth and width using a compound scaling technique. The compound scaling allows the model to outperform conventional CNNs in terms of accuracy and efficiency while reducing parameter size [46]. The selection of EfficientNet-B7 over other variants B0-B6 is driven by our empirical analysis of the trade-off between model complexity and computational requirements for medical tube detection. While smaller variants B0-B4 showed faster inference times, they could not capture fine-grained differences in tube positioning. B7 provided the optimal balance between accuracy and computational complexity

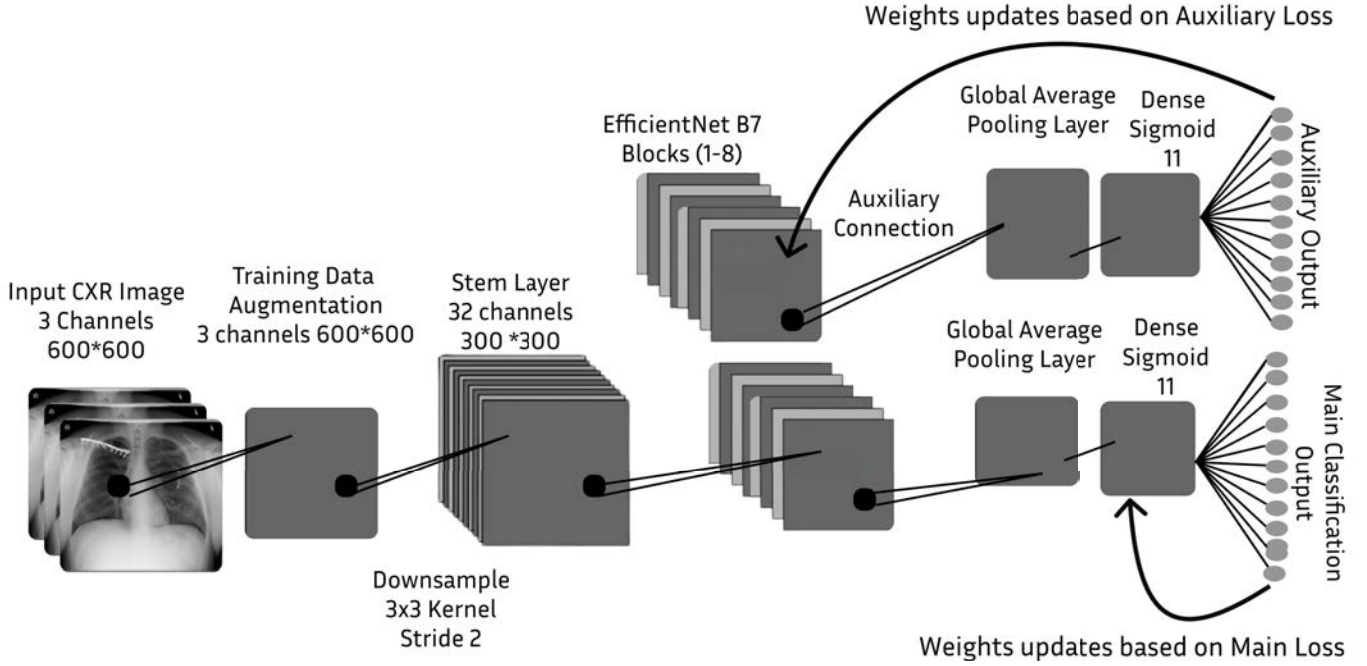


Fig. 3: The proposed Conv-MTD model, featuring an auxiliary connection head designed to detect and classify multi-label tubes. This auxiliary head is implemented by attaching a lightweight network to the fourth layer of EfficientNet-B7’s feature extraction process. The auxiliary head enhances gradient flow, improves feature learning, and stabilizes the overall training process.

for our specific use case. The main concept of compound scaling is to balance all the dimensions with a constant ratio. The constant ratio is determined by coefficients γ , α , and β . Resolution ($r: \gamma^\phi$), Depth ($d: \alpha^\phi$), Width ($w: \beta^\phi$).

$$f = \alpha \cdot \beta^\phi \cdot \gamma^\phi \quad (6)$$

$$f = d \cdot w^\phi \cdot r^\phi \quad (7)$$

The values of α and β γ are fixed by a grid search. The values of these coefficients were: $\alpha = 1.2$, $\beta = 1.1$, and $\gamma = 1.15$, the Constant values of α , β , and γ indicate that if the resolution of an image is increased by 15%, then the width of the model will be increased by 10%, and the depth will be increased by 20%. Depth coefficient ($\alpha = 1.2$) was chosen to allow sufficient network depth for capturing hierarchical features while avoiding the diminishing returns observed with deeper architectures. The value of ϕ may be changed to upscale the model depending on the available resources. The detailed training configuration hyper-parameters and architectural parameters are summarized in Table III and Table IV.

2) *Auxiliary Head*: An auxiliary head is added to the EfficientNet-B7 architecture to act as a regularizer and improve model stability by enhancing gradient flow during training, as shown in Fig. 3. This connection introduces an auxiliary prediction head, denoted as Auxiliary Head, attached to an intermediary layer 4 output, h_{aux} . By generating an additional loss term, \mathcal{L}_{aux} , this auxiliary head provides supplementary gradient flow through the entire network, addressing challenges with vanishing gradients in DNNs. The EfficientNet-

TABLE III: EfficientNet-B7 Configuration Parameters

Parameter	Value
Depth Coefficient (α)	1.2
Width Coefficient (β)	1.1
Resolution Coefficient (γ)	1.15
Dropout Rate	0.5
Number of Parameters	66M
Input Image Size	600 × 600
Model FLOPs	37B
Training Dataset	ImageNet
Number of Classes	1000

TABLE IV: EfficientNet-B7 Architectural Parameters

Stage	Operator	Resolution / Channels / Layers
1	Conv3x3	600 × 600 / 64 / 1
2	MBConv1, k3x3	300 × 300 / 32 / 3
3	MBConv6, k3x3	150 × 150 / 48 / 5
4	MBConv6, k5x5	75 × 75 / 80 / 7
5	MBConv6, k3x3	38 × 38 / 160 / 14
6	MBConv6, k5x5	19 × 19 / 224 / 18
7	MBConv6, k3x3	19 × 19 / 384 / 5
8	Conv1x1	19 × 19 / 1536 / 1
9	Pooling	1 × 1 / 1536 / 1
10	Fully Connected	1 × 1 / 1000 / 1

B7 model consists of 813 layers, formed by stacking multiple modules. While this deep structure enables complex feature learning, it complicates weight updates in deeper layers during backpropagation, which can result in diminished accuracy. The auxiliary head predicts the same classes as the main model output. The overall loss function, \mathcal{L}_{total} , combines the main

output loss $\mathcal{L}_{\text{main}}$ with the auxiliary loss \mathcal{L}_{aux} :

$$\mathcal{L}_{\text{total}} = \alpha \cdot \mathcal{L}_{\text{main}} + \beta \cdot \mathcal{L}_{\text{aux}} \quad (8)$$

where $\alpha = 0.9$ and $\beta = 0.1$ are weights that prioritize the main output loss while leveraging auxiliary feedback to reinforce training stability.

The auxiliary head itself is designed with dense and dropout layers to refine intermediate features and reduce overfitting. Let h_{aux} denote the feature representation from the intermediary layer. The auxiliary head computes the intermediate classification logits, z_{aux} , as follows:

$$z_{\text{aux}} = W_{\text{aux}} \cdot h_{\text{aux}} + b_{\text{aux}} \quad (9)$$

where W_{aux} and b_{aux} represent the weight matrix and bias vector of the auxiliary head, respectively. The final auxiliary prediction \hat{y}_{aux} is produced using a softmax activation function:

$$\hat{y}_{\text{aux}} = \text{softmax}(z_{\text{aux}}) \quad (10)$$

As a result, the model produces two predictions: one from the auxiliary head and one from the main EfficientNet-B7 output. The final classification relies on the main EfficientNet-B7 output, with the auxiliary head acting as a stabilizing regularizer to maintain robust gradient flow through the network.

D. Quantization

Quantization is the process of converting a real-valued number into a lower-precision format, typically an integer multiple of a base unit. This approach is commonly used to reduce model size, increase inference speed, and decrease memory requirements, which in turn enhances model efficiency and throughput.

In conventional deep learning models, mathematical operations are performed using 32-bit floating-point (FP32) numbers with millions of additions and multiplications during inference. This becomes computationally expensive and time-consuming. Quantization addresses this by using lower precision, such as 16-bit floating point (FP16) or 8-bit integers (INT8), to approximate the original FP32 weights and activations.

In this research, FP16 quantization is applied to the trained Conv-MTD model weights to ensure compatibility with low-powered edge devices. The quantization process can be mathematically represented as follows:

$$q(x) = \text{round}\left(\frac{x}{s}\right) \times s \quad (11)$$

where:

- x is the original FP32 value,
- s is the quantization scale factor, and
- $q(x)$ is the quantized representation of x .

For FP16 quantization, each 32-bit floating point value x is converted to a 16-bit representation, reducing memory usage by half. The error introduced by quantization can be minimized by selecting an optimal scale factor s , which can be defined based on the maximum absolute value of the weights or activations:

$$s = \frac{\max(|x|)}{2^{n-1} - 1} \quad (12)$$

where n is the number of bits used for quantization (e.g., 16 for FP16). Algorithm 1 shows that the details of the quantization process. By performing quantization, we not only reduce the model size but also increase inference speed, making the model more suitable for deployment on resource-constrained edge devices.

Algorithm 1 Quantization Algorithm for FP32 to FP16 Conversion

- 1: **Input:** Input values $x = \{x_1, x_2, \dots, x_m\}$, number of bits n
 - 2: **Output:** Quantized representation $q(x) = \{q(x_1), q(x_2), \dots, q(x_m)\}$
 - 3: **Step 1:** Calculate the maximum absolute value of the input set x :
 - 4: $x_{\text{max}} = \max(|x_1|, |x_2|, \dots, |x_m|)$
 - 5: **Step 2:** Compute the scale factor s based on the maximum value and number of bits:
 - 6: $s = \frac{x_{\text{max}}}{2^{n-1} - 1}$
 - 7: **for** $i = 1$ **to** m **do**
 - 8: Normalize x_i by dividing it by the scale factor s :
 - 9: $\hat{x}_i = \frac{x_i}{s}$
 - 10: Quantize \hat{x}_i by rounding to the nearest integer:
 - 11: $q(\hat{x}_i) = \text{round}(\hat{x}_i)$
 - 12: De-normalize $q(\hat{x}_i)$ back to the original scale by multiplying by s :
 - 13: $q(x_i) = q(\hat{x}_i) \times s$
 - 14: **end for**
 - 15: **Return** quantized values $q(x) = \{q(x_1), q(x_2), \dots, q(x_m)\}$
-

E. Experimentation Setup

All experiments used the TensorFlow framework and the Kaggle TPU hardware accelerator. The selection of TensorFlow is driven by its robust support for TPU acceleration and efficient handling of large datasets. Initial hyperparameter values are set based on recommendations in the literature [55], [56]. The choice of optimization strategy and associated hyperparameters are guided by extensive experimentation and theoretical considerations: Model optimization is achieved using the Adam optimizer, with an initial learning rate of 0.001. Adam is selected over other optimizers (SGD, RMSprop) due to its adaptive learning rate capabilities and superior convergence properties as demonstrated in our ablation studies Table IX. The learning rate gradually reduced to 5×10^{-4} by applying a decay factor of $\lambda = 0.5$ every two epochs. The decay schedule is empirically determined through experiments showing that more aggressive decay rates led to premature convergence, while slower decay resulted in training instability. The learning rate at epoch (n) is defined as:

$$\text{Learning Rate}(n) = \text{Learning Rate}(n - 2) \times \lambda \quad (13)$$

Focal loss is chosen to compute prediction errors, with default parameters ($\alpha = 0.25$ and $\gamma = 2$). These specific values were selected after conducting a grid search over $\alpha \in [0.15, 0.35]$ and $\gamma \in [1, 3]$. The chosen parameters provided optimal

handling of class imbalance in our medical tube dataset as demonstrated in Table XI. The focal loss function is defined as:

$$FL(p_t) = -\alpha(1 - p_t)^\gamma \log(p_t) \quad (14)$$

Initial training is set to 30 epochs. However, peak model performance is observed between the fifth and eighth epochs. The 30-epoch limit is selected based on our observation that training beyond this point showed no significant improvement in validation metrics while increasing the risk of overfitting. Model checkpoints are saved at each epoch to allow restoration to the weights associated with the lowest validation loss. Early stopping is also implemented to prevent overfitting by terminating training if validation loss does not improve after five consecutive epochs. The patience value of 5 epochs is selected as a balance between allowing sufficient time for escape from local minima while preventing unnecessary computational overhead. The early stopping criterion is defined as:

$$\text{Early Stop} = \begin{cases} \text{Stop} & \text{if } \text{val_loss}_i \geq \min(\text{val_loss}_{1:i-1}) \\ & \text{and } i \geq p \\ \text{Continue} & \text{otherwise} \end{cases} \quad (15)$$

where ($p = 5$) is the patience parameter.

The details of the model parameters are presented in Table V.

TABLE V: Classifier Parameters

Parameter	Value
Image input size	600x600
Training parameters	63,825,702
Batch Size	16
Epochs	30
Stopping patience	5
Reduce Learning Rate patience	2
Learning rate	5×10^{-4}
Optimizer	ADAM
Evaluation Metric	ROC AUC

1) Evaluation Metrics: The area under the ROC curve (AUC) is used to evaluate the performance of the trained model, calculated by averaging the AUC for each of the 11 labels. AUC measures the ability of the model to distinguish between classes, with higher values indicating better discrimination. For a single label, the AUC is calculated as the integral of the True Positive Rate (TPR) over the False Positive Rate (FPR) along the ROC curve:

$$\text{AUC} = \int_0^1 \text{TPR} d(\text{FPR}) \quad (16)$$

The definitions for TPR, TNR, and FPR are given as follows:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (17)$$

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (18)$$

$$\text{FPR} = 1 - \text{TNR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (19)$$

For multi-label classification, True Positive (TP) refers to the correct identification of misplaced tubes, and False Positive (FP) represents normally placed tubes that are incorrectly classified as misplaced. TPR and FPR enable the evaluation of classification accuracy across all labels. The final AUC score is calculated by averaging the individual AUCs across all labels, providing an overall measure of model performance.

III. RESULTS AND DISCUSSIONS

The applicability and feasibility of a Conv-MTD have extensively examined throughout this study. This section details the results of the proposed Conv-MTD, its applicability in the malposition medical tube detection task and its feasibility in a resource-constrained clinical setting.

A. Results Evaluation

To rigorously evaluate the proposed model, we employed a 5-fold cross-validation to ensure robust and unbiased assessment. The Conv-MTD demonstrated the best performance in the detection and classification of medical tubes using CXRs, achieving a mean AUC of 0.95 ± 0.0078 . This high accuracy highlights the model's potential as an effective tool in clinical settings, where rapid and accurate tube verification is crucial. The evaluation of the trained model has carried out in a Three-fold approach: 1) Base Evaluation, 2) Evaluation with data augmentation, and 3) Evaluation with the addition of an auxiliary head.

1) Baseline Performance: The initial evaluation of the Conv-MTD, utilizing the EfficientNet-B7 backbone, has resulted in a mean AUC of 0.92 ± 0.0024 . This robust performance has validated the feature extraction capabilities of the architecture, which helps in effective differentiation between tube classes.

2) Data Augmentation Impact: Data augmentation introduces variability in the training dataset, which enhances the model's generalization across diverse X-ray imaging conditions, such as varying brightness, angles, and patient anatomies. By incorporating data augmentation techniques, the model's performance has improved significantly from a mean AUC of 0.92 ± 0.0024 to 0.93 ± 0.0037 . Fig. 4 (a) and Fig. 4(b) illustrate the learning curves for models trained with and without augmentation. The augmented model Fig. 4 (a) tended to overfit the training data, and 4 (b) demonstrated smoother convergence, suggesting that data augmentation contributed to improved generalization.

3) Auxiliary Head Contribution: The proposed Conv-MTD has further enhanced by adding auxiliary head connected to hidden layers to improve the training stability. Adding an auxiliary head aids in mitigating vanishing gradient issues in the base network, enabling enhanced feature learning. This architecture refinement improved the classification accuracy for subtle distinctions in tube positioning resulting in improved learning to capture nuanced differences among normal, borderline, and abnormal categories. This improvement is evident in the model's

TABLE VI: Performance Comparison of Classifier Models in Terms of AUC, Hardware Support, and Computational Efficiency. The table presents the model size, mean AUC with standard deviation (STD), supported hardware platforms, and the number of floating-point operations per second (Flops).

Model	Model Size	Mean AUC \pm STD	Supported Hardware	Flops
Base EfficientNet-B7	244 MB	0.92 \pm 0.0024	Only GPU supported	32 bits
Conv-MTD without Augmentation	244 MB	0.93 \pm 0.0037	Only GPU supported	32 bits
Conv-MTD with Augmentation	244 MB	0.95 \pm 0.0080	Only GPU supported	32 bits
FP16 based Conv-MTD	122 MB	0.95 \pm 0.0078	GPU, CPU supported	16 bits

TABLE VII: Comparison of the proposed Conv-MTD methodology with existing techniques. The proposed Conv-MTD approach demonstrates a balanced performance with an AUC of 0.95 on a large dataset (30,083 CXRs) while supporting multi-class detection (ETT, NGT, CVC, SGC) and edge device deployment using 16-bit quantization. Although some existing methods achieve slightly higher AUC (e.g., Cascaded CNNs [37], AUC: 0.99), they are limited to single-tube detection and lack support for multi-class scenarios and edge devices.

Methods	Dataset Size	Target Tubes	Multi-class	Edge Device Support	Results	Notes
Neural network ensemble [30]	7,081 CXRs	NGT only	No	No	AUC: 0.86	Single tube detection
DenseNet-121 CNN [40]	4,693 CXRs	NGT only	No	No	AUC: 0.92	Limited dataset
EfficientNet B0 with Mask R-CNN [36]	1,985 CXRs	ETT only	No	No	F1: 0.88	Small dataset
Cascaded CNNs [37]	16,000 CXRs	ETT only	No	No	AUC: 0.99	Single tube focus
Weakly-supervised CNN [39]	175 CXRs	NGT only	No	No	AUC: 0.76	Very limited data
DeepLabv3+ResNeSt50 [44]	7,378 CXRs	NGT only	No	No	AUC: 0.96	Single modality
ResNet-50 CNN [35]	777 CXRs	ETT,NGT,CVC	Yes	No	AP: 0.97*	Extremely small dataset
ResNet50V2 DCNN [33]	30,083 CXRs	ETT,NGT,CVC,SGC	Yes	No	AUC: 0.80	No quantization
EfficientNet + Segmentation masks [45]	30,083 CXRs	ETT,NGT,CVC,SGC	Yes	No	AUC: 0.96	No edge device support
Proposed Conv-MTD	30,083 CXRs	CVC, NGT, ETT, SGC	Yes	Yes (16-bit)	AUC: 0.95	Edge Device Support, FP16 bit computation

*AP (Average Precision) is not directly comparable with AUC metrics

overall performance, as shown in Table VI. The enhanced model achieved a mean AUC of 0.95 ± 0.0080 , outperforming both the baseline and augmented variants, which highlight the effectiveness of these architectural adjustments in addressing the complexities of multilabel classification tasks.

4) *Baseline Comparison:* To establish the incremental benefits of Conv-MTD, We conducted comprehensive experiments with several baseline architectures. The evaluation included ResNet152v2, Inception V3, and Xception, As shown in Table VIII, ResNet152v2, despite its substantial parameter count 58.2M, achieved a baseline AUC of 0.88. Inception V3, with a more efficient architecture of 21.7M parameters, performed better with AUC of 0.92. Xception, utilizing 20.8M parameters, demonstrated strong performance with AUCs of 0.93. The proposed Conv-MTD, while having the largest parameter count 63.8M, justified its complexity by achieving superior performance with AUCs of 0.95. These comparisons demonstrate that the architectural choices in Conv-MTD contribute to meaningful performance improvements over simpler approaches, with a consistent 2-9 percentage point advantage in AUC over baseline models.

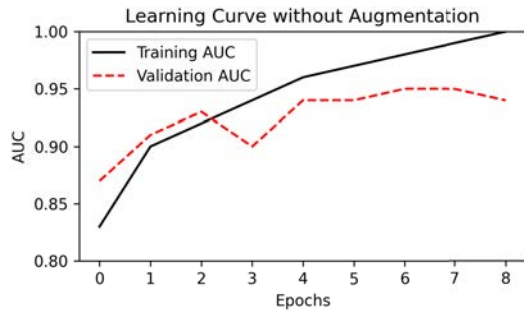
TABLE VIII: Comparison with Baseline Models

Model	Parameters	AUC
ResNet152v2	58.2M	0.88
Inception V3	21.7M	0.92
Xception	20.8M	0.93
Conv-MTD (Proposed)	63.8M	0.95

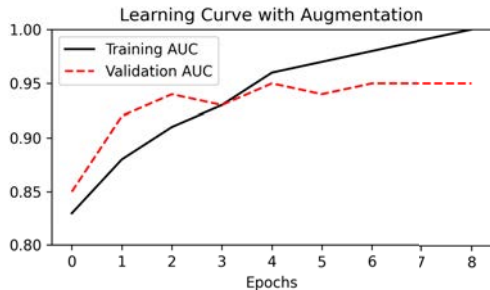
The confusion matrix of the proposed Conv-Mtd, illustrated in Fig. 5, shows detailed insights into the model’s classification

behaviour across various tube types and conditions. The model exhibits strong performance in identifying normal cases, with 1386 correct classifications for ETT and 2266 for CVC. However, challenges arise in distinguishing borderline cases from normal ones, particularly in CVC classification, where 246 borderline cases are misclassified as normal. Minimal confusion across tube types is observed, as most misclassifications occur within the same tube type but under different conditions. This demonstrates that while the model effectively differentiates between tube types, there is a need for improvement in distinguishing subtle variations within each tube category, especially for borderline cases. In NGT classification, the model performs well for normal cases, achieving 142 correct classifications, but shows limitations in classifying incompletely imaged cases, with only 43 correct classifications and several misclassifications across other categories. These patterns suggest that the model’s performance is influenced by the completeness and quality of the imaging, particularly in NGT cases. Furthermore, the confusion patterns underscore the impact of class imbalance, with the model showing stronger performance in categories with a higher abundance of training examples, particularly in normal cases across all tube types.

The multilabel ROC curves in Fig. 6 further illustrate the model’s performance across various tube classifications. Fig. 6(a) highlights the performance for Endotracheal Tube (ETT) and Nasogastric Tube (NGT) classifications, achieving AUC values of 0.95 for ETT-Abnormal, 0.955 for ETT-Borderline, 0.99 for ETT-Normal, and 0.96 for NGT-Abnormal. These results reflect the model’s strong ability to differentiate between normal and abnormal states while maintaining high sensitivity in borderline cases. Similarly, Fig. 6(b) focuses on additional NGT classifications and Central Venous Catheter



(a)



(b)

Fig. 4: (a) Train and validation learning curves without augmentation. (b) Train and validation learning curves with augmentation.

(CVC), showing robust performance with AUC values of 0.97 for NGT-Borderline, 0.98 for NGT-Incompletely Imaged, 0.98 for NGT-Normal, and 0.90 for CVC-Abnormal.

Finally, Fig. 6(c) emphasizes the performance for CVC and Swan Ganz Catheter classifications. The model maintained an AUC of 0.84 for CVC-Borderline, 0.90 for CVC-Normal, and achieved a perfect AUC of 1.00 for Swan Ganz Catheter Present, highlighting exceptional precision in identifying this specific category. Across all categories, the AUC values indicate the model’s effectiveness in handling subtle distinctions, even in borderline and incomplete cases, which are inherently more challenging to classify.

B. Weights Quantization

The quantized Conv-MTD model maintained a competitive AUC of 0.95 ± 0.0080 while achieving a 50% reduction in model size and offering a 2x speedup. Our analysis of FP16 precision impact revealed varying effects across different tube placement scenarios. For common tube placements (CVC-Normal), FP16 maintained a similar performance to FP32 with an AUC difference < 0.01 . However, for rare cases like NGT-Abnormal and Swan Ganz Catheter, FP16 showed slight performance degradation in AUC decrease of 0.02-0.03, which is within acceptable clinical margins. The negligible performance impact can be attributed to the inherent resilience of the EfficientNet-B7 architecture to reduced numerical precision. Our experiments showed that the quantized model requires only 8GB of RAM compared to the 16GB needed for the FP32 model, making it suitable for deployment on edge devices commonly available in clinical settings. The results of

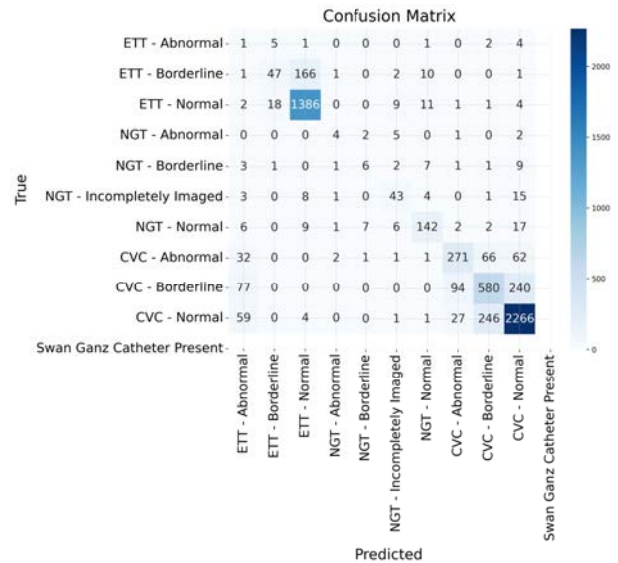


Fig. 5: Confusion matrix showing the classification results for the multi-label medical tube classification task, illustrating the frequency of true positive, false positive, true negative, and false negative predictions for each category of tube class with confidence threshold > 75 .

optimized Conv-MTD have presented in Table VI.

C. Scalability and Limitations

The Conv-MTD architecture demonstrates potential for scaling to larger systems through its modular design and efficient resource utilization. The model’s architecture can be adapted to handle additional tube types by modifying the output layers while maintaining the core feature extraction capabilities. However, scaling to larger systems would require careful consideration of computational resources and potential trade-offs between model complexity and inference speed. Several limitations of the current implementation warrant discussion. First, while Conv-MTD performs well on CXR, its application to other imaging modalities such as CT or MRI would require significant architectural modifications to handle 3D data and different image characteristics. Second, the model’s performance may be impacted when dealing with novel tube types not represented in the training data, particularly those with unique positioning requirements or visual characteristics. Third, the current implementation is optimized for specific hardware configurations, and deployment on different platforms may require additional optimization work.

D. Ablation study

We conducted an ablation study comprising five experiments with different optimizers and loss functions, and detailed results of the ablation study are shown in Table IX, and Table XI. The Adam optimizer with a learning rate of 0.001 provided the highest AUC of 0.95, outperforming AdaGrad and RMSProp, suggesting that Adam is particularly effective given the data characteristics and network architecture.

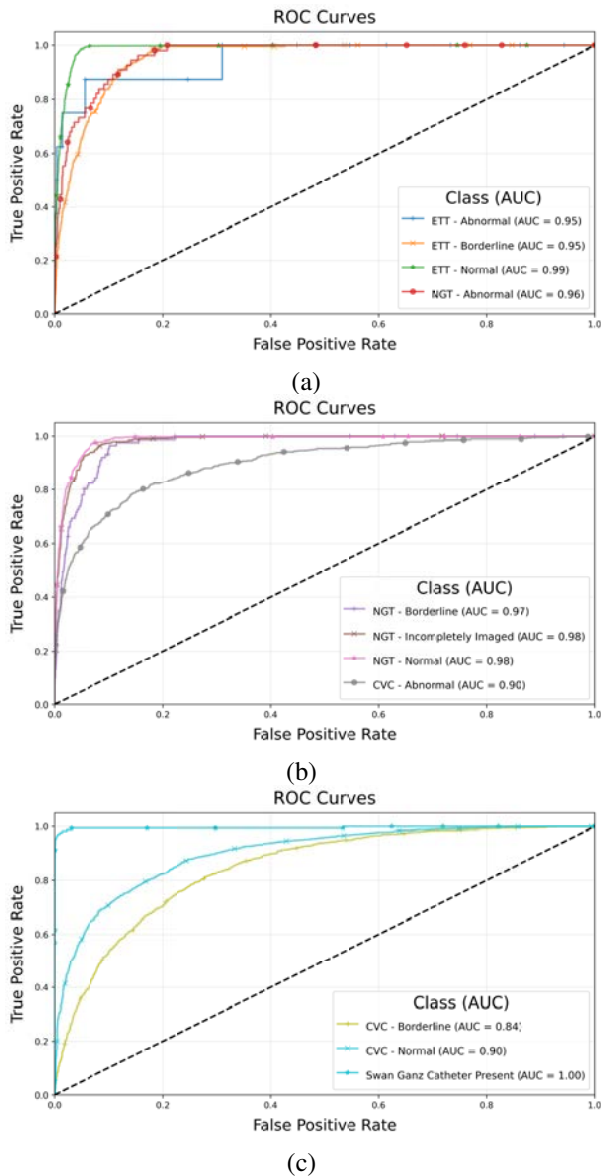


Fig. 6: Receiver Operating Characteristic (ROC) curves for the multi-label medical tube classification task, illustrating the performance of the model in distinguishing different classes of tubes.

We further conducted architectural ablation studies to quantify the impact of key model components. Table X presents these results. The base model without auxiliary head achieved an AUC of 0.92, while adding the auxiliary head improved performance to 0.95, demonstrating its effectiveness in enhancing feature learning. The auxiliary head’s contribution particularly notable in improving the detection of borderline cases, where the AUC increased from 0.90 to 0.95. We also evaluated the impact of quantization, and deployment feasibility. Initially, the model is trained on TPU for optimal training efficiency. Table X presents the results across different configurations and hardware settings. The 16-bit quantized model reduced size by 50% (from 244.64 to 122.32) with only a minimal performance drop of 0.01. We evaluated the quantized model’s practical deployment metrics across different resource-constrained en-

TABLE IX: Ablation study of Proposed Model on RANZCR CLIP dataset with varying optimizers

Optimizer	Adam	AdaGrad	RMSProp
LR	0.001	0.001	0.001
AUC	0.95	0.82	0.83

vironments on 32GB RAM systems, it achieved an average inference time of 48ms per image on 32GB systems, 50ms on 16GB and on 8GB systems 55ms. These results demonstrate the model’s ability to maintain clinically viable performance even on hardware configurations typical of point-of-care devices.

TABLE X: Model Performance and Deployment Metrics Across Different Configurations

Configuration	Parameters Size	AUC	Inference Time (ms)
Base Model (32GB RAM)	244.63	0.92	86
With Auxiliary Head (32GB RAM)	244.64	0.95	88
Quantized Model (32GB RAM)	122.32	0.949	48
Quantized Model (16GB RAM)	122.32	0.949	50
Quantized Model (8GB RAM)	122.32	0.949	55

Table XI presents the model’s performance using Focal Loss and Binary Cross-Entropy (BCE). The effectiveness of Focal Loss is particularly evident in handling minority classes. For NGT-Abnormal cases, Focal Loss improved the detection AUC from 0.90 (with BCE) to 0.96. For ETT-Abnormal, Swan Ganz, and CVC-Abnormal, the AUCs improved from 0.91 to 0.95, 0.95 to 0.99, and 0.86 to 0.90, respectively. This improvement stems from Focal Loss’s ability to assign higher weights to minority class samples during training.

TABLE XI: Class-wise Performance Comparison: BCE vs Focal Loss

Tube Class	BCE (AUC)	Focal Loss (AUC)
NGT - Abnormal	0.90	0.96
ETT - Abnormal	0.91	0.95
Swan Ganz	0.95	0.99
CVC - AbNormal	0.86	0.90

Furthermore, it is crucial to acknowledge a limitation in our methodology arising from inconsistent CVC-Normal and CVC-Abnormal, NGT-Normal and NGT-Abnormal labels for same class applied by radiologists, which is impossible. Future work will focus on the integration of multi-modal AI approaches by combining images from different modalities, such as CT scans and MRI scans, to enhance model performance. In summary, the Conv-MTD model successfully addresses several challenges in medical tube detection and classification. Its high AUC score, robustness against class imbalance, and compatibility with edge devices underscore its potential as a valuable tool in clinical diagnostics. While certain limitations exist in terms of scalability and cross-modality applications, the identified future directions provide a clear pathway for advancing this technology toward more comprehensive clin-

ical solutions. Future studies could build on this work by incorporating more diverse datasets and exploring multi-modal capabilities, ultimately aiming to further enhance patient care in high-demand healthcare settings.

IV. CONCLUSION

In this study, we developed and evaluated Conv-MTD, a deep learning-based model for the detection and classification of medical tube placements in CXRs. Conv-MTD demonstrated robust feature extraction, achieving high accuracy with an AUC of 0.95. The auxiliary head effectively mitigated the vanishing gradient problem in the base EfficientNet-B7 architecture, ensuring stable and efficient training. To enhance the model's applicability in real-world healthcare settings, we implemented a quantization process that reduced memory and computational requirements through FP16 quantization. This optimization allows Conv-MTD to operate efficiently on resource-limited edge devices, making it suitable for real-time and point-of-care diagnostics. The ability to deploy Conv-MTD on such devices significantly expands its potential impact, enabling timely and accurate medical tube placement assessments in diverse clinical environments. For future work, we aim to explore multi-modal data integration by combining CXR images with corresponding doctor diagnostic reports, which could provide additional context for more precise decision-making. We also plan to enhance the model's capabilities by integrating a functionality to calculate the precise distance from the ideal tube placement position in centimeters, offering actionable insights for clinical interventions. Furthermore, we intend to address dataset limitations by incorporating a larger and more diverse dataset that includes rare tube misplacements scenarios and variations across different imaging modalities. These advancements will improve the robustness, accuracy, and generalizability of Conv-MTD, ensuring broader applicability in varied healthcare settings.

REFERENCES

- [1] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical image analysis using convolutional neural networks: a review," *Journal of medical systems*, vol. 42, pp. 1–13, 2018.
- [2] M. A. Abdou, "Literature review: Efficient deep neural networks techniques for medical image analysis," *Neural Computing and Applications*, vol. 34, no. 8, pp. 5791–5812, 2022.
- [3] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 4–21, 2016.
- [4] G. Papanastasiou, N. Dikaios, J. Huang, C. Wang, and G. Yang, "Is attention all you need in medical image analysis? a review," *IEEE Journal of Biomedical and Health Informatics*, 2023.
- [5] Z. Guo, Y. Shen, S. Wan, W.-L. Shang, and K. Yu, "Hybrid intelligence-driven medical image recognition for remote patient diagnosis in internet of medical things," *IEEE journal of biomedical and health informatics*, vol. 26, no. 12, pp. 5817–5828, 2021.
- [6] A. K. Sahu, S. Bhoi, P. Aggarwal, R. Mathew, J. Nayer, P. R. Mishra, T. P. Sinha *et al.*, "Endotracheal tube placement confirmation by ultrasonography: A systematic review and meta-analysis of more than 2500 patients," *The Journal of Emergency Medicine*, vol. 59, no. 2, pp. 254–264, 2020.
- [7] K. A. Miller, A. Kimia, M. C. Monuteaux, and J. Nagler, "Factors associated with misplaced endotracheal tubes during intubation in pediatric patients," *The Journal of Emergency Medicine*, vol. 51, no. 1, pp. 9–18, 2016.
- [8] M. Tuna, R. Latifi, A. El-Menyar, and H. Al Thani, "Gastrointestinal tract access for enteral nutrition in critically ill and trauma patients: indications, techniques, and complications," *European Journal of Trauma and Emergency Surgery*, vol. 39, pp. 235–242, 2013.
- [9] S. Nayak, M. Sherchan, S. D. Paudel, J. C. Rai, R. Shrestha, B. Shrestha, S. Shrestha, and P. Bhattacharyya, "Assessing placement of nasoduodenal tube and its usefulness in maintaining nutrition in critically ill patients," *Nepal Med Coll J*, vol. 10, no. 4, pp. 249–253, 2008.
- [10] D. A. Raptis, K. Neal, and S. Balla, "Imaging approach to misplaced central venous catheters," *Radiologic Clinics*, vol. 58, no. 1, pp. 105–117, 2020.
- [11] G. Aydin and Z. Akcaboy, "Extracavally malpositioned central venous catheter," *Journal of the College of Physicians and Surgeons Pakistan*, vol. 30, no. 4, pp. 459–461, 2020.
- [12] Y. Chen, E. Shlofmitz, N. Khalid, N. L. Bernardo, I. Ben-Dor, W. S. Weintraub, and R. Waksman, "Right heart catheterization-related complications: a review of the literature and best practices," *Cardiology in Review*, vol. 28, no. 1, pp. 36–41, 2020.
- [13] H. R. Gilbertson, E. J. Rogers, and O. C. Ukoumunne, "Determination of a practical ph cutoff level for reliable confirmation of nasogastric tube placement," *Journal of Parenteral and Enteral Nutrition*, vol. 35, no. 4, pp. 540–544, 2011.
- [14] A. Stock, H. Gilbertson, and F. E. Babl, "Confirming nasogastric tube position in the emergency department: ph testing is reliable," *Pediatric emergency care*, vol. 24, no. 12, pp. 805–809, 2008.
- [15] J. Li, "Capnography alone is imperfect for endotracheal tube placement confirmation during emergency intubation," *The Journal of emergency medicine*, vol. 20, no. 3, pp. 223–229, 2001.
- [16] K. Glen, C. E. Weekes, M. Banks, I. Arbi, and M. Hannan-Jones, "A prospective observational study of ph testing to confirm ongoing nasogastric tube position," *Journal of Clinical Nursing*, 2024.
- [17] C. Cumming and J. McFADZEAN, "A survey of the use of capnography for the confirmation of correct placement of tracheal tubes in pediatric intensive care units in the uk," *Pediatric Anesthesia*, vol. 15, no. 7, pp. 591–596, 2005.
- [18] C. Sitzwohl, A. Langheinrich, A. Schober, P. Krafft, D. I. Sessler, H. Herkner, C. Gonano, C. Weinstabl, and S. C. Kettner, "Endobronchial intubation detected by insertion depth of endotracheal tube, bilateral auscultation, or observation of chest movements: randomised trial," *Bmj*, vol. 341, 2010.
- [19] T. Nejo, S. Oya, T. Tsukasa, N. Yamaguchi, and T. Matsui, "Limitations of routine verification of nasogastric tube insertion using x-ray and auscultation: two case reports of life-threatening complications," *Nutrition in Clinical Practice*, vol. 31, no. 6, pp. 780–784, 2016.
- [20] S. Taylor and A. R. Manara, "X-ray checks of ng tube position: a case for guided tube placement," *The British Journal of Radiology*, vol. 94, no. 1124, p. 20210432, 2021.
- [21] S. J. Taylor, T. Karpasiti, and D. Milne, "Safety of blind versus guided feeding tube placement: misplacement and pneumothorax risk," *Intensive and Critical Care Nursing*, vol. 76, p. 103387, 2023.
- [22] A. S. Ebenezer, S. D. Kanmani, M. Sivakumar, and S. J. Priya, "Effect of image transformation on efficientnet model for covid-19 ct image classification," *Materials Today: Proceedings*, vol. 51, pp. 2512–2519, 2022.
- [23] B. Kim, T. S. Mathai, K. Helm, P. A. Pinto, and R. M. Summers, "Classification of multi-parametric body mri series using deep learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 11, pp. 6791–6802, 2024.
- [24] Y. Cheng, Q. Guo, F. Juefei-Xu, H. Fu, S.-W. Lin, and W. Lin, "Adversarial exposure attack on diabetic retinopathy imagery grading," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–13, 2024.
- [25] W. Dai, R. Liu, T. Wu, M. Wang, J. Yin, and J. Liu, "Deeply supervised skin lesions diagnosis with stage and branch attention," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 2, pp. 719–729, 2024.
- [26] G. Marques, D. Agarwal, and I. De la Torre Díez, "Automated

- medical diagnosis of covid-19 through efficientnet convolutional neural network,” *Applied soft computing*, vol. 96, p. 106691, 2020.
- [27] C. Huang, W. Wang, X. Zhang, S.-H. Wang, and Y.-D. Zhang, “Tuberculosis diagnosis using deep transferred efficientnet,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 5, pp. 2639–2646, 2022.
 - [28] A. Rafay and W. Hussain, “Efficientnet-based classification model for a large manually curated dataset of 31 skin diseases,” *Biomedical Signal Processing and Control*, vol. 85, p. 104869, 2023.
 - [29] M. B. Hossain, S. H. S. Iqbal, M. M. Islam, M. N. Akhtar, and I. H. Sarker, “Transfer learning with fine-tuned deep cnn resnet50 model for classifying covid-19 from chest x-ray images,” *Informatics in Medicine Unlocked*, vol. 30, p. 100916, 2022.
 - [30] I. Drozhdov, R. Dixon, B. Szubert, J. Dunn, D. Green, N. Hall, A. Shirandami, S. Rosas, R. Grech, S. Puttagunta *et al.*, “An artificial neural network for nasogastric tube position decision support,” *Radiology: Artificial Intelligence*, vol. 5, no. 2, p. e220165, 2023.
 - [31] M. Abbas, A. A. Salam, and J. Zeb, “Automatic detection and classification of correct placement of tubes on chest x-rays using deep learning with efficientnet,” in *2022 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2)*. IEEE, 2022, pp. 1–6.
 - [32] M. Aryal and N. Yahyasoltani, “Identifying catheter and line position in chest x-rays using gans,” in *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2021, pp. 122–127.
 - [33] A. Elaanba, M. Ridouani, and L. Hassouni, “Automatic detection using deep convolutional neural networks for 11 abnormal positioning of tubes and catheters in chest x-ray images,” in *2021 IEEE World AI IoT Congress (AIoT)*. IEEE, 2021, pp. 0007–0012.
 - [34] W. B. Geftter and H. Hatabu, “Reducing errors resulting from commonly missed chest radiography findings,” *Chest*, vol. 163, no. 3, pp. 634–649, 2023.
 - [35] R. D. Henderson, X. Yi, S. J. Adams, and P. Babyn, “Automatic detection and classification of multiple catheters in neonatal radiographs with deep learning,” *Journal of digital imaging*, vol. 34, no. 4, pp. 888–897, 2021.
 - [36] H. C. Jung, C. Kim, J. Oh, T. H. Kim, B. Kim, J. Lee, J. H. Chung, H. Byun, M. S. Yoon, and D. K. Lee, “Position classification of the endotracheal tube with automatic segmentation of the trachea and the tube on plain chest radiography using deep convolutional neural network,” *Journal of Personalized Medicine*, vol. 12, no. 9, p. 1363, 2022.
 - [37] S. Kara, J. Y. Akers, and P. D. Chang, “Identification and localization of endotracheal tube on chest radiographs using a cascaded convolutional neural network approach,” *Journal of Digital Imaging*, vol. 34, pp. 898–904, 2021.
 - [38] P. Lakhani, A. Flanders, and R. Gorniak, “Endotracheal tube position assessment on chest radiographs using deep learning,” *Radiology: Artificial Intelligence*, vol. 3, no. 1, p. e200026, 2020.
 - [39] G. Liang, H. Ganesh, D. Steffe, L. Liu, N. Jacobs, and J. Zhang, “Development of cnn models for the enteral feeding tube positioning assessment on a small scale data set,” *BMC Medical Imaging*, vol. 22, no. 1, p. 52, 2022.
 - [40] D. Mallon, C. McNamara, G. Rahmani, D. O’Regan, and D. Amiras, “Automated detection of enteric tubes misplaced in the respiratory tract on chest radiographs using deep learning with two centre validation,” *Clinical Radiology*, vol. 77, no. 10, pp. e758–e764, 2022.
 - [41] I. Park, H.-S. Choi, G. Moon, J. Y. Hong, J. Heo, H. Ko, D. Lee, Y. Kim, W. J. Kim, and K. M. Moong, “Deep learning-based dual-stage model for accurate nasogastric tube positioning in chest radiographs,” *Available at SSRN 4965848*.
 - [42] J. Rueckel, C. Huemmer, C. Shahidi, G. Buizza, B. F. Hoppe, T. Liebig, J. Ricke, J. Rudolph, and B. O. Sabel, “Artificial intelligence to assess tracheal tubes and central venous catheters in chest radiographs using an algorithmic approach with adjustable positioning definitions,” *Investigative Radiology*, vol. 59, no. 4, pp. 306–313, 2024.
 - [43] J. Stroeder, M. Multusch, L. Berkel, L. Hansen, A. Saalbach, H. Schulz, M. P. Heinrich, Y. Elser, J. Barkhausen, and M. M. Sieren, “Optimizing catheter verification: An understandable ai model for efficient assessment of central venous catheter placement in chest radiography,” *Investigative Radiology*, pp. 10–1097, 2023.
 - [44] C.-H. Wang, T. Hwang, Y.-S. Huang, J. Tay, C.-Y. Wu, M.-C. Wu, H. R. Roth, D. Yang, C. Zhao, W. Wang *et al.*, “Deep learning-based localization and detection of malpositioned nasogastric tubes on portable supine chest x-rays in intensive care and emergency medicine: A multi-center retrospective study,” *Journal of Imaging Informatics in Medicine*, pp. 1–11, 2024.
 - [45] P. Wongveerasin, T. Tongdee, and P. Saiviroonporn, “Deep learning for tubes and lines detection in critical illness: Generalizability and comparison with residents,” *European Journal of Radiology Open*, vol. 13, p. 100593, 2024.
 - [46] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.
 - [47] J. S. Tang, J. C. Seah, A. Zia, J. Gajera, R. N. Schlegel, A. J. Wong, D. Gai, S. Su, T. Bose, M. L. Kok *et al.*, “Clip, catheter and line position dataset,” *Scientific Data*, vol. 8, no. 1, p. 285, 2021.
 - [48] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.
 - [49] A. Kebaili, J. Lapuyade-Lahorgue, and S. Ruan, “Deep learning approaches for data augmentation in medical imaging: a review,” *Journal of Imaging*, vol. 9, no. 4, p. 81, 2023.
 - [50] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, “A review of medical image data augmentation techniques for deep learning applications,” *Journal of Medical Imaging and Radiation Oncology*, vol. 65, no. 5, pp. 545–563, 2021.
 - [51] E. Goceri, “Medical image data augmentation: techniques, comparisons and interpretations,” *Artificial Intelligence Review*, vol. 56, no. 11, pp. 12561–12605, 2023.
 - [52] J. Mukhoti, V. Kulharia, A. Sanyal, S. Golodetz, P. Torr, and P. Dokania, “Calibrating deep neural networks using focal loss,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 15288–15299, 2020.
 - [53] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, “Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation,” *Computerized Medical Imaging and Graphics*, vol. 95, p. 102026, 2022.
 - [54] M. Mulyanto, M. Faisal, S. W. Prakosa, and J.-S. Leu, “Effectiveness of focal loss for minority classification in network intrusion detection systems,” *Symmetry*, vol. 13, no. 1, p. 4, 2020.
 - [55] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.
 - [56] Y. Bengio, “Practical recommendations for gradient-based training of deep architectures,” *Neural Networks: Tricks of the Trade: Second Edition*, pp. 437–478, 2012.