ISSN ONLINE: 3102-0712
www.divyagyan.edu.np

DIVYA GYAN JOURNAL OF BUSINESS AND COMPUTING

Volume 1 | Issue 1

RESEARCH MANAGEMENT CELL
DIVYA GYAN COLLEGE

About the Journal:
Divya Gyan Journal of Business and Computing

Divya Gyan College has reached a milestone in scholarly excellence with the launch of the Divya Gyan Journal of Business and Computing (DGJBC). As a premier interdisciplinary platform, DGJBC is dedicated to the dissemination of high-quality research that explores the dynamic intersection of Management, Information Technology, and Computer Science. Published bi-annually in English, the journal is released in a modern online format, ensuring a global reach that extends far beyond the college walls to the wider academic and professional community in Nepal and abroad.

With a firm commitment to fostering a culture of rigorous inquiry, Divya Gyan Journal of Business and Computing encourages faculty, independent researchers, and industry experts to contribute original scholarly articles that bridge the gap between theoretical frameworks and practical applications. The journal serves as a catalyst for innovation, focusing on how computational advancements and data-driven strategies are reshaping modern business landscapes. By embracing a meticulous peer-review process, Divya Gyan Journal of Business and Computing upholds the highest standards of academic integrity and credibility, ensuring that every published piece provides significant value to its respective field. As Divya Gyan College continues to champion research-driven education, the Divya Gyan Journal of Business and Computing stands poised to be a cornerstone of innovation, providing a fertile ground for the ideas that will define the future of business and technology.

Vision:

Divya Gyan College aspires to be a leading educational institution advancing research, innovation, and industry collaboration. We envision an academic ecosystem where students sustain their education through real-world research and development projects, turning learning into meaningful contribution and opportunity.

Mission:

- To provide high-quality, industry-relevant education in technology and management, equipping students with solid core computing, software development skills, and up-to-date knowledge of the latest technologies.
- To integrate technology with management education to develop practical skills for effective decision-making and operations in today's digital world.
- To cultivate a culture of research, innovation, and lifelong learning among students and faculty.
- To develop state-of-the-art infrastructure and a vibrant academic environment.
- To build strong linkages with industry and community, preparing students to meet industry challenges and enhancing learning and career opportunities.

Goals:

1. To develop and upgrade physical and IT infrastructure to support program expansion and academic excellence.
2. To achieve academic excellence through quality improvement and the introduction of new, relevant programs.
3. To promote a robust culture of research, development, and innovation.
4. To enhance the overall student experience through comprehensive welfare programs and extracurricular activities.
5. To attract, develop, and retain highly qualified and motivated human resources.
6. To strengthen partnerships with industry, alumni, and the community.

Publisher
Research Management Cell, Divya Gyan College, Affiliated to Tribhuvan University
Kamaladi Mode, Kathmandu
Contact: +977 5346070, 9801169141
Email: info@divyagyan.edu.np

Disclaimers:

The opinions expressed in the articles and research papers published in this journal are those of the authors and do not reflect the official policy or position of Divya Gyan College or the Divya Gyan Journal of Business and Computing. The college shall not be held responsible for any errors, omissions, or consequences arising from the use of the information contained herein.

Table of Contents

Performance Analysis of InceptionV3-RCNN and CGAN Models for Image Forgery Detection

Babita Rimal[1], Balkrishna Subedi[2], Sagar Timalsena[3]

[1]Faculty Member, Divya Gyan College, Tribhuvan University

[2]Asst. Professor, Central Department of Computer Science and Information Technology, Tribhuvan University

[3]Department of Computer Science, Banaras Hindu University

Article History

Received: 08/12/2025

Revised: 11/01/2026

Accepted: 05/01/2026

Published: 15/01/2026

Corresponding Author

Name: Babita Rimal

Email: babitarimal@divyagyan.edu.np

ORCiD: https://orcid.org/0009-0007-4885-1137

Abstract

Due to the usage of advanced editing software and rising of AI-generated content, there seems to be the great problem with the forgery detection. This paper is a comparison of two deep learning models InceptionV3-RCNN and a Conditional GAN (CGAN) using CASIA v2.0 dataset which is benchmark for forgery detection. The result concluded that InceptionV3-RCNN network was highly effective on the augmented dataset with an accuracy of 98.61 while CGAN model had an accuracy of 76.52%. The training time for InceptionV3-RCNN is less and inferred time is more, whereas CGAN has taken more training time and less inferred time. On the whole, the two models bring out different strengths. The combined approach of InceptionV3-RCNN and CGAN be useful to improve the effectiveness of image forgery detection which could be explored in future work. As the results suggest, by integrating the strong classification properties of InceptionV3-RCNN and localization fine-grained properties of CGAN the combined features can be utilized.

Keywords: ***Image Forgery Detection, InceptionV3-RCNN, Conditional GAN, Deep Learning, Digital Forensics.***

Introduction

The use of advanced digital image editing tools, which are powered by AI has mostly used in the manipulation of visual content. In most cases, the manipulations can be highly challenging to get identified by the detection system regardless of whether the manipulation is splicing, copy-move, or deepfakes. Significant issues can be seen due to Forged images in the field of journalism, the law enforcement, medicine, scientific studies, and even social media. The traditional detection methods are insufficient to detect the subtle or highly sophisticated forgeries that make use of metadata or manual features. Due to this, a reliable system for detection of forgery has become an need in digital image forensics(Ross et al., 2020).

The recent advancement in deep learning have provided new in opportunities to ove4rcome the challenges pose in the digital forensics. Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) models have proven to be the powerful tools. They can directly acquire meaningful patterns and spatial inconsistencies as dictated by the image data. Two of these models, namely 1) InceptionV3-RCNN, a multi-scale feature extraction-based architecture which utilizes adversarial learning to produce tampering masks and classify images whether it is authentic or manipulated; and (2) Conditional GAN, a model that employs adversarial learning to produce tampering masks and classify image as authentic or manipulated. InceptionV3-RCNN is a multiscale feature extraction-based system that integrates region-based detection and this allows it to classify and localize the areas of manipulation with great efficiency. Conditional GANs, by comparison, follow adversarial learning, which designs pixel-level tampering masks, but at the expense of more intricate training. These complementary features encourage a comparative analysis of the two architectures to detect forgery of images.

This work is aimed at comparing these two models by their classification accuracy and computational efficiency and the overall capability to help in the image forgery detection.

Literature Review

Image forgeries detection methods has developed widely due to the use of deep learning approaches as opposed to the traditional feature-based technique. The initial copy moves and splicing detection systems were based on block-based extracting features or key point-based. They performed well on simple manipulations. When subsequent post-processing functions such as compression, rotation, and noise were used then these methods were not so effective.

So, using convolutional neural networks (CNNs), helped learning discriminative features directly from image data (Abidin et al., 2019).

In the forgery image detection, CNN based copy move / splicing manipulation got a decent accuracy in CASIA data but the results were worse with complicated manipulations. (Mallick et al., 2022). Further stability and accuracy of transformation are achieved by using models based on the dense Nets that convoluted by counting on the dense feature (Alzahrani, 2024), (Hingrajiya and Patel, 2023). When applied together with the Error Level Analysis (ELA), the preprocessing Pretrained networks, e.g., VGG-19, Inception-V3, ResNet, and EfficientNet, have also proven to be effective (Joshi et al., 2022).

The region-based approaches tend to have higher computational complexity in inference. Inception combined with Faster R-CNN that has multi-scale feature extraction to detect and localize the forged region. The CNN models based on region are the extensions of classification that involve localization to improve forgery localization accuracy (Vijiyakumar et al., 2024). Another approach which has become prominent in pixel-level forgery localization is Generative Adversarial Networks (GANs). The CGAN model that was proposed by (Abdalla et al., 2019) has the ability to detect the minor copy move based image forgeries. Recent GAN-based models are more successful in mask generation. But they have poor training stability and invariance due to data augmentation which results in hidden distortions (Pham and Park, 2023). More recent studies investigate the concept of robustness and generalization. (Wu et al., 2023) presented contrastive clustering to enhance resilience to post-processing operations while (Shi et al., 2023) presented discrepancy-guided reconstruction learning to incorporate semantic and forgery-specific cues. Regardless of these developments, there is still an evident trade-off between the robustness in classifications (CNN-based models) and fine-grained localization (GAN-based models).

However, this study offers a direct comparative study of a region-based CNN (InceptionV3-RCNN) and a Conditional GAN under the similar experimental conditions using CASIA v2.0 dataset unlike previous researchers who usually are the part of one. Through combined assessment of classification and localization and computational efficiency, the research paper raises real-world image forgery detection applicability trade-offs and contributes to the successful choice of appropriate models for image forgery detection.

Methodology

A. Dataset

The CASIA v2.0 Image Tampering Detection Dataset (Dong et al., 2013), is the dataset used for this study which consists of 3 folders namely authentic, tampered and ground-truth. It includes mainly tampered image with splicing, copy-move, and region editing. All the images are in JPEG format and with different sizes.

Experiments were done on an augmented version. Flips, rotations, and introducing noise were among the augmented data transformations with which the models trained on a more diverse set of distortions. For the both models, data set is split in 80:20 ratio for train and test split.

B. Data Preprocessing

Both the approaches required different preprocessing setup. In the case of InceptionV3-RCNN, the images were reduced to the size of 299x299x3. They were then normalized, and augmentation is performed on the dataset. For assessing the generalization capability of the model, the augmentation played an important role. It helps in changing lighting, orientation, and texture.

Similarly, In the case of the Conditional GAN (CGAN), image size 256x256x3 was taken which matched with their ground-truth masks. The spatial structure remains unchanged. It helps keeping both the image and the corresponding mask to transform identically since mask-aligned augmentation was used. As CGAN generator, generates to pixel level tampering masks.

C. InceptionV3-RCNN

InceptionV3-RCNN model combines the features of multi-level feature extraction with regions-based approach. It helps to identify regions of images they are manipulated. InceptionV3 backbone, as Shown in Figure 1 below, initially works on the input image i.e., augmented data and obtains multi scale feature maps. Parallel convolutional filters of different sizes are used in each layer. Convolutions filters helps the model to detect minor as well as high-level semantic manipulations that are often used to tamper the image.

The feature that are obtained then given fed to region-based convolutional neural network (RCNN) detection head. RCNN module primarily produces region-based area of interest that

has potential regions of manipulation. RCNN head the performs binary classification of each region proposal as tampered or authentic. Also bound-box regression is performed to localize forged regions. Multi-task loss is used in training that is a combination of classification loss and localization loss to enable the network to trade-off detection accuracy and spatial precision.

This combined architecture of InceptionV3 and RCNN helps locating global contextual cues but are also sensitive to localized discrepancies. So, it is highly suitable to localized various manipulation in detecting robust image forgery detection.

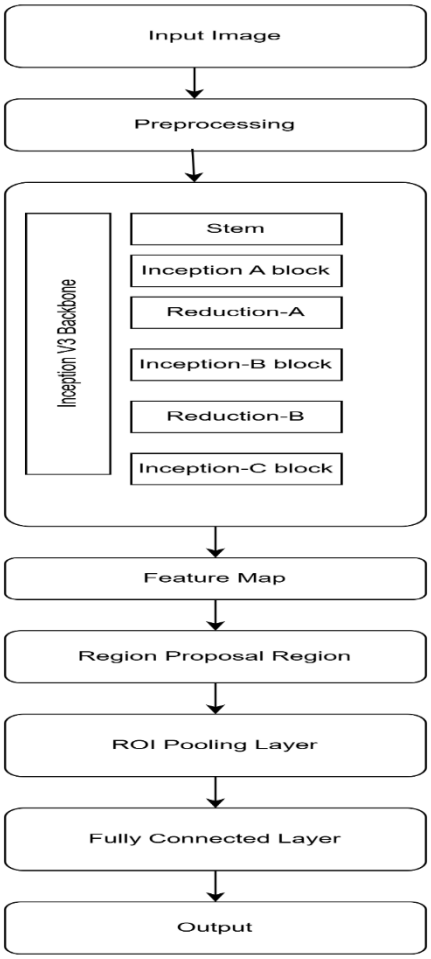


Figure 1: Inception V3 RCNN

.

D. Conditional GAN

The Conditional Generative Adversarial Network (CGAN) that will be used in the given research paper is aimed at pixel-wise forgery localization and authenticity classification. As shown in Figure 2 below, the CGAN is made up of both a generator and a discriminator both

of which are conditioned to the input image. Conditioning enables the model to gain manipulation patterns that are directly related to image content.

The generator has a U-Net architecture encoder-decoder model. The encoder gradually generates multi-scale feature representations that reflect not only global structure but also local artifacts and the decoder loads back a tampering mask that identifies manipulated areas. Jump connections between similar encoder and decoder layers retain spatial information and enhance boundary precision of the produced masks.

The image-mask pair is sent to the discriminator which compares two goals: adversarial discrimination of the real and generated masks, and binary classification of image authenticity. Adversarial loss, Dice loss and L1 loss are used to guide training, which together promote accuracy of mask generation, stable convergence and accurate localization.

The CGAN architecture has a more localized localization compared to region-based detection, but has a more complicated training compared to the adversarial optimization.

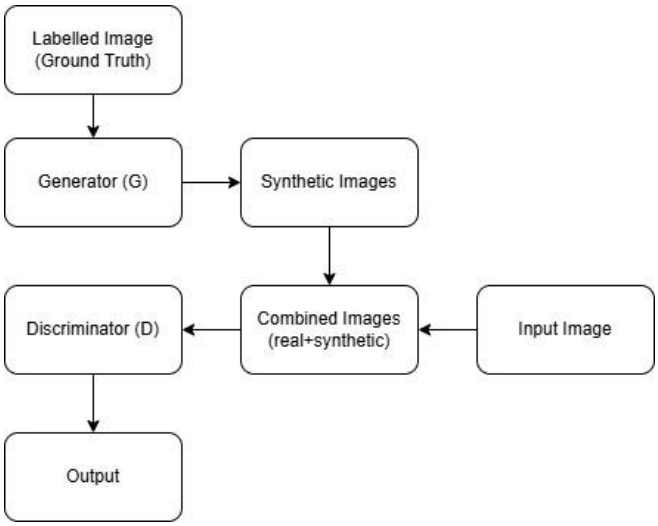


Figure 2: CGAN Architecture

E. Evaluation Metrics

To make certain, that the two models were fairly and comprehensively compared, set of conventional classification and computational measures were used to assess the performance of each under augmented and non-augmented dataset conditions.

For **classification assessment**, the following measures were used:

- **Accuracy**: It represents the overall ratio of correctly identified authentic and tampered images.
- **Precision**: It reflects how reliably the model identifies tampered images
- **Recall**: It measures the model's ability to detect actual tampered images.
- **F1-score**: It is harmonic mean of precision and recall.
- **AUC-ROC**: It evaluates the model's capability to distinguish between authentic and tampered images

Besides classification measures, the performance of computation using training time and inference time was also evaluated in order to get the clear idea of various models.

F. Training Configuration and Hyperparameter

In case of the InceptionV3-RCNN model, transfer learning was used which included the pretrained ImageNet versions of the InceptionV3 backbone. This model was trained end-to-end by means of the Adam optimizer with a starting learning rate of 0.0001 and a batch size of 16. 30 epochs were trained using early stopping which relied on a validation loss to avoid overfitting. In binary classification, categorical cross-entropy loss was utilized; in the RCNN detection head, Smooth L1 loss was utilized with bounding-box regression.

In the case of the Conditional GAN (CGAN), the discriminator and generator networks were both trained via the Adam optimizer with a learning rate of 0.0002 and $\beta_1 = 0.5$. Adversarial training has memory limits that led to the use of a batch size of 8. 50 epochs were used to train to get convergence between the discriminator and the generator. The adversarial learning binary cross-entropy loss combined with generator loss till the generator focuses on minimizing its loss, Dice Loss of coverage with ground-truth tampering mask, and L1 Loss of reconstruction of the model. The discriminator minimized binary cross-entropy loss on real/fake discrimination as well as authenticity classification.

G. Proposed Hybrid Approach (Future Work)

Although this study assesses InceptionV3-RCNN and CGAN separately, findings of the experiments suggest that they complement each other. InceptionV3-RCNN shows simple distinguish performance and stable training performance, and CGAN captures the fine pixel-level location information with faster calculation speed for inference.

A possible solution could be a hybrid approach by using InceptionV3-RCNN as the main detector to classify images and spot suspicious regions, and then use a CGAN-based module to obtain pixel-level localization in the detected suspicious regions. Such a pipeline would have the advantage of minimizing false positives while increasing the localization precision. This combined framework was not used in the current study and suggested as a future research direction. The comparative results presented in this work provide empirical justification of such integration and serve as a basement for future development of such hybrid models.

Result And Discussion

A. Classification Report

The following table shows accuracy, precision, recall, f1 score for each class of the respective models:

Table 1: Table showing Classification report of Inception V3-RCNN and CGAN

Classification Metrics	Accuracy	Precision	Recall	F1 Score	AUC-ROC
Inception V3-RCNN	98.61%	0.9640	0.9721	0.9680	0.8449
CGAN	76.59%	0.4686	0.6244	0.5354	0.7926

Table 1 represents the results of both models using the CASIA v2.0 dataset. The data presented in the table above is for precision, recall and F1-score are macro-averaged values between the classes 'authentic' and 'tampered'.

The InceptionV3-RCNN model performed with the accuracy of 98.61%, strong precision (0.9640), recall (0.9721) and F1-score (0.9680). With a AUC-ROC value of 0.8449, it has a reliable discriminative ability between true and fraudulent images under different thresholds. These results also indicate a good generalization of the model in spite of the variations introduced by data augmentation.

In contrast, the CGAN obtained a lower accuracy 76.59%, with lower precision value (0.4686) and moderate recall (0.6244). The lower F1-score indicates difficulties in ensuring that

classification is consistent under the conditions of augmentation. Although the CGAN shows reasonable performance of AUC-ROC (0.7926), it still has an inferior result of classification reliability than the region-based CNN method.

B. Accuracy Graph

The following graph shows the training and validation accuracy

Figure 3: Training Vs Validation Accuracy for Inception V3RCNN

Figure 3 give insight for the training and validation accuracy plots for the InceptionV3-RCNN model. The curves suggest that there is stable convergence and no overfitting.

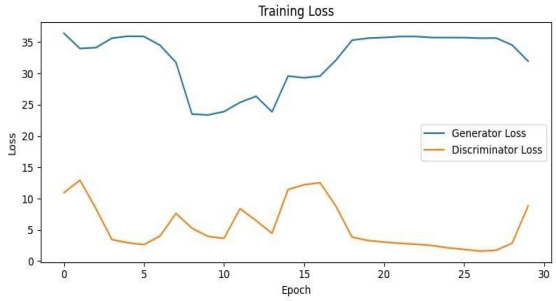


Figure 4: Generator and Discriminator Loss for CGAN

Figure 4 provides the insight of the generator and discriminator loss trends for CGA. It represents that adversarial stabilization is gradually achieved during the training epoch with higher variance than the CNN-based model.

C. AUC-ROC Graph

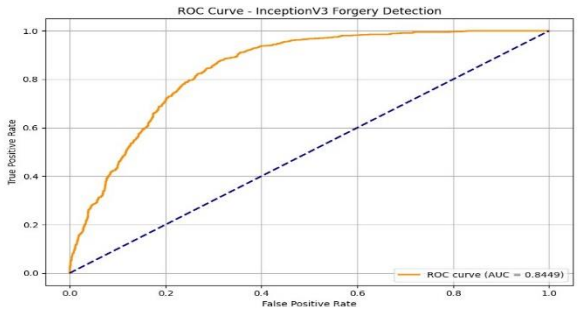


Figure 5: AUC-ROC For InceptionV3-RCNN

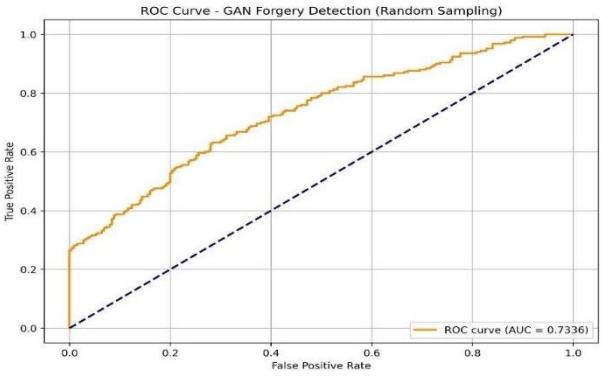


Figure 6: AUC-ROC For CGAN

Figures 5 and 6 show the Area under the Curve ROC (AUC) for InceptionV3-RCNN and CGAN respectively. Here InceptionV3-RCNN achieves a better AUC-ROC value than CGAN. The outcome of show that its value is satisfactory. Since it has no perfect separability which leaves a room for improvement by using hybrid or ensemble methods.

D. Computational Performance

Regarding the efficiency of the computations, InceptionV3-RCNN has less training time as it uses transfer learning. But it has more time to get the inference result. On the other hand, CGAN has more training time as a result of adversarial optimization. But it has a faster inference which may be useful in real-time applications of localization.

In general, the result indicates that a distinct trade-off between InceptionV3-RCN, having higher classification robustness and CGAN having finer-grained localization will be more efficient which can be explored in future work.

Conclusion

Performance analysis of InceptionV3-RCNN and Conditional GAN models were conducted for detecting image forgery using the CASIA v2.0 dataset. Through experimentation, it is concluded that InceptionV3-RCNN has better classification accuracy, better ratings of precision and recall, and more robust training characteristics for augmented data whereas CGAN shows weaker in the classification aspect, but has better result in inference speed and pixel-wise localization.

These results suggest that Inception V3RCNN models are beneficial where the classification is in priority, and the use of GAN-based models as beneficial in the tasks where localization is in priority. One area of future research could involve the studies of hybrid architectures that would unite the strength of CNN-based detectors and the localization capability of GAN-based schemes.

References

Abdalla, Y., Iqbal, M. T., & Shehata, M. (2019). Copy-Move Forgery Detection and Localization Using a Generative Adversarial Network and Convolutional Neural-Network. In *Information* (Vol. 10, Issue 9). https://doi.org/10.3390/info10090286

Abidin, A., Majid, H., A Samah, A., & Hashim, H. (2019). *Copy-Move Image Forgery Detection Using Deep Learning Methods: A Review*. https://doi.org/10.1109/ICRIIS48246.2019.9073569

Alzahrani, A. (2024). Digital Image Forensics: An Improved DenseNet Architecture for Forged Image Detection. *Engineering, Technology & Applied Science Research*, *14*, 13671–13680. https://doi.org/10.48084/etasr.7029

Dong, J., Wang, W., & Tan, T. (2013). *CASIA-2.0 Image Tampering Detection Dataset*. https://www.kaggle.com/datasets/divg07/casia-20-image-tampering-detection-dataset

Hingrajiya, K. H., & Patel, C. (2023). An Approach for Copy-Move and Image Splicing Forgery Detection using Automated Deep Learning. *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 1–5. https://doi.org/10.1109/ESCI56872.2023.10100202

Joshi, R., Gupta, A., Kanvinde, N., & Ghonge, P. (2022). Forged Image Detection using SOTA Image Classification Deep Learning Methods for Image Forensics with Error Level Analysis.

2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), 1–6. https://doi.org/10.1109/ICCCNT54827.2022.9984489

Mallick, D., Shaikh, M., Gulhane, A., & Maktum, T. (2022). Copy Move and Splicing Image Forgery Detection using CNN. *ITM Web Conf.*, *44*. https://doi.org/10.1051/itmconf/20224403052

Pham, N., & Park, C.-S. (2023). Toward Deep-Learning-Based Methods in Image Forgery Detection: A Survey. *IEEE Access*, *PP*, 1. https://doi.org/10.1109/ACCESS.2023.3241837

Ross, A., Banerjee, S., & Chowdhury, A. (2020). Security in smart cities: A brief review of digital forensic schemes for biometric data. *Pattern Recognition Letters*, *138*, 346–354. https://doi.org/https://doi.org/10.1016/j.patrec.2020.07.009

Shi, Z., Chen, H., Chen, L., & Zhang, D. (2023). Discrepancy-guided reconstruction learning for image forgery detection. *ArXiv Preprint ArXiv:2304.13349*.

Vijiyakumar, K., Govindasamy, V., & Akila, V. (2024). An effective object detection and tracking using automated image annotation with inception based faster R-CNN model. *International Journal of Cognitive Computing in Engineering*, *5*, 343–356. https://doi.org/https://doi.org/10.1016/j.ijcce.2024.07.006

Wu, H., Chen, Y., & Zhou, J. (2023). Rethinking image forgery detection via contrastive learning and unsupervised clustering. *ArXiv Preprint ArXiv:2308.09307*.

Anomaly Detection in System Logs Through Contrastive Self-Supervised Learning Integrated with the Wazuh SIEM Platform

Aakash Singh[1], Sujit Shrestha[2]

[1]Faculty Member, Divya Gyan College, Tribhuvan University

[2]Faculty Member, Divya Gyan College, Tribhuvan University

Article History

Received: 14/12/2025

Revised: 7/1/2026

Accepted: 10/01/2026

Published: 15/01/2026

Corresponding Author

Name: Aakash Singh

Email: aakashsingh@divyagyan.edu.np

ORCiD: https://orcid.org/0009-0006-2271-7354

Abstract

Anomalies in system logs nowadays are very hard and difficult to identify due to their nature and originations from accounts of legitimate users. Traditional security systems seem to be very struggling to detect because threats depend on explicit attack signatures and the complex behavioral patterns of insider persons. This study gives us a framework which integrates contrastive self-supervised learning with the Security Information and Event Management (SIEM) platform to improve the detection of anomalies in system logs. The proposed system is using a data preprocessing pipeline, contrastive learning engine, and also integration interface which is capable of analyzing logs without hampering operational works. To evaluate the performance this study evaluated unsupervised and supervised algorithms. The results gain a high accuracy and an F1-score in favor of Random Forest algorithm. This research shows if we combine temporal activity patterns with organizational context in open source SIEM platform we can find improved threat detection capabilities. The research focuses on modern SIEM platforms for better detection of anomalies in real time environments, showing better results which are based on different evaluation techniques.

Keywords: ***Anomaly Detection, Contrastive Learning, Random Forest, Isolation Forest, Wazuh SIEM, NSL-KDD.***

Introduction

Security Information and Event Management (SIEM) systems are crucial for modern cybersecurity operations. As organizations are nowadays expanding their digital infrastructure by adopting cloud services as well as distributed systems and remote work models. The volume and complexity of system logs have increased in a devastating way. SIEM platforms were developed to address the different challenges given by centralizing logs collection by enabling real-time monitoring and correlating different events across the systems and also generating notifications when suspicious activity is found (Muhammad et al., 2023; Khayat et al., 2025). Traditional SIEM platforms are largely driven by rule-based logic and signature matching methods. These mechanisms perform good when identifying known attack patterns [2]. The real problem arises when attackers come from predictable behaviors. Modern threat actors continuously modify their techniques and exploit already detected unknown vulnerabilities. In cases like that, static detection rules are often not sufficient (Schindler, 2018).

Enterprise networks like applications and also servers including different node from user's perspectives frequently produces vast streams of data. Analysts who work as a Security Operations Centers must deal with thousands and sometimes lakhs of alerts per day (Chamkar et al., 2025). This will create burden for analyst.

Among open-source SIEM solutions the Wazuh system has gained good adoption nowadays due to its flexibility and extensible feature set. It supports intrusion detection with file integrity monitoring as well as vulnerability assessment and compliance auditing (Muhammad et al., 2023). It integrates better with other tools in the ecosystem of security and also it provides organizations with cost-effective real time monitoring features. If we compare traditional SIEM systems with Wazuh it primarily depends on predefined rules of detection and database of signature. Also, these rules can be customized but they still need manual maintenance and continuous updates (Khayat et al., 2025). This dependency on static rules hinders Wazuh's effectiveness against attacks like zero-day exploits and hinders behavioral anomalies which may not violate patterns of existing detections techniques (Schindler, 2018).

The recent progress in machine learning have introduced different alternatives. Self-supervised learning has come up as a powerful mechanism for extracting meaningful patterns from datas that are unlabeled (Liu et al., 2021). Contrastive self-supervised learning is seen training models by contrasting between similar and dissimilar data pairs without the need of manual labels and also it has gained noticeable success among field like computer vision and also natural language processing including cybersecurity (Hojjati & Armanfard, 2022; Liu et al., 2021).

Log anomaly detection is found best suited for the approaches like this. In real world environments there is scarcity of log data and also there is imbalanced datas and they are expensive to obtain (Aziz & Munir, 2024; Grover, 2018). Self-supervised contrastive learning gives us a way to learn normal behavioral representations straight from log(raw) sequences. Also, the deviations from these representations can be represented as anomalies (Le & Zhang, 2021).

This study is seen exploring on how the contrastive self-supervised learning can also be integrated with the SIEM platform that is found enhancing anomaly detection capacities.

Statement of the Problem

Despite the heavy use the existing SIEM systems is found struggling to detect threat and face many limitations. One of the most popular challenge which is seen is the dependence on predefined detection rules (Muhammad et al., 2023). Also, they are seen struggling to detect novel attack techniques or versatile behavior of attacking. Advanced persistent threats (APTs) and zero-day exploits seems removing this gap by mimicking activity of genuine or legitimate users (Schindler, 2018; Dumitrasc, 2023).

Alert overload is another critical issue. Modern IT infrastructures obtains enormous amount of log volumes and traditional SIEM systems continuously flag normal activity as anomalies or suspicious (Chamkar et al., 2025). The massive false positives increase cost of operations and allows real incidents to be overlooked (Khayat et al., 2025).

Maintaining detection rules is seems to have demanding the good expertise and continuous updates. As threats evolve the security analysts must do revision of rule which sets to remain

effective [14]. This process is very time consuming and also it introduces the risk of configuration errors and also in organizations where there are limited cybersecurity resources (Ahmad, 2025).

Although we have machine learning that are proposed as a good solution there are many existing approaches that depend on supervised techniques which also require training data that are labeled (Aziz & Munir, 2024). In practice, labeled data that are high quality and good datasets of anomalies are rarely available (Grover, 2018; Le & Zhang, 2021). So exactly this scarcity in data stops the scalability and applicability of solutions that are supervised.

So, there is a clear need for detection frameworks which are capable of learning from datas that are unlabeled and also integrating with existing SIEM platforms, and which reduces both false positives and rule maintenance which are manual (Liu et al., 2021; Khayat et al., 2025).

This research has following questions:

1. How contrastive self-supervised learning can be effectively integrated into the Wazuh SIEM platform that can enhance accuracy of log anomaly detection?
2. What architectural and implementation strategies is found enabling integration without hampering the existing workflows?
3. How does the proposed approach will compare rule-based detection and supervised learning methods in terms of accuracy and also false positive rate including computational efficiency?
4. What training strategies, hyperparameter configurations will optimize the performance in enterprise log formats?

The objectives of this study are as follows:

- To design and evaluate a contrastive self-supervised anomaly detection framework by integrating with Wazuh.
- To benchmark its performance against rule-based detection and supervised learning models with different metrics like precision, recall, F1 score and utilization of resource

Literature Review

The integration of machine learning with SIEM systems nowadays is an active area of research. Traditional log anomaly detection mainly focuses on statistical methods and rule-based systems. Schindler (2018) had showed the disadvantages of commercial SIEMs which handles very complex attack patterns.

Le and Zhang (2021) proposed NeuralLog, a BERT-based approach which is found eliminating the need for log parsing.

Liu et al. (2021) presented CoLA, which is a contrastive framework for detection of anomalies on attributed networks. Zheng et al. (2021) elaborated the field by adding contrastive and generative learning which is used for graph anomaly detection. Hojjati and Armanfard (2022) showed that contrastive learning principles are used in anomaly detection which can be transferred to log-based analysis.

Despite these factors, limited research has been done on the practical integration of advanced techniques of machine learning with SIEM platforms. Many studies are only theoretical which showed a kind of gap in understanding actual deployment in real-world environments (Grover, 2018).

Research Methodology

This study shows a mixed methods techniques which combines experimental research with practical implementation. The system used has three parts: a part that gets the data ready a part that helps the system learn on its own and a part that connects with the Wazuh SIEM platform. The part that gets the data ready takes care of a things. It looks at the logs that come in makes sure they are all in the format and pulls out the important information. The logs from different places are taken by it and it makes them useful by focusing on the good things and content. This part is also good at figuring out how to read kinds of logs that you might find in a big company. The part that helps the system learn on its own is showing the part of the system which they are proposing. The system they are talking about uses this learning part to make it work. The learning part is really important for the system to be good at its job.

The system is made up of the data preparation part the learning part and the connection part, with the Wazuh SIEM platform and the learning part is what makes the system special. This component implements a novel contrastive learning architecture specifically designed for sequential log data. The model learns meaningful representations by contrasting normal operational patterns against potential anomalies through carefully designed positive and negative pair generation strategies. The integration interface provides good connectivity with the SIEM platform which enables the real time anomaly detection and alert generation without hampering existing operational workflows.

System Architecture

The architecture of this study consists of three primary components:

1. **Pipelining of Data Preprocessing:** It handles the ingestion of log, normalization, and feature extraction using technique like adaptive parsing.
2. **Contrastive Self-Supervised Learning Engine:** It uses a architecture of contrastive learning which learns by contrasting normal operational patterns against potential anomalies.
3. **Integration Interface:** This connects with the SIEM platform via APIs to enable detection in real-time.

Algorithms Implemented

To evaluate the performance and the differences between supervised and unsupervised methods following algorithm was implemented by the study:

- **Random Forest:** An ensemble learning method using bagging which is used to build multiple decision trees.
- **Isolation Forest:** An unsupervised algorithm which isolates anomalies rather than showing normal data points.
- **InfoNCE:** A loss function is used to train models to find differences between similar and dissimilar data pairs.

Data Preparation

The study used the NSL-KDD dataset which is a balanced version of the KDD dataset. It is widely used for intrusion detection systems. Data preprocessing involved parsing, normalization, feature extraction, and temporal sequencing.

Integration Strategy

An Ubuntu node with SIEM agent was configured to gain resource usage and log data to the Wazuh server.

Finding and Result

The system was tested in Ubuntu with the SIEM server as a real-world scenario. Various anomalies were simulated which include failed login attempts and also abnormal utilization of Linux resource.

Operational Results

The integration allowed for real-time monitoring on the dashboard of SIEM platform.

- **Failure in Logins:** The anomaly detector showed irregular login patterns.
- **Utilization of Resource:** Anomalies in CPU and also in memory usage were found and visualized on the dashboard.
- **Threat Detection:** The system captured alerts for vulnerability and also malware detection within the last 24-hour.

Performance Evaluation

These models were evaluated which are based on Accuracy, Precision, Recall, and F1-Score.

Table 1: Model Comparison based on performances

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	0.997	0.970	0.980	0.970
Isolation Forest	0.806	0.555	0.147	0.249

Discussion

The Random Forest model showed very good performance with high accuracy and a high F1-score. This showed that when labeled data is available, supervised methods are highly effective.

The Isolation Forest showed significantly lower performance mainly in Recall. Whereas the accuracy in initial stage was seen acceptable but it greatly reflects the dominance of normal transactions in the dataset. Also, the low recall indicates that the unsupervised model missed a number of real anomalies, and the low precision indicates a high rate of false positives.

A scatter plot analysis showed us that the algorithm could identify outliers (red points) far from dense clusters of normal logs (blue points) where the unsupervised approach struggled with the data imbalance which leads to high flagging of normal logs as threats which is also called false positives.

Conclusion

This study generates anomaly framework identification that is based on contrastive self-supervised learning which is integrated with the SIEM platform. The evaluation results shows that supervised methods like Random Forest significantly outperform unsupervised methods like Isolation Forest in detecting anomalies. The integration with Wazuh proves that when we combine temporal activity patterns with organizational context it gives better detection performance and offers a scalable solution for enterprise environments. This approach reduces the dependency on static rules and manual analysis.

Traditional models which seem to be dependent on labeled data or static rules, this approach is flexible, less resource required and also better suited for real-world deployment mainly in the organizations or offices where there are limited cybersecurity capabilities.

In dynamic network environments where attack patterns are constantly changing, the suggested framework exhibits strong adaptability. Without requiring a lot of manual labeling, the system can learn meaningful representations of typical behavior straight from raw log sequences by utilizing contrastive self-supervised learning. This greatly lessens the need for frequent rule updates and expert-driven rule engineering, which are frequent problems in traditional SIEM deployments.

Improvements in detection accuracy, precision, and recall are also highlighted by the experimental analysis, which also shows a decrease in false positive rates, which usually

overwhelm Security Operations Center (SOC) analysts. Improved operational efficiency and quicker incident response times are two benefits of fewer false alarms. Furthermore, compatibility with current enterprise infrastructures is guaranteed by the modular integration with Wazuh, which makes deployment feasible and economical. Overall, the framework provides a balanced solution with the combination of intelligent anomaly detection and real time monitoring which shows the way for adaptive mechanisms for cyber security defense in modern organizations.

References

Ahmad, W. (2025). Unveiling Anomalies: Leveraging Machine Learning for Internal User Behaviour Analysis–Top 10 Use Cases. *International Journal of Innovative Technology and Interdisciplinary Sciences, 8*(1), 1789-1805.

Aziz, A., & Munir, K. (2024). Anomaly Detection in Logs using Deep Learning. *IEEE Access, 12*, 176129-176145. https://doi.org/10.1109/ACCESS.2024.3512847

Chamkar, S. A., Zaydi, M., Maleh, Y., & Gherabi, N. (2025). ML-Driven Log Analysis for Real-Time Cyber Threat Detection in Security Operation Centers. *Preprints*. https://doi.org/10.20944/preprints202412.0123.v1

Dumitrasc, V. (2023). *Anomaly Detection Through User Behaviour Analysis* [Master's thesis, Universitat Politècnica de Catalunya].

Grover, A. (2018). *Anomaly detection for application log data* [Master's thesis, San José State University].

Hojjati, H., & Armanfard, N. (2022). Self-supervised acoustic anomaly detection via contrastive learning. *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 3371-3375). IEEE. https://doi.org/10.1109/ICASSP43922.2022.9746207

Khayat, M., Barka, E., Serhani, M. A., Sallabi, F., Elmedany, M., & Bentahar, H. (2025). Empowering Security Operation Center with Artificial Intelligence and Machine Learning–A Systematic Literature Review. *IEEE Access, 13*, 12045-12067. https://doi.org/10.1109/ACCESS.2025.3523901

Le, V. H., & Zhang, H. (2021). Log-based anomaly detection without log parsing. *Proceedings of the 36th IEEE/ACM International Conference on Automated Software Engineering* (pp. 492-504). https://doi.org/10.1109/ASE51524.2021.9678773

Liu, Y., Li, Z., Pan, S., Gong, C., Zhou, C., & Karypis, G. (2021). Anomaly detection on attributed networks via contrastive self-supervised learning. *IEEE Transactions on Neural Networks and Learning Systems, 33*(8), 8270-8282. https://doi.org/10.1109/TNNLS.2021.3068344

Muhammad, A. R., Sukarno, P., & Wardana, A. A. (2023). Integrated security information and event management (SIEM) with intrusion detection system (IDS) for live analysis based on machine learning. *Procedia Computer Science, 217*, 1528-1537. https://doi.org/10.1016/j.procs.2022.12.352

Schindler, T. (2018). Anomaly detection in log data using graph databases and machine learning to defend advanced persistent threats. *arXiv preprint arXiv:1802.00259*.

Efficient Fine-Tuning of Vision Transformers for Histopathological Image Classification via Low-Rank Adaptation

Keshab Bashyal[1]

[1]Faculty Member, Divya Gyan College, Tribhuvan University

Article History

Received: 21/12/2025

Revised: 29/12/2025

Accepted: 07/01/2026

Published: 15/01/2026

Corresponding Author

Name: Keshab Bashyal

Email: keshabbashyal@divyagyan.edu.np

ORCiD: https://orcid.org/0009-0000-8461-4824

Abstract

Excessive computational and memory requirements associated with traditional full-fine tuning, despite their remarkable performance, significantly hinder the pragmatic application of modern vision transformers especially for histopathological image analysis. To alleviate this problem, modern transformers like Swin and DeiT are systematically evaluated using Low Rank Adaption (LoRA) technique, which is a parameter efficient fine-tuning technique, especially designed to shorten training time in natural language processing. When LoRA is applied to histopathological image classification, surprisingly, LoRA adapted Swin and DeiT models performs comparable performance across all evaluation metrics: accuracy, precision, specificity and F1 score, compared to their full- fine-tuned counterparts by updating less than 2% of the model's parameters. The results show that LoRA not only accelerate training speed by updating fewer than 2% of the model's parameters but also achieves superior accuracy for both Swin (99.42% vs. 99.21%) and DeiT (99.27% vs. 98.91%) compared to their fully fine-tuned counterparts on NCT-CRC-HE dataset. Consequently, efficient fine-tuning using LoRA can provide an alternative way to traditional full fine-tuning without scarifying performance while boosting training speed, opening new avenues for various medical image classification problems.

Keywords: ***Vision Transformers, Parameter-efficient, Histopathological image, Fine-tuning, Superior accuracy, Performance, Classification.***

Introduction

A fundamental component of contemporary medicine is histopathology, the microscopic analysis of tissue to identify illness. This field is especially important for the diagnosis and treatment of cancers, especially colorectal cancer (CRC). Being the third most common cancer to be diagnosed and a major cause of cancer-related death, colorectal cancer (CRC) is one of the most common and deadly cancers in the world and a major public health concern.(Mármol et al., 2017). The histopathological examination of biopsied or surgically removed tissue, in which pathologists examine cellular morphology and tissue architecture to ascertain the presence and severity of the disease, is the gold standard for diagnosing colorectal cancer. Nevertheless, this manual process has many challenges despite its fundamental role. It is a labor-intensive and time-consuming task with significantly inter-observer variability that may affect the performance of the diagnosis.(Demir & Yener., 2005).

Current developments in artificial intelligence, especially in the area of deep learning, have provided computational tools that can automate and improve digital pathology diagnostic accuracy. Despite being fundamental, Convolutional Neural Networks (CNNs) are not able to capture the long-range dependencies within the image which are essential for understanding complex tissue structures due to their high inductive biases (Romero et al., 2022). Vision Transformers (ViTs), on the other hand, can capture long-term dependencies in the images using self-attention mechanisms(Dosovitskiy et al., 2020). Since then, this concept has developed into a new class of advanced transformers, such as the Swin Transformer and DeiT. The requirements of medical imaging are perfectly met by their multi-scale designs. Nevertheless, the colossal computational cost of these large-scale models severely limits their practical application, thereby, making conventional fine-tuning a major bottleneck for many research and clinical settings.

In this paper, the multiclass classification of colorectal histopathological tissue across nine classes has been implemented and evaluated using the Parameter-Efficient Fine-Tuning of two top vision transformers, Swin and DeiT. Each of these models has been modified using the Low-Rank Adaptation (LoRA) technique using the NCT-CRC-HE dataset. The LoRA-adapted models' performance has been compared to that of their counterparts that have undergone standard fine-tuning. The comparison is thorough and includes important indicators of

computational efficiency like training and inference times in addition to common classification metrics like accuracy, recall, precision, F1-score, and specificity.

Literature Review

The usage of neural networks and deep learning in colorectal cancer detection and classification task has currently been leveraged by the integration of recent transformer architectures to mitigate the limitations of conventional convolution neural networks. Current studies have demonstrated powerful methods that facilitates these implementations to enhance medical diagnostic accuracies and precision. For instance, the standard U-Net architecture is enhanced by using skip connection. Which has used a Swin transformer for feature extraction and achieved a 95.8% accuracy on the NCT_CRC_HE_100K dataset (Qin et al., 2024). Using this, the Colorectal cancer detection network was introduced to integrate dilated convolutions with coordinate attention model with a cross-shaped window transformer in order to capture local, global and subtle tissue changes, thereby, reaching an accuracy of 98.96% on the same dataset.

The excessive computational and memory requirements of these large-scale ViTs models have demanded the development of more efficient fine-tuning strategies for fast training. Scaling and Shifting Features (SSF) has been developed as a highly efficient method for fine-tuning pretrained models by learning only to scale and shift features within frozen network blocks (Tay et al., 2023). This approach frequently outperforms full fine-tuning in various domain and problems while training significantly fewer parameters and providing the benefits of being mathematically merged back into the model to ensure no inference latency. Likewise, the application of Adapters and Low-Rank Adaptation (LoRA) to large models like SEEM and Mask DINO has a competitive performance is achievable by updating around 1–6% of model's total parameters. That can significantly reduce the costs to train the model compared to full-fine tuning (Abou Baker et al., 2024). Addressing the heuristic nature of many Parameter-Efficient Fine-Tuning (PEFT) methods. Sensitivity-aware PEFT (SPT) was designed to identify and prioritize most essential parameters for specific problems (Yin et al., 2023). This technique effectively allocates tuning budget to the most critical and sensitive weight matrices, which in turn, boost performance and achieving the state-of-the-art performance across major benchmarks (Xin et al., 2024).

The practical usage of these efficient fine-tuning methods are particularly vital in specialized medical imaging tasks like classification and segmentation. In cervical cancer detection task, LoRA-based models have solved data scarcity challenges and also outperformed standard CNNs by reducing trainable parameters to less than 1% of the base model (Hong et al., 2024). LoRA has been successfully utilized to adapt large vision models for lung nodule malignancy classification tasks which has achieved around 3% higher ROC AUC than previous state-of-the-art methods while using 89.9% fewer parameters and reducing training time by 36.5% (Veasey & Amini, 2025). The scalability of those models is supported by other techniques such as federated learning via Kubernetes for privacy-preserving image synthesis (Preda et al., 2025) and introduction of hierarchical cell transformers that model spatial interdependencies within whole slide images to give superior performance for survival prediction and cancer classification problems (Yang et al., 2024).

Vision Transformers (ViTs) are known for being their fundamental processing characteristics. ViTs shows a greater shape bias and closely resembles with the nature of human error patterns compared to ResNets based on evaluations on various datasets. That's why they can capture long-term dependencies within the images (Tuli et al., 2021). Research also indicates that general-purpose models like Swin Transformer V2 can outperform specialized medical encoders in various tasks like cell segmentation (Vadori et al., 2025). Those architectural advantages have been also evaluated in novel domains such as deep ultraviolet fluorescence breast cancer imaging in which patch-level transformers achieved around 98.3% accuracy and surpassed other comparable models up to nearly 13% (Afshin et al., 2025).

Despite these successes, this field continues to integrate the benefits of both transformers and specialized convolutional frameworks with hybrid designs. While transformers excel in capturing long-range dependencies, comparative analyses on datasets like BraTS 2020 show that specialized models like nnU-Net can still outperform them in certain segmentation tasks (Träff, 2023). Some research has implemented dataset shortcuts that demonstrates simpler models like EfficientNet-B0 can occasionally outperform complex transformers on clean data, which underscores the importance of combining high-quality data with innovative architectures like the CViTS-Net hybrid (Ignatov & Malivenko, 2025; Kanadath et al., 2024). Multi-modal integration has also seen progress through models like TransMed, that uses a cross-modal

transformer encoder to fuse PET and CT data for nasopharyngeal carcinoma prognosis. Those model's performance has surpassed traditional CNN-based fusion techniques (Dai et al., 2021).

The research has expanded in clinical interpretability and practical implication as those models perform exceptionally well. Hierarchical Vision Transformers have used prototype-based learning in order to provide clear explanations for their high-performance decisions, which helps to build clinical usability (Gallée et al., 2025). Hybrid models like EffNetV2_ViT and the BreaST_Net ensemble have achieved accuracies as high as 99.8% by processing images in multiple magnifications (Hayat et al., 2024; Tummala et al., 2022). To address the constraints of clinical environments, RMT-Net integrates ResNet-50 with transformers to maintain a condensed model size suitable for quick diagnosis (Ren et al., 2023). However, as noted in studies of lung disease detection and skin cancer classification, the high computational demands of the most powerful transformers still pose several challenges for resource-constrained edge devices (Aladhadh et al., 2022; Uparkar et al., 2022). This highlights the ongoing need for optimization techniques like pruning, quantization, and parameter-efficient fine-tuning to ensure that powerful diagnostic tools can be equitably deployed in diverse medical settings.

Methodology

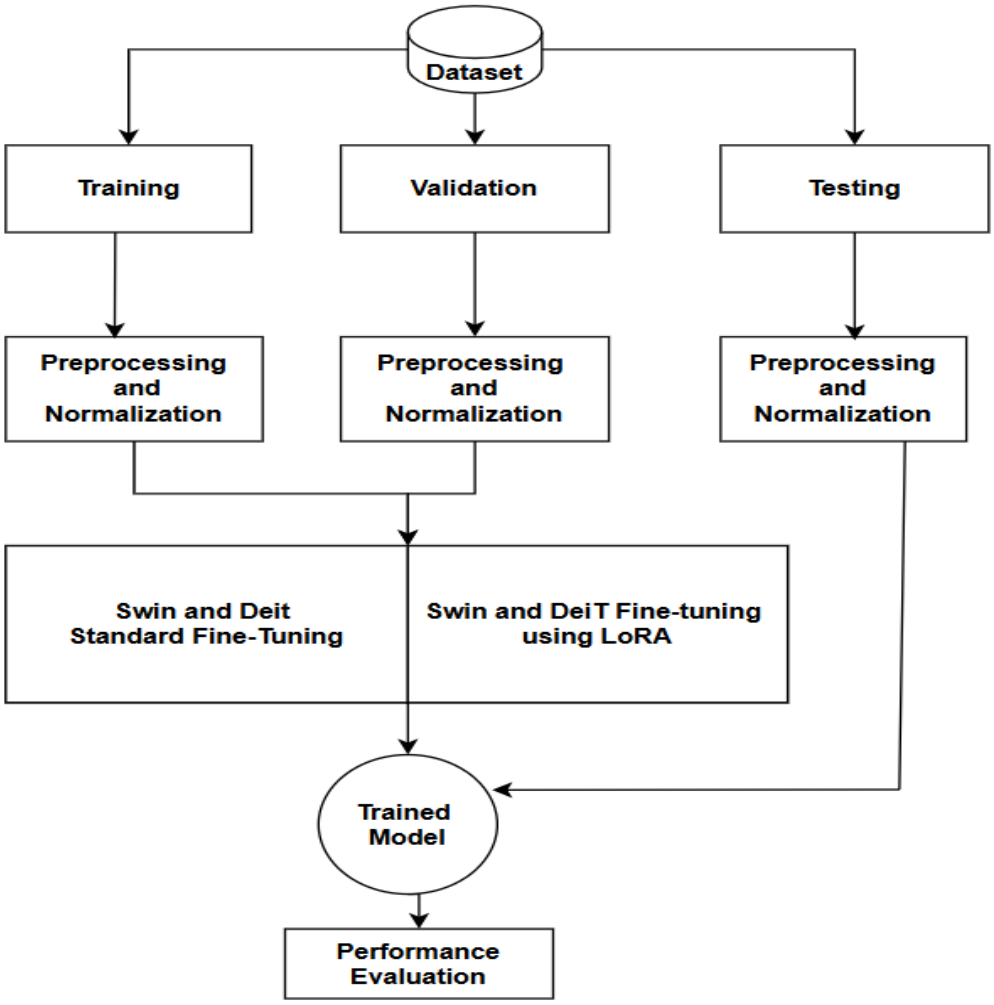


Figure 1: Research Methodology

Dataset Description

In this research the NCT-CRC-HE dataset collected from the National Center for Tumor Disease in Heidelberg is used, which consists of 100000 H&E stained images of colorectal cancer tissue parts. And the images are collected from 86 patients. All the images are standardized into 224*224 pixels for uniformity in the models. This dataset consists of nine different classes of different tissues. For better performance and evaluation, the dataset is partitioned into 80/10/10 split, leaving 80,000 images for training, 10,000 for validation tasks such as hyperparameter tuning and early stopping, and 10,000 as a strictly held-out test set for fair model evaluation.

Explanation of each 9 classes

- **Adipose (ADI)** -- A fat tissues, in which a large and empty appearing adipocytes that store lipids and provides structural cushioning.
- **Background (BACK)** -- It is a non-tissue area on the slide such as empty glass or regions without cellular material.
- **Debris (DEB)** -- It accumulates dead cells necrotic fragments and cellular breakdown products that is found in damaged tissue regions.
- **Lymphocytes (LYM)** -- It is a cluster of small and round immune cells that represent the body's immune response within or around the tumor.
- **Mucus (MUC)** -- Pools of mucin secreted by glands, which is often found in normal tissue or associated with mucinous tumors. It is a healthy tissue.
- **Smooth Muscle (MUS)** --- A long and spindle-shaped muscle fibers that form a part of the bowel wall and provide contractile function.
- **Normal Colon Mucosa (NORM)** -- Healthy epithelial glands and supporting tissue of the colon, which shows typical glandular architecture.
- **Cancer-associated Stroma (STR)** -- Fibrous and dense connective tissue that surrounds and supports tumor cells and often remodeled by cancer.

Table 1: Number of Training, Validation and Testing Images per class

Class	Training Images(80k)	Validation Images(10K)	Testing Images(10K)
ADI	8,295	1,080	1,032
BACK	8,470	1,020	1,076
DEB	9,295	1,130	1,087
LYM	9,322	1,090	1,145
MUC	7,079	930	887
MUS	10,891	1,290	1,355
NORM	6,927	940	896
STR	8,261	1,060	1,125
TUM	11,460	1,460	1,397

Data Preprocessing

For better model evaluation, a uniform preprocessing protocol without data augmentation is applied across all data partitions. To provide stable and numerical optimization, the pipeline first converts 224x224 images to 32-bit floating-point PyTorch tensors. Furthermore, for efficient transfer learning, the tensors have been normalized using standard ImageNet mean and standard deviation statistics.

Low Rank Adaptation (LoRA)

Rather than fine-tuning all the model's parameters, the central idea of the LoRA technique is in training time in which the original pre-trained weights remain frozen. Trainable low-rank decomposition matrices, the LoRA adapters-are injected into the self-attention blocks of the models to adapt the model for the problem specific task. These LoRA modules are applied to the query (Wq) and value (Wv) projection matrices within all multi-head self-attention block across all of the model's stages. To reduce training time, gradient updates are only computed and applied to these newly introduced LoRA adapters. However, the rest of the model's parameters including the MLP layers and other architectural components remain unchanged.

The final classification process follows a fully connected layer. After the last stage, the output vectors are aggregated by global average pooling layer to produce a single feature vector of the entire image for stable and compact output representation. The vector is then passed to a final fully connected classification head that projects it to a logit vector with a dimension equal to the number of target classes. This classification head and the LoRA adapters are the only trainable components of the model which drastically reduce the training time. Lastly, a softmax function is applied to the logits to generate a normalized probability distribution over the classes, from which the final prediction is obtained.

The core equation of LoRA is,

$$h = W_0 x + \frac{a}{r}(B A) x, \quad \text{where}$$

- "h" is the output vector of the layer
- "x" is input vector of the layer
- "a" is hyperparameter that modules the magnitude of the adaptation
- "r" is rank of the adaptation, a key hyperparameter

- "W_0" is the original, pre-trained weight matrix, which is frozen and does not get updated during training time.
- "A" and "B" are the two new small, low-rank adaptation matrices. These are the only matrices that are trained.

The Equation for Merging (Inference): Matrices A and B are fixed during inference after the training. To avoid any inference latency, one time merge is performed.

$W_{Final} = W_0 + \frac{a}{r}$(B A)

Here, the output matrix W_{Final} has the same dimension as the original W_0. and A and B are discarded and deployed a model with this single, unified weight matrix. The inference pass then becomes a simple h = W_{Final} * x , which is identical in speed to a standard fine-tuned model.

Swin Model

The Swin base model classifies 224×224 colorectal images using a hierarchical design which divides the input image into 4×4 non-overlapping image patches. This process first generates 3136 tokens, to effectively capture both the local details and global context, layers are projected into 128-dimensional embedding vectors. The architecture is structured into four progressive stages in which the number of Swin Transformer blocks per stage is set to [2, 2, 18, 2]. Each block contains two key components.

1. In Swin transformer multi-head self-attention works within local 7×7 windows which employs a shifted window mechanism so as to provide cross-window communication to detect long-range dependencies. The number of self-attention heads is increased hierarchically [4, 8, 16, 32] across stages to progressively enrich feature representation of the model.
2. A two multi-layer perceptron (MLP) is used with the GELU activation function, which introduces non-linearity and helps in learning involved hierarchical transformations. Each MLP sub-layer is preceded by layer normalization and followed by residual connections, offering stable training and gradient flow.

By doubling feature dimensionality to 1024, patch merging layers down sample spatial resolution that balance computational efficiency with semantic depth. Eventually, global

average pooling aggregates these features for a fully connected classification head, which applies softmax to predict the final nine tissue subtypes.

Multi-head Self-attention within 7×7 Shifted Windows: In multi-head self-attention the 7*7 shifted windows divide the image tokens into fixed sized local windows, which computes self-attention separately in each window in order to reduce computational complexity. To enable cross-window connections and better capture global context, those windows are shifted between consecutive layers. This process balances efficiency and expressive power in formulating spatial relationships.

DeiT Model

By dividing 224×224 RGB images to 196 non-overlapping 16×16 patches, the DeiT base_16 model classifies colorectal histopathology tissue. These patches are linearly projected into 768_ dimensional embedding vectors. That generates a token sequence which optimizes the transformer architecture for making training more efficient on limited medical datasets. In this model two different special learnable tokens are prepended.

1. A CLS token whose output embedding is used during inference to perform final class prediction.
2. A DIST token which is introduced uniquely in the DeiT framework that is designed to absorb signals from a teacher model through knowledge distillation technique during training. This enables the model to mimic the behavior of a stronger and typically pre-trained CNN-based teacher.

Learnable positional embeddings are appended to the sequence of 198 tokens to ensure spatial awareness during training so that the model can track the position of each pixel. Which includes the special “CLS” and “DIST” tokens. And then after, input is processed through 12 Transformer encoder blocks in a non-hierarchical design. This maintains constant feature dimensionality and resolution throughout the network, which is different from architectures like Swin. Each encoder block is composed of two main sub-layers.

1. A global multi-head self-attention module with 12 attention heads which allows each token to interact with all others globally, which in turn captures better long-range dependencies.

2. A two-layer multilayer perceptron “MLP” that introduces non-linearity via the GELU activation function, enhancing the network's capacity to learn intricate feature transformations.

The architecture uses residual connections and layer normalization, while maintaining a constant 768-hidden-dimensional state. For instance, the “CLS” token drives a linear classifier to identify the nine tissue subtypes. The “DIST” token facilitates knowledge distillation from a teacher CNN that enables the model to achieve high performance and data efficiency despite limited annotated training samples.

Experimental Setup & Environment

The models are implemented using PyTorch, which is the most popular deep learning library in the ML research community. NumPy is used for numerical computation. For visualization and preprocessing Scikit- learn and matplotlib are used. The experiments have been conducted on Google colab by utilizing NVDIA T4 GPU of around 16 GB RAM to accelerate training and testing through CUDA optimized kernels.

Configuration for Standard Fine-tuned models

For both Swin base and DeiT base architectures, a conventional fine-tuning strategy has been adopted to update their pretrained weights for the 9_class classification task. The original classification heads are replaced with a new custom head consisting of a dropout layer with a probability of 0.20. This is followed by a fully connected linear layer. To provide stable convergence while preserving discriminative representations learned by the pretrained backbones, the AdamW optimizer has been used with discriminative learning rates: $\beta_1 = 0.9$, $\beta_2 = 0.99$, $\varepsilon = 1e^{-5}$. The base parameters are optimized using a small learning rate of 1×10^{-6} whereas the newly initialized classification head is trained with higher learning rate of 1×10^{-5} along with a weight decay of 0.05 for regularization.

Both models have been trained over 10 epochs. Which use a OneCycleLR learning rate scheduler with cosine annealing. The scheduler incorporates a warm up phase occupying 30% of the total training steps “pct_start = 0.3” and utilizes peak learning rates of $[1 \times 10^{-6}, 1 \times 10^{-5}]$ for backbone and head respectively. The learning rate is initialized at one-tenth of the maximum value “div_factor =10” and gradually decays to one-hundredth of the initial rate

"final_div_factor = 100". Which ensures a smooth transition from warm up to cooldown. This scheduling strategy enables precise weight updates while preventing disruption of the pretrained spatial and semantic hierarchies during fine-tuning.

Configuration for LoRA adapted models

Both Swin and DeiT models have been optimized using the AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.99$, $\varepsilon = 1e^{-5}$ and a weight decay of 0.05. Separate learning rates of 3×10^{-5} and 1×10^{-4} is assigned to the LoRA parameters and the classification head for both models. A cosine annealing learning rate scheduler with warm restarts has been employed with an initial restart period $T_0 = 15$ and T_mult = 1. At first, the minimum learning rate is set to 5×10^{-6} to prevent premature convergence. To reduce memory requirement the validation phase is conducted in evaluation mode without gradient computation. Mixed precision inference has been applied using automatic casting to improve computational efficiency and the validation loss is accumulated over all batches. While accuracy is computed by comparing predicted class labels with ground truth labels, final validation loss and accuracy is obtained by normalizing over the total number of batches and samples respectively.

A Swin base Transformer model pretrained on ImageNet-22K is used as the backbone model. Low-Rank Adaptation (LoRA) blocks with rank r = 8 and scaling factor $\alpha = 16$ have been integrated to the attention (QKV) and MLP layers of each transformer block. All original backbone parameters are kept frozen to preserve pretrained pattern representations. Only the LoRA parameters and the modified classification head with dropout of 0.20 are trainable. This design has significantly reduced the number of trainable parameters while maintaining strong fine-tuning capability. The Swin model is trained over 10 epochs.

A pretrained DeiT base Vision Transformer has been employed as the backbone model. Low-Rank Adaptation (LoRA) modules with rank r = 8 and scaling factor $\alpha = 16$ have been injected into the attention (QKV) and MLP layers of each transformer block. All original DeiT parameters are kept frozen to preserve pretrained knowledge. Only the LoRA parameters and a modified classification head with dropout of 0.20 is updated during training to reduce number of trainable parameters by offering efficient fine-tuning.

Result Analysis

This data depicts the performance of LoRA adapted fine-tuning and full fine- tuning, LoRA significantly reduces training time by around 16% and 12% for both Swin and DeiT models respectively by reducing the number of trainable parameters from approximately 86-85 million to just 1.3-1.2 million. This result is achieved with no impact on inference speed and even leads to a slight better performance than full-fine tuning. As compared to the base models, the results show the Swin model with LoRA technique achieves a highest accuracy with slight margin.

The macro averaged results in the charts are calculated, the practical application of LoRA offers a dramatic performance improvement in both Swin and DeiT models. The LoRA adapted Swin model has achieved the best results among 4 models in all evaluation metrices: accuracy, precision, recall and F1 score. All of which are around 99.4%. This points a slight improvement than baseline Swin model. Likewise, similar results is obtained in LoRA adapted DeiT model in which LoRA boosts its key metrics from around98.9% to 99.27%, making LoRA an alternative method to full-fine tuning.

Swin Model Evaluation

Swin Standard Fine-Tuning Results

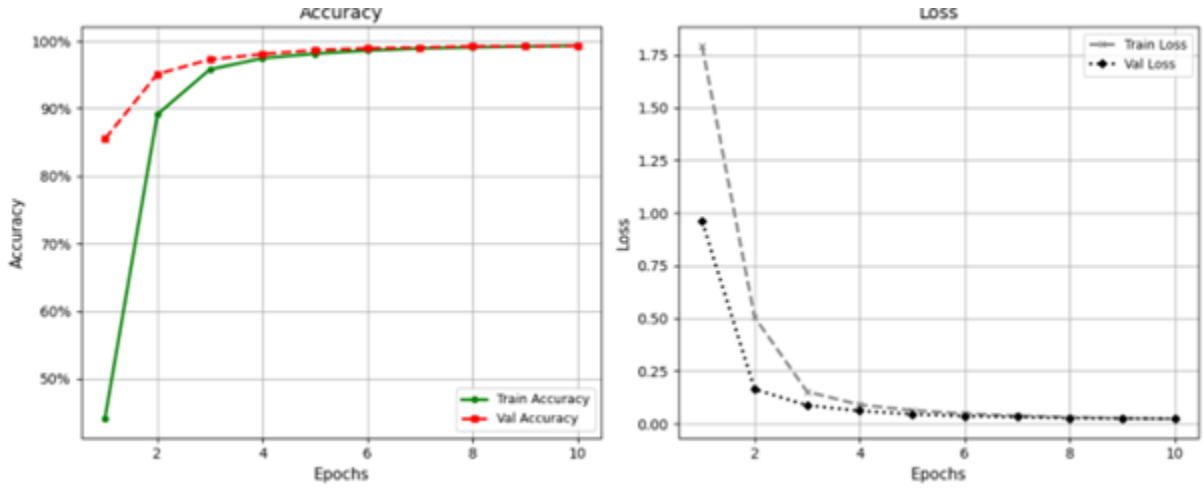


Figure 2: Swin Fine-tuned Model Accuracy and Loss curves

Swin Model with LoRA-Adapted Results

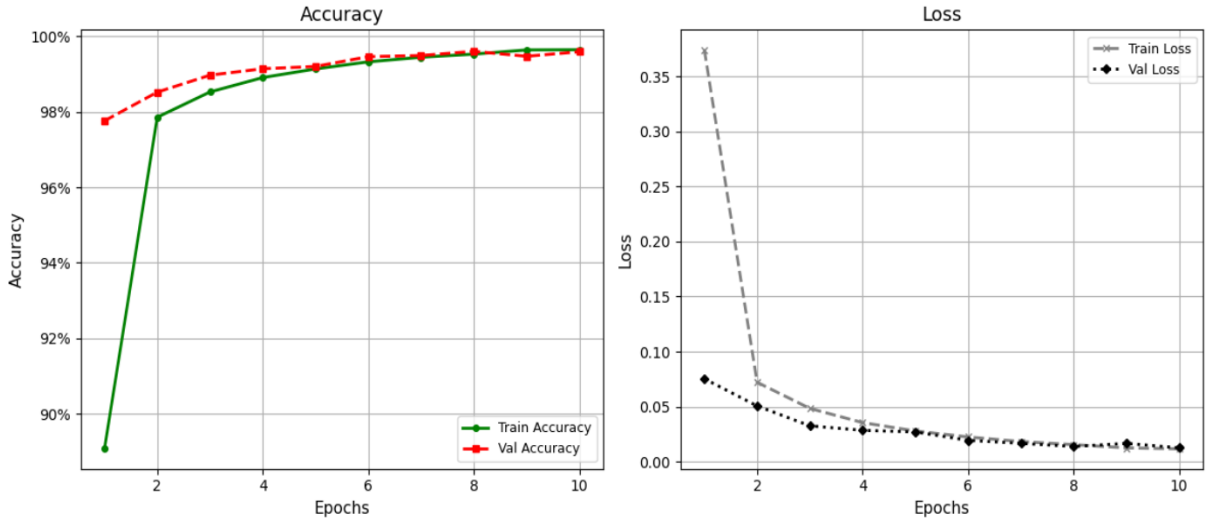


Figure 3: Swin Model with LoRA Accuracy and Loss Curves

DeiT Model Evaluation

DeiT Standard Fine-Tuning

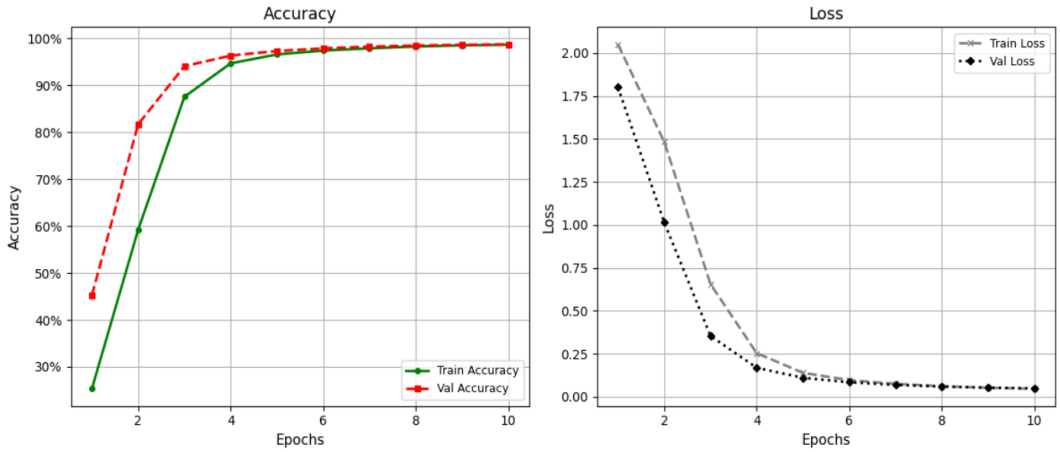


Figure 4: DeiT Fine-tuned Model Accuracy and Loss Curves.

DeiT with LoRA-Adapted Results

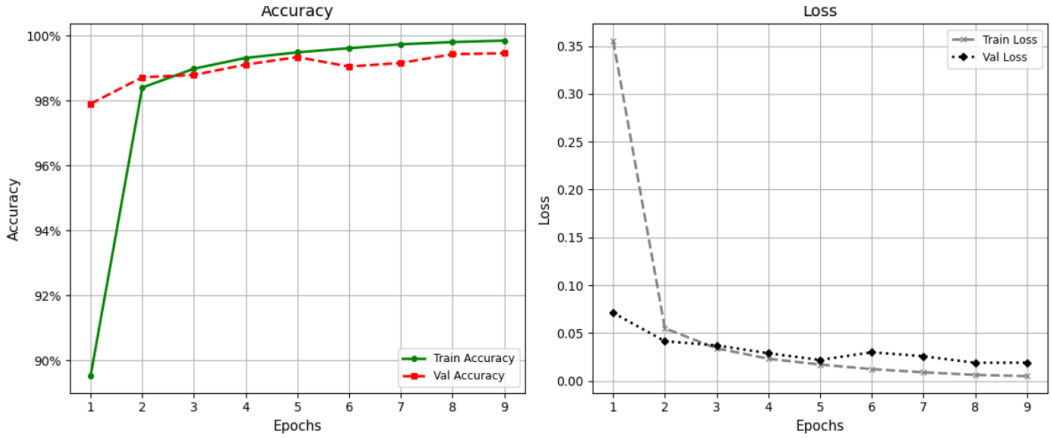
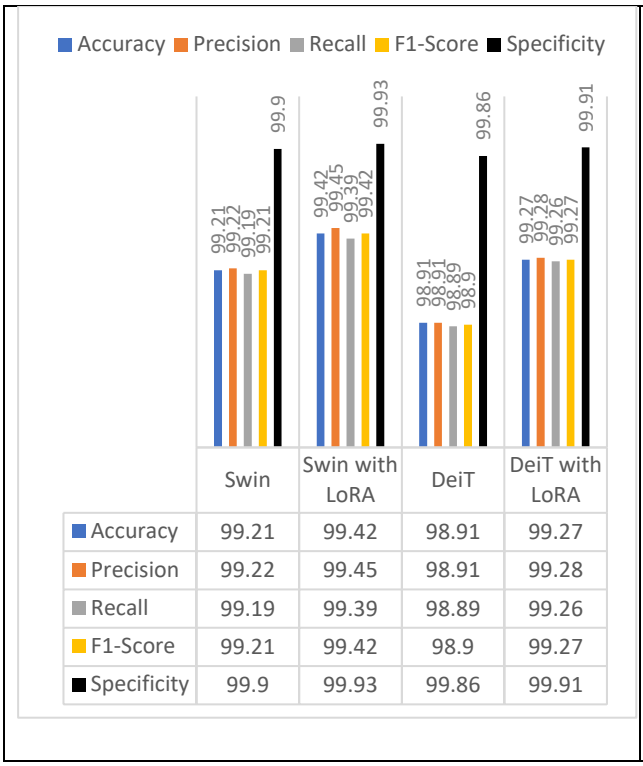


Figure 5: DeiT with Lora Accuracy and Losses Curves.

Models comparison

Table 2: Accuracy, Precision, Recall and Specificity using Test data

Model	Accuracy	ADI Precision, Recall, Specificity	BACK Precision, Recall, Specificity	DEB Precision, Recall, Specificity	LYM Precision, Recall, Specificity	MUC Precision, Recall, Specificity	MUS Precision, Recall, Specificity	NORM Precision, Recall, Specificity	STR Precision, Recall, Specificity	TUM Precision, Recall, Specificity
Swin Base	0.9921	0.9971, 0.9990, 0.9997	0.9991, 1, 0.9999	0.9940, 0.9898, 0.9992	0.9982, 0.9964, 0.9998	0.9921, 0.9843, 0.9992	0.9949, 0.9927, 0.9992	0.9850, 0.9895, 0.9986	0.9833, 0.9888, 0.9980	0.9853, 0.9880, 0.9976
Swin Base with LORA	0.9942	0.9990, 1.0000, 0.9999	1.0000, 1.0000, 1.0000	0.9907, 0.9941, 0.9988	1.0000, 0.9964, 1.0000	0.9955, 0.9854, 0.9996	0.9956, 0.9978, 0.9993	0.9919, 0.9907, 0.9992	0.9916, 0.9879, 0.9990	0.9860, 0.9930, 0.9977
DeiT Base	0.9891	0.9971, 0.9971, 0.9997	0.9972, 0.9991, 0.9997	0.9932, 0.9890, 0.9991	0.9982, 0.9964, 0.9998	0.9842, 0.9809, 0.9985	0.9868, 0.9897, 0.9979	0.9791, 0.9814, 0.9980	0.9831, 0.9795, 0.9980	0.9825, 0.9866, 0.9971
DeiT Base with LoRA	0.9927	0.9981, 0.9990, 0.9998	0.9991, 1.0000, 0.9999	0.9983, 0.9847, 0.9998	1.0000, 0.9991, 1.0000	0.9899, 0.9876, 0.9990	0.9941, 0.9941, 0.9991	0.9930, 0.9895, 0.9993	0.9760, 0.9879, 0.9971	0.9867, 0.9916, 0.9978



Model Name	Total Parameters	Trainable Parameters	Training time per epoch	Inference time	Accuracy (%)
Swin	86 M	86 M	17.55 mins	55 sec	99.21
Swin with LoRA	87 M	1.3 M	14.45 mins	55 sec	99.42
DeIT	85 M	85 M	10.9 mins	29 sec	98.91
DeIT with LoRA	86 M	1.2 M	9.5 mins	29 sec	99.27

Figure 6: Performance Evaluation

Conclusion and Future Works

Conclusion

It is found that the LoRA technique achieves even superior performance as compared to their full-fined tuned models for both Swin and DeiT transformers by reducing trainable parameters by more than 98%, and reducing training time significantly. For the Swin and DeiT models accuracy improved from 99.21% to 99.42% and from 98.91% to 99.27% respectively. The tiny added LoRA metrices during training can be merged back during testing time, having no impact on inference time as compared to conventional fine tuning. By using LoRA techniques rather than managing various large models for every problem domain, hospitals and labs can utilize a unique base model with its several smaller and lightweight LoRA adapters, which can considerably reduce computational costs, thereby, making it practically feasible for numerous medical tasks.

Future Works

Based on the success of LoRA for histopathological images, it is crucial to validate its efficiency and efficacy across other imaging domains such as radiological scans or satellite

images. This helps find out its generalizability. Hybrid methodological techniques are recommended to strengthen its accuracy in all imaging domains along with the standard LoRA, in which LoRA modifies the internal patterns of existing weight matrices where as other method like Adapters reshape the information flow between two consecutive layers, allowing orthogonal improvements. Furthermore, a dynamic rank allocation technique where some layers perform better in different rank while others excel at different values of rank because pretrained weights of some layers are more important than others for the specific problem domains.

References

Abou Baker, N., Rohrschneider, D., & Handmann, U. (2024). Parameter-Efficient Fine-Tuning of Large Pretrained Models for Instance Segmentation Tasks. *Machine Learning and Knowledge Extraction*, *6*(4), 2783–2807. https://doi.org/10.3390/make6040133

Afshin, P., Helminiak, D., Lu, T., Yen, T., Jorns, J. M., Patton, M., Yu, B., & Ye, D. H. (2025). *Breast Cancer Classification in Deep Ultraviolet Fluorescence Images Using a Patch-Level Vision Transformer Framework*. http://arxiv.org/abs/2505.07654

Aladhadh, S., Alsanea, M., Aloraini, M., Khan, T., Habib, S., & Islam, M. (2022). An Effective Skin Cancer Classification Mechanism via Medical Vision Transformer. *Sensors*, *22*(11). https://doi.org/10.3390/s22114008

Dai, Y., Gao, Y., & Liu, F. (2021). Transmed: Transformers advance multi-modal medical image classification. *Diagnostics*, *11*(8). https://doi.org/10.3390/diagnostics11081384

Demir, C., & Yener, B. (2005). *Automated cancer diagnosis based on histopathological images: a systematic survey*.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. http://arxiv.org/abs/2010.11929

Gallée, L., Lisson, C. S., Beer, M., & Götz, M. (2025). *Hierarchical Vision Transformer with Prototypes for Interpretable Medical Image Classification*. http://arxiv.org/abs/2502.08997

Hayat, M., Ahmad, N., Nasir, A., & Tariq, Z. A. (2024). Hybrid Deep Learning EfficientNetV2 and Vision Transformer (EffNetV2-ViT) Model for Breast Cancer Histopathological Image Classification. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3503413

Hong, Z., Xiong, J., Yang, H., & Mo, Y. K. (2024). Lightweight Low-Rank Adaptation Vision Transformer Framework for Cervical Cancer Detection and Cervix Type Classification. *Bioengineering*, *11*(5). https://doi.org/10.3390/bioengineering11050468

Ignatov, A., & Malivenko, G. (2025). *NCT-CRC-HE: Not All Histopathological Datasets Are Equally Useful*. https://github.com/gmalivenko/NCT-CRC-HE-experiments.

Kanadath, A., Angel Arul Jothi, J., & Urolagin, S. (2024). CViTS-Net: A CNN-ViT Network With Skip Connections for Histopathology Image Classification. *IEEE Access*, *12*, 117627–117649. https://doi.org/10.1109/ACCESS.2024.3448302

Khalid, M., Deivasigamani, S., V, S., & Rajendran, S. (2024). An efficient colorectal cancer detection network using atrous convolution with coordinate attention transformer and histopathological images. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-70117-y

Li, H., Zhang, Y., Chen, P., Shui, Z., Zhu, C., & Yang, L. (2024). *Rethinking Transformer for Long Contextual Histopathology Whole Slide Image Analysis*. http://arxiv.org/abs/2410.14195

Malekmohammadi, A., Badiezadeh, A., Mirhassani, S. M., Gifani, P., & Vafaeezadeh, M. (2024). *Classification of Gleason Grading in Prostate Cancer Histopathology Images Using Deep Learning Techniques: YOLO, Vision Transformers, and Vision Mamba*. http://arxiv.org/abs/2409.17122

Mármol, I., Sánchez-de-Diego, C., Dieste, A. P., Cerrada, E., & Yoldi, M. J. R. (2017). Colorectal carcinoma: A general overview and future perspectives in colorectal cancer. In *International Journal of Molecular Sciences* (Vol. 18, Issue 1). MDPI AG. https://doi.org/10.3390/ijms18010197

Preda, A.-A., Tăiatu, I.-M., & Cercel, D.-C. (2025). *Scaling Federated Learning Solutions with Kubernetes for Synthesizing Histopathology Images*. http://arxiv.org/abs/2504.04130

Qin, Z., Sun, W., Guo, T., & Lu, G. (2024). Colorectal cancer image recognition algorithm based on improved transformer. *Discover Applied Sciences*, *6*(8). https://doi.org/10.1007/s42452-024-06127-2

Ren, K., Hong, G., Chen, X., & Wang, Z. (2023). A COVID-19 medical image classification algorithm based on Transformer. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-32462-2

Romero, D. W., Knigge, D. M., Gu, A., Bekkers, E. J., Gavves, E., Tomczak, J. M., & Hoogendoorn, M. (2022). *Towards a General Purpose CNN for Long Range Dependencies in ND*. http://arxiv.org/abs/2206.03398

Träff, H. (2023). *Comparative Analysis of Trans-former and CNN Based Models for 2D Brain Tumor Segmentation*. www.liu.se

Tuli, S., Dasgupta, I., Grant, E., & Griffiths, T. L. (2021). *Are Convolutional Neural Networks or Transformers more like human vision?* http://arxiv.org/abs/2105.07197

Tummala, S., Kim, J., & Kadry, S. (2022). BreaST-Net: Multi-Class Classification of Breast Cancer from Histopathological Images Using Ensemble of Swin Transformers. *Mathematics*, *10*(21). https://doi.org/10.3390/math10214109

Uparkar, O., Bharti, J., Pateriya, R. K., Gupta, R. K., & Sharma, A. (2022). Vision Transformer Outperforms Deep Convolutional Neural Network-based Model in Classifying X-ray Images. *Procedia Computer Science*, *218*, 2338–2349. https://doi.org/10.1016/j.procs.2023.01.209

Vadori, V., Peruffo, A., Graïc, J.-M., Finos, L., & Grisan, E. (2025). *Mind the Gap: Evaluating Patch Embeddings from General-Purpose and Histopathology Foundation Models for Cell Segmentation and Classification*. http://arxiv.org/abs/2502.02471

Veasey, B. P., & Amini, A. A. (2025). Low-Rank Adaptation of Pre-trained Large Vision Models for Improved Lung Nodule Malignancy Classification. *IEEE Open Journal of Engineering in Medicine and Biology*. https://doi.org/10.1109/OJEMB.2025.3530841

Xin, Y., Yang, J., Luo, S., Zhou, H., Du, J., Liu, X., Fan, Y., Li, Q., & Du, Y. (2024). *Parameter-Efficient Fine-Tuning for Pre-Trained Vision Models: A Survey*. http://arxiv.org/abs/2402.02242

Yang, Z., Qiu, Z., Lin, T., Chao, H., Chang, W., Yang, Y., Zhang, Y., Jiao, W., Shen, Y., Liu, W., Fu, D., Jin, D., Yan, K., Lu, L., Jiang, H., & Bian, Y. (2024). *From Histopathology Images to Cell Clouds: Learning Slide Representations with Hierarchical Cell Transformer*. http://arxiv.org/abs/2412.16715

Yin, D., Han, X., Li, B., Feng, H., & Bai, J. (2023). *Parameter-efficient is not sufficient: Exploring Parameter, Memory, and Time Efficient Adapter Tuning for Dense Predictions*. https://doi.org/10.1145/3664647.3680940

Semantic Fidelity Over Speed: A Comparative Study of Conditional vs. Unconditional GANs for Few-Shot Skin Disease Classification

Prabeen Singh Airee [1], Sarbin Sayami[2]

[1]Faculty Member, Divya Gyan College, Tribhuvan University

[2]Central Department of Computer Science and Information Technology, Tribhuvan University

Article History

Received: 11/12/2025

Revised: 27/12/2025

Accepted: 10/01/2026

Published: 15/01/2026

Corresponding Author

Name: Prabeen Singh Airee

Email: prabeenairee@divyagyan.edu.np

ORCiD: https://orcid.org/0009-0002-0131-0791

Abstract

This research aims to break the crucial bottleneck of data deficiency in developing AI for skin disease diagnosis by presenting a clear and evidence-based comparison of two different generative AI models for data augmentation. By using a small clinical data set, this research systematically compared a fast but unconditional model of generative AI, FastGAN, with a semantic-aware model of generative AI, cGAN. Synthetic data generated by these models was used for training and testing different classifiers for different kinds of medical diagnoses. The results of this research clearly show that there is a crucial dependency of AI models for medical data augmentation on semantic fidelity. The cGAN model, which is semantic-aware and preserves class-specific features of skin diseases, enabled classifiers to retain high accuracy (up to 93%) for different kinds of medical diagnoses. However, in stark contrast, the unconditional FastGAN model, despite being much faster in generating synthetic data, catastrophically failed in retaining accuracy as low as 49% due to semantic inconsistency in data augmentation.

Keywords: ***Skin Disease Classification, Data Augmentation, Generative Adversarial Networks (GAN), Conditional GAN, FastGAN, Few-Shot Learning, Medical Image Analysis, Deep Learning.***

Introduction

A lot of people worldwide suffer from skin conditions; therefore, precise diagnosis is essential to successful treatment. The availability of large datasets with annotations is vital for the success of deep learning techniques, particularly for Convolutional Neural Networks (CNNs), which have shown revolutionary possibilities for automating dermatological image processing. Due to the restrictions on patient privacies, the expensive cost of expert annotations, and the inherent scarcity of many diseases, medical imaging is marked by considerable data scarcity. As a consequence, the development of dependable and broadly applicable models is hindered by the small and unbalanced sets of data available. Data Augmentation plays a critical role in alleviating this problem, and this need is met by the presence of an effective replacement, termed as "Generative Adversarial Networks" (GANs), which are capable of generating new data that is similar to the existing data, based on the underlying distribution of the data available. Not all GANs are suitable replacements, however.

The main architectural difference is that, in the case of unconditional GANs, the image is generated using random noise, while in the case of conditional GANs (cGANs), the image is generated using class labels, thus allowing for targeted generation.

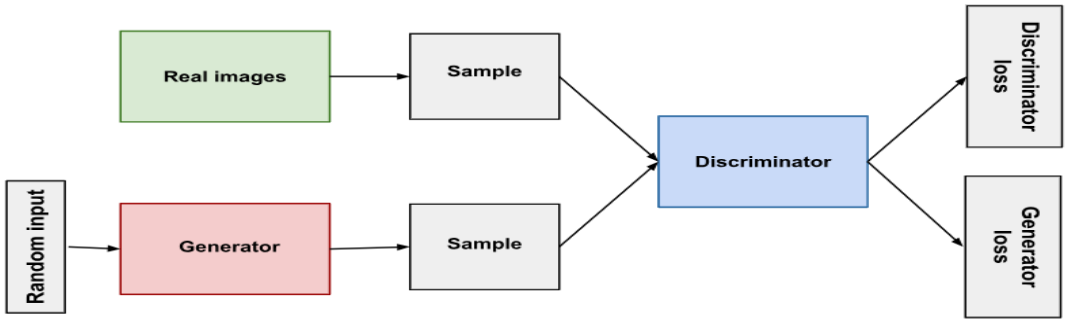


Figure 1: Architecture of Generative Adversial Network (GAN)

Recent developments have resulted in the formulation of new GAN variants, such as the FastGAN, designed for the stability and training speed of the model for use in small datasets (Biswas, 2023). Although the model is effective for the intended purpose, the absence of conditionality makes one ponder the applicability of the model in medical image synthesis, given the importance of disease-specific features. The cGAN, in turn, can be used as a DAGAN to address the class imbalance problem. Although the models have proved to be beneficial in

the literature, there is no empirical study that uses the models for augmenting the same small clinical skin dataset using both traditional and data-efficient learning strategies.

This study addresses this gap by posing the following research questions:

1. What is the comparative impact of FastGAN (unconditional) and cGAN (conditional) augmentation on the classification performance of ResNet models and a Few-Shot Learning framework?
2. Which paradigm supervised learning or Few-Shot Learning is more effective at leveraging GAN-generated synthetic data under extreme data limitations?
3. What is the optimal combination of generative strategy and classifier for small-scale clinical skin disease classification?

Our contribution is a controlled comparison research that proves that the speed of unconditional generation is not remotely as useful for medical picture augmentation as the semantic fidelity ensured by conditional generation. We prove that FastGAN-generated data can even be detrimental when used, especially for Few-Shot Learning, while cGAN-generated data maintains classifier performance.

The rest of the sections of this paper will be structured in the following manner: Section 2 will deal with Related work. The Appendix outlines the methodology in Section 3. Section 4 presents and discusses its results. Section 5 concludes the study and gives recommendations to the ongoing research.

Related Work

Medical Data Augmentation using GANs: Medical data is missing, which has elicited the application of GANs. (Montenegro et al., 2021) introduced a privacy-preserving GAN, and it is case-based explain ability. An example of a lightweight GAN suitable to few-shot synthesis was released (Liu et al., 2021) and can process high-resolution image consumption using fewer data samples. To enhance the data efficiency of GANs, Zhao et al. introduced Differentiable Augmentation also known as DiffAugment. The studies demonstrate the potential of GANs, though the authors do not provide the comparison of the two approaches.

Few-Shot Learning in Dermatology: Few-Shot Learning (FSL) is a low-data-setting algorithm. Meta-transfer learning was used in the skin disease classification in the study of (Özdemir et al., 2024) under long-tailed settings. A few-shot model named CDD-Net was introduced in (Chen et al., 2024) and entails the use of multi-scale feature fusion to diagnose clinical data. Although FSL is applicable to dermatology, past studies tend to use original or otherwise augmented data, rather than GAN-generated ones. CNNs will stay in the core of skin lesion classification using deep learning. used a hybrid ResUNet++ and AlexNet-RF model on classification and segmentation challenges. Concatenated VGG-16 and ResNet-50 to detect both benign/malignant skin lesions. ResNet is renowned in that it is a depth-based model, and addresses the issue of vanishing gradients.

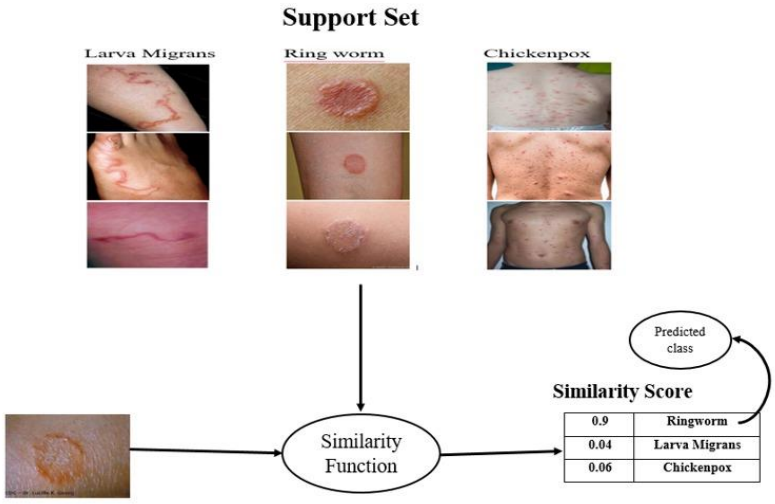


Figure 2: Architecture of FSL

The efficacy of the new unconditional GANs against their conditional counterparts, including their usage in supervised classifiers as well as in FSL classifiers, on the same clinical image set, has not been investigated although this has been demonstrated in earlier works regarding image augmentation and FSL using GANs and conditional GANs respectively. The paper addresses this research gap by offering guidelines of the selection of GAN variants, as the variants are used to generate FSL classifiers, to the cases of medical image datasets, which are affected by a lack of data.

Literature Review

The research (Özdemir et al., 2024) is focused on the classification of rare skin diseases in long-tailed data distributions by drawing a comparison between episodic and conventional training and few-shot learning and transfer learning. Using ISIC2018, Derm7pt, and SD-198 samples, the ImageNet pre-trained versions of DenseNet121 and a MobileNetV2 demonstrated significant improvements on limited labeled samples, demonstrating that traditional transfer learning using data augmentation performed better than the current methods, especially at 5-way classification and SD-198, and with a larger number of training examples, the classification rate increased.

CSDD-Net (Chen et al., 2024) is a plug-in module suggested by the authors to classify skin diseases on a few-shot basis, incorporating a context feature-fusion module to localize the details of lesions and a dual-attention mechanism to improve the relevant regions and channels and decrease the irrelevant areas. CDD-Net was validated by achieving an accuracy of +9.14 percentage points over baselines on the Derm104 dataset and ablation studies demonstrate the efficacy of both of its components.

The authors proposed a medical image classification privacy-preserving GAN that offers case-based interpretability in the form of factual explanations and counterfactual explanations and safeguards patient privacy by generating realistic images. The technique does not compromise personal identity as its explanatory power is found to be robust on both biometric and medical data. (Montenegro et al., 2021)

The authors (Liu et al., 2021) have created a lightweight GAN to perform few-shot image-based synthesis, which can produce high-resolution 1024x1024 images with low levels of data and computational-power. It trained in hours with a skip-layer excitation module and a self-supervised discriminator and achieved higher performance on thirteen datasets with limited data amounts than StyleGAN2.

In their work, the authors suggested Differentiable Augmentation (DiffAugment) (Zhao et al., 2020) which utilizes both real and generated images in differentiable augmentation to stabilize GAN training with small amounts of data. This avoids discriminator memorization and attains

state-of-the-art results, such as large dividing FID on datasets such as FFHQ and LSUN, with only 20 percent of training data.

The authors (Mustafa Et al., 2024) introduced a hybrid deep learning method of skin lesion analysis with hair removal preprocessing, ResUNet++ segmentation, and AlexNet-Random Forest classification. They demonstrated a higher level of segmentation and classification performance on HAM10000 as their approach combined the medical segmentation capabilities of ResUNet++ with the powerful classification capabilities of AlexNet-RF.

The authors developed a hybrid CNN model for skin lesion classification. They combined the optimized VGG-16 and ResNet-50 architectures. The hybrid model achieved an accuracy of 85.65% in the classification of dermoscopic images from the ISIC 2016-17 data set. This proves the concept of combining different models for better performance. (Gupta et.al, 2022)
The article (Mukherjee, 2022) gives an overview of ResNet-50, a neural network model. It explains the importance of ResNet in the field of deep learning. It also gives a detailed explanation of the ResNet residual connection, which prevents the problem of vanishing gradients.

Methodology

1. Overall Workflow

Figure 3 showed that within a comparative experimental design, the research was carried out. Preprocessing of clinical data set was first performed. Subsequently, FastGAN and cGAN two variants of GANs were trained to perform image synthesis. Three other classifiers, namely ResNet-18, ResNet-50, and Prototypical Network, were then trained in three conditions of data namely Baseline, +FastGAN, and +cGAN.

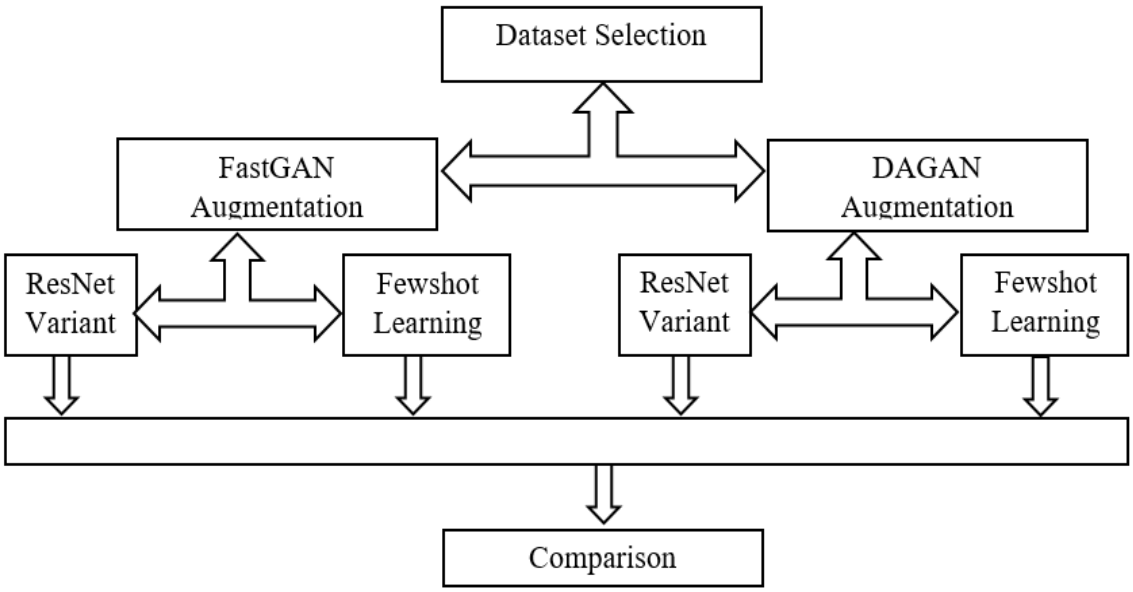


Figure 3: Research Workflow

2. Dataset and Preprocessing

It was based on the Skin Disease Dataset provided by Kaggle (Biswas, 2023), which included clinical images that have eight categories: bacterial diseases (Cellulitis, Impetigo), fungal diseases (Athlete's Foot, Nail Fungus, Ringworm), parasitic disease (Cutaneous Larva Migrans), and viral diseases (Chickenpox, Shingles). There were around 2,000 images that made up the dataset. The images were all resized to 224x224 and normalized. Training and testing were done according to an 80/20 split.

Table 1: Summary of skin Disease Dataset Categories

Disease Category	**Representative Diseases**
Bacterial infections	Cellulitis, Impetigo
Fungal infections	Athlete Foot, Nail Fungus, Ringworm
Parasitic infections	Cutaneous Larva Migrans
Viral Skin infections	Chickenpox, Shingles
Total Classes	8 distinct categories

To ensure consistency and compatibility with the deep learning models, all images underwent the following preprocessing pipeline:

Resizing: All images were resized to a uniform dimension of 224x224 pixels.

Normalization: Pixel values were normalized to a range of [0, 1] or [-1, 1] to stabilize and speed up the training process for the GANs and CNNs.

Data Splitting: The dataset was split into three subsets:

- Training Set (80%): Used to train the GANs and the classification models.
- Test Set (20%): Used only once for the final evaluation to report unbiased results. This set was completely hidden during all training phases

3. GAN Architectures for Augmentation

FastGAN Implementation: we used FastGAN architecture that was reported to rely on skip-layer channel-wise excitation in the generator and a multi-scale discriminator to train with small datasets with naturalness and efficiency. It was trained in the unconditioned mode with 50 epochs and Adam (lr=0.0002, 81=0.5, 82=0.999) to produce a balanced image set of all classes.

Conditional GAN (cGAN) Implementation:

We adopted a cGAN that was used in which class labels (embedded vectors) were inputted to both the generator and the discriminator. This conditions the generation on certain diseases to allow imbalanced classes to be oversampled. It was trained on the same hyper parameters as FastGAN to compare with it fairly with attention to minority classes generating supplemental images.

4. Classification Models

ResNet based supervised learning:

We used ResNet-18 and ResNet-50, which were fined tuned on our task of 8 classes, using ImageNet pre-training. The last fully-connected layer was substituted. The models were trained on Cross-Entropy loss and Adam (lr=1e-4, batch size=32) in 30 epochs.

Few-Shot Learning: Prototypical Networks:

Our training regime was a 5-way 5-shot episodic training scheme (Snell et al., 2017). The embedding network was a 4 layer CNN. The support embeddings of the prototypes were

computed on mean. In training, synthetic images were provided to help sets of underrepresented classes.

$$\boldsymbol{P_c = \frac{1}{|Sc|}\sum_{(x_i,y_i)\in Sc} f_0(x_i)}\text{.....(1)}$$

P_c -: Prototype vector for class c

S_c -: Support set for class c

The classification is done by choosing the closest prototype to the query point in the embedding space.

This metric space is used to query a sample of a prototype to classify it. Training of the model was done in stages. The support sets were used to add synthetic images on training to make them more diverse, especially on rare classes.

5. Experimental Setup & Evaluation

Three experimental data conditions were established:

1. Baseline: Original training data with traditional augmentations (flips, rotations).
2. FastGAN: Original data + synthetic images from the unconditional FastGAN.
3. cGAN: Original data + targeted synthetic images from the conditional GAN.

All models were evaluated on the same unseen test set using Accuracy, Precision, Recall, F1-Score, and Confusion Matrices.

Results and Discussion

1. Overall Performance Summary

The test accuracy for all models under the three data conditions is summarized in Table 1. The results reveal a consistent and striking pattern: performance with cGAN-augmented data, while lower than the high baseline, was consistently and significantly superior to performance with FastGAN-augmented data across all classifiers.

Table 2: Test Accuracy (%) for All Models Across Data Conditions. (Mean ± SD over 5 runs)

Models	**FewShot**	**ResNET 50**	**ResNET 18**
On Original Data	99% ± 0.3	94% ± 0.5	88% ± 1.1
On cGAN Augmented Data	93% ± 0.8	80% ± 1.2	76% ± 1.5

On Fast GAN Augmented Data	49% ± 2.1	68% ± 1.8	62% ± 2.0

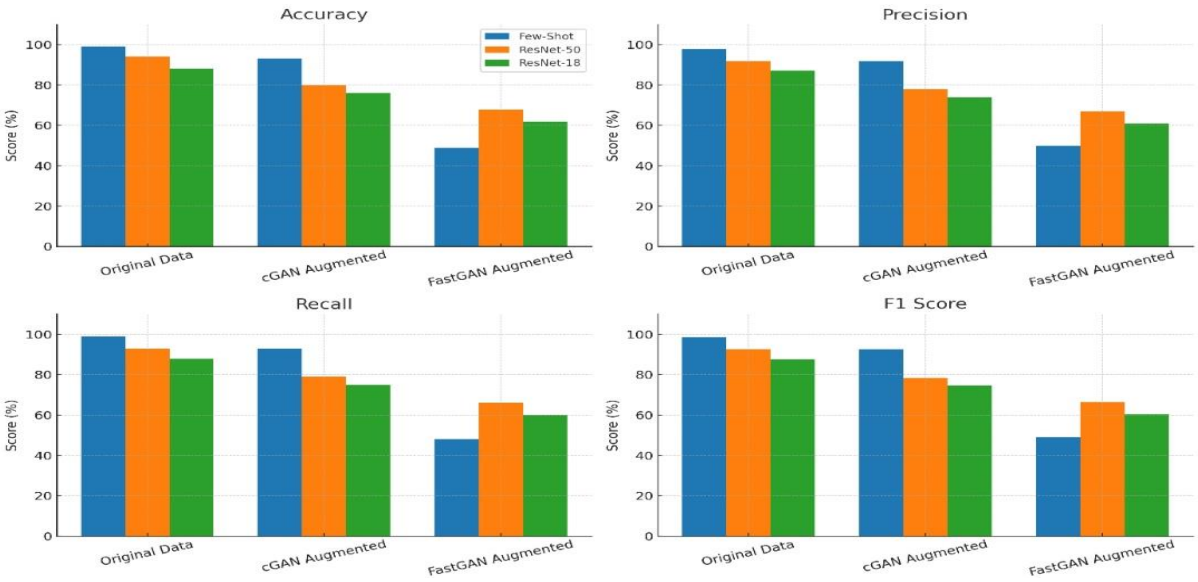


Figure 4: Comparision of Accuracy, Precision, Recall and F1-Score across models

2. Impact on Few-Shot Learning Performance

The highest extreme variance was found in the Prototypical Network (Fig. 4). It was almost perfect with 99% accuracy on the original data, and the confusion matrix is clean and there is great separation between classes. Accuracy using cGAN-augmented data was 93, but the confusion matrix showed that there were minor changes in inter-class confusion. Importantly, when using FastGAN-augmented data, performance reduced to 49% which is close to mere chance when dealing with an 8-class problem and the confusion matrix presented with many misclassifications. It means that the unconditional FastGAN produced features that were semantically incompatible with actual morphology of diseases, poisoning the metric embedding space on which few-shot classification was to occur.

3. Impact on Supervised Learning Performance

The ResNet models were similar in terms of the ordinal pattern: Baseline > cGAN > FastGAN (Fig. 4). As was anticipated, ResNet-50 performed better than ResNet-18 under all conditions because it is more profound. The fact that the performance between baseline and cGAN-augmented data declined by 13 percent and 12 percent, respectively, in ResNet-50 andResNet-18, indicates that cGAN data achieved valuable variance but might have introduced some

distributional shift or noise. Nonetheless, the decrease in accuracy with FastGAN-augmented data was much more dramatic (-26% with ResNet-50, -27% with ResNet-18), which validates the unconditional synthetic images had an element of features harmful to learning that distincts a robust decision boundary.

4. Discussion

The Excellence of Semantic Fidelity: The higher results of the cGAN compared to the FastGAN makes the paramount importance of semantic fidelity to medical AI. FastGAN is a fast and stable model that can generate unconstrained and feasible semantically plausible images. This may appear in the medical field as biologically implausible lesion textures, boundaries or color distributions fantasy features that are damaging noise during training. The model is a clean-conditioned cGAN and can be regularised to generate images based on the actual appearance of a particular class, thus producing more pedagogically useful synthetic data.

The High Sensitivity of Few-Shot Learners: Prototypical Network with FastGAN data catastrophic failure underscores the extreme sensitivity of metric-based learning to the quality of the data. The FSL models are based on the building of a continuous embedding space in which distance is related to semantic similarity. Loud or distracting artificial data disastrously interferes with this space, giving rise to incorrect prototypes and unsuccessful generalization. This renders FSL an effective but challenging paradigm, which needs genuinely pristine or sincerely augmented information.

Practical Implications and the challenge of High Baseline: The high baseline accuracies (89%-99%) on the original dataset would suggest that it was well structured and learnable This resulted in a ceiling effect so that it was hard to depict the simple improvement by augmentation. Thus, it is not important to find that augmentation was effective, but rather that cGAN augmentation did not affect performance but FastGAN augmentation was highly deleterious. This is an important lesson to practitioners: the conditional, targeted, generative model is a less risky, more stable option to use when augmenting small medical datasets than an unconditional one.

Limitations and Future Work: The research presented is limited to one clinical dataset and two GAN architectures. The findings on dermoscopic images and other areas of medicine should be supported by future work. It is reasonable to move on to more developed conditional generators, such as StyleGAN2-ADA or Latent Diffusion Models. Moreover, augmentation pipelines could be improved by creating ways of filtering or determining the usefulness of individual synthetic samples prior to incorporating them into training.

The excessive high baseline accuracies (i.e., up to 99 percent) indicate that there is a possibility of either the simplicity of the data sets, a lack of variability, or that the data is leaked, thus, affecting the validity and generalizability of the results. Even the absence of cross-dataset validation or stringent cross-validation prevents the confidence of the model to be applied to the real world.

Conclusion

The present research gives a direct, empirical response to the question concerning the GAN selection in medical image augmentation. Conditional GAN (cGAN) was superior to an unconditional, speed-optimized GAN (FastGAN) in a direct comparison on a small-scale, clinically-relevant task of classifying small clinical skin diseases, in both settings of standard supervised (ResNet) and Few-Shot Learning (Prototypical Network) classifiers.

The most important conclusion is that in medical use, the semantic faithfulness of conditional generation is inalienable. Although unconditional GANs provide performance, it is prone to be filled with semantically inconsistent samples that adversely affect model performance. This is especially disastrous to Few-Shot Learning paradigms. Thus, we firmly believe in the application of the selective, conditional generative models in increasing the limited medical data in constructing robust and dependable AI-based diagnostic instruments.

The intended work in the future will involve incorporating the most recent conditional generative models and building smart methods of synthetic data curation.

References

Biswas, S. (2023). *Skin-Disease-Dataset*. (Kaggle) Retrieved March 16, 2025, from https://www.kaggle.com/datasets/subirbiswas19/skin-disease-dataset

Chen et al. (2024). *Few-Shot Classification with Multiscale Feature Fusion for Clinical Skin Disease Diagnosis.* USA: National Library of Medicine (National Center for Biotechnology Information).

Gupta et.al. (2022). Skin Lesion Detection using VGG-16 and ResNet-50 based Hybrid CNN Model. *International Journal Of Creative Research Thoughts(IJCRT), 10*(6).

Liu et al. (2021). *TOWARDS FASTER AND STABILIZED GAN TRAINING FOR HIGH-FIDELITY FEW-SHOT IMAGE SYNTHESIS.* USA: Department of Computer Science, Rutgers University.

Montenegro et al. (2021). *Privacy-Preserving Generative Adversarial Network for Case-Based Explainability in Medical Image Analysis.* Porto, Portugal: Faculty of Engineering, Univercity of Porto.

Mukherjee, S. (2022, August 18). *The Annotated ResNet-50.* Retrieved June 18, 2025, from https://medium.com/data-science/the-annotated-resnet-50-a6c536034758

Mustafa Et al. (2024). *Deep learning-based skin lesion analysis using hybrid ResUNet++ and modified AlexNet-Random Forest for enhanced segmentation and classification.* Lahore, Pakistan: PLOS ONE.

Özdemir et al. (2024, April 25). *META-TRANSFER DERM-DIAGNOSIS: EXPLORING FEW-SHOT LEARNING AND TRANSFER LEARNING FOR SKIN DISEASE CLASSIFICATION IN LONG-TAIL DISTRIBUTION.* Retrieved from https://arxiv.org/abs/2404.16814

Zhao et al. (2020, June 18). *Differentiable Augmentation for Data-Efficient GAN Training.* Retrieved July 18, 2025, from arXiv:2006.10738

Predicting Channel Quality Indicator (CQI) in LTE Using Ensemble Learning Approaches

Marshal Babu Basnet[1]

[1]Faculty Member, Divya Gyan College, Tribhuvan University

Article History

Received: 18/12/2025

Revised: 07/01/2025

Accepted: 10/01/2026

Published: 15/01/2026

Corresponding Author

Name: Marshal Basnet

Email: marshalbasnet@divyagyan.edu.np

ORCiD: https://orcid.org/0009-0000-1430-2681

Abstract

Correct prediction of Channel Quality Indicator (CQI) is a key to successful link adaptation and resource allocation in Long-Term Evolution (LTE) and 5G New Radio (NR) networks. The paper presents a CQI prediction method based on an ensemble-based technique by using measurements on the LTE radio signals like Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), Received Signal Strength Indicator (RSSI), and Signal to Noise Ratio (SNR). The proposed method is evaluated against with individual machine-learning models since it combines complementary learners to improve prediction stability and accuracy. Result obtained from publicly available datasets of LTE show the proposed framework outperforms other competitors with an MAE of 0.66 and R^2 of 0.93. The ensemble would add to training and inference time, but prediction latency per sample (0.052 ms/sample) is much lower than LTE timing requirements, and thus quite practical to use in real time. In general, this work demonstrates that an ensemble-based method can provide a significant boost to the efficacy of CQI estimation, which can become a promising solution to effective scheduling and adaptive modulation of LTE systems.

Keywords: ***Channel Quality Indicator, LTE, ensemble learning, stacking, machine learning, link adaptation, CQI prediction***.

Introduction

Reliable Channel Quality Indicator (CQI) estimation is fundamental to efficient link adaptation, resource allocation, and modulation and coding selection in Long-Term Evolution (LTE) systems. Accurate CQI ensures that base stations select appropriate transmission parameters necessary for maintaining Quality of Service (QoS). In contrast, inaccurate or delayed CQI reporting leads to retransmissions, throughput degradation, and poor user experience, especially under fast-varying wireless conditions.(Elsherbiny et al., 2020; Rappaport, 2013).

LTE Traditional CQI estimate is founded on user equipment direct recording and reporting. Though effective, such methods can be ineffective in situations where fading is rapid, produce feedback load, and cannot rely on historical and contextual signal patterns. Consequently, learning-based approaches have achieved popularity since they could deduce CQI by using signal-quality measures that are measurable to the observer and nonlinear trends that are neglected by conventional estimators (Cordina & Debono, 2017; Vankayala & Shenoy, 2020). The literature such as Osiezagha, Mishra, and Fagbola (2025), Kim and Han (2023) are mainly about individual machine-learning models such as regression-based predictors, tree-based, or neural networks that demonstrate good results but are not very robust and generalized under different channel conditions. It is the gap, however, in building a framework that will not only increase accuracy, but also guarantee a consistent performance of prediction across situations.

To settle this, the current paper proposed a stacking-based ensemble model of CQI prediction and compared that with commonly used single learning models. The significant findings of this paper are:

1. Comparison of several machine-learning models based on publicly available LTE datasets.
2. establishment of a stacking group of learners that are complementary to each other and will be used to augment the robustness of CQI predictions.
3. performance evaluation based on accuracy and computational issues, and
4. evidence that the proposed ensemble is capable of enhancing prediction accuracy, but does not increase inference latency beyond the LTE timing constraints, which can be applied in real-time.

Literature Review

Machine learning is another widely used tool in predicting communication quality metrics of wireless network. The first methods used statistical techniques and rudimentary machine learning to predict Channel Quality Indicator (CQI) or similar measures, but tended to fail in nonlinear channel behavior and dynamic conditions (Soret et al., 2015). New developments prove that more advanced models can be used to capture more complex relationships between radio signal features and CQI.

Ensemble models like Random Forest and Gradient Boosting have been used in several studies like Pavan, M., & Reddy, B. R. (2022), Yang, Y., Zhang, H., Wang, W., and Li, Z. (2022), and have helped in quality prediction of wireless better than simple models. Nevertheless, these techniques continue to suffer a scaling or generalization deficit in a wide range of channel scenarios. In a few studies, ensemble-based models have been used to predict the quality of a wireless.

Okiemute Osiezagha et al. (2025) investigated supervised ensemble algorithms in the prediction of mobile network coverage and signal quality and found a better performance compared to that of single learners. Nevertheless, they concentrated on estimation of coverage in a university level test setting, but not CQI regression in a variety of LTE conditions. Similarly, Sharma et al. (2025) used machine learning to predict and optimize channel quality in 5G, indicating that learning based methods can be used to increase the accuracy of predictions. They however did not investigate stacked meta-learning strategies or latency in-depth scrutiny of real-time applicability in their study.

Bakri et al. (2020) established channel stability prediction using deep learning in 5G with a heavy emphasis on the flexibility to changing time-dependent wireless conditions. Most recently, Kim and Han (2023) designed a CNN, LSTM based CQI prediction model in vehicular communication and demonstrated an improvement in accuracy through the model with the help of temporal information. On a similar note, Ahmad (2025) developed a CQI prediction technique unique to 5G NR by comparing SNR to CQI mapping, but on a small scale, to one model and network condition, but it primarily focused on the design of algorithms as opposed to comparison of the models as families or trade-offs between computation and cost.

However, despite the research into deep learning and ensemble-based methods of CQI prediction in LTE systems, much of the current literature still adopts individual models, including CNNs or LSTMs, or simple ensemble strategies of averaging predictions. These approaches do not exploit the opportunities of integrating various categories of learners within one frame to the fullest. Moreover, previous research has mostly focused on the accuracy of prediction but rendered little focus on the training complexity and inference delay despite the fact that they are important in real-time scheduling and link adaptation in LTE networks. As a result, the literature on the use of stacked ensemble models to perform CQI prediction that combine a variety of base learners and are not only assessed based on their accuracy but also the appropriateness to use in real-time application has a distinct gap.

The gap that is filled in the current study is the suggestion of a stacking-based ensemble framework that uses any type of models such as gradient boosting models and neural networks to boost robustness and generalization. In contrast to the previous literature where either a deep learning approach was considered or the ensembles were not supplemented by any meta learning, the study measures the predictive performance and the computational efficiency to determine the feasibility in LTE transmission timing budgets. This puts the contribution as a viable and scalable solution towards real-time CQI prediction and adaptive modulation strategies in LTE systems.

METHODOLOGY

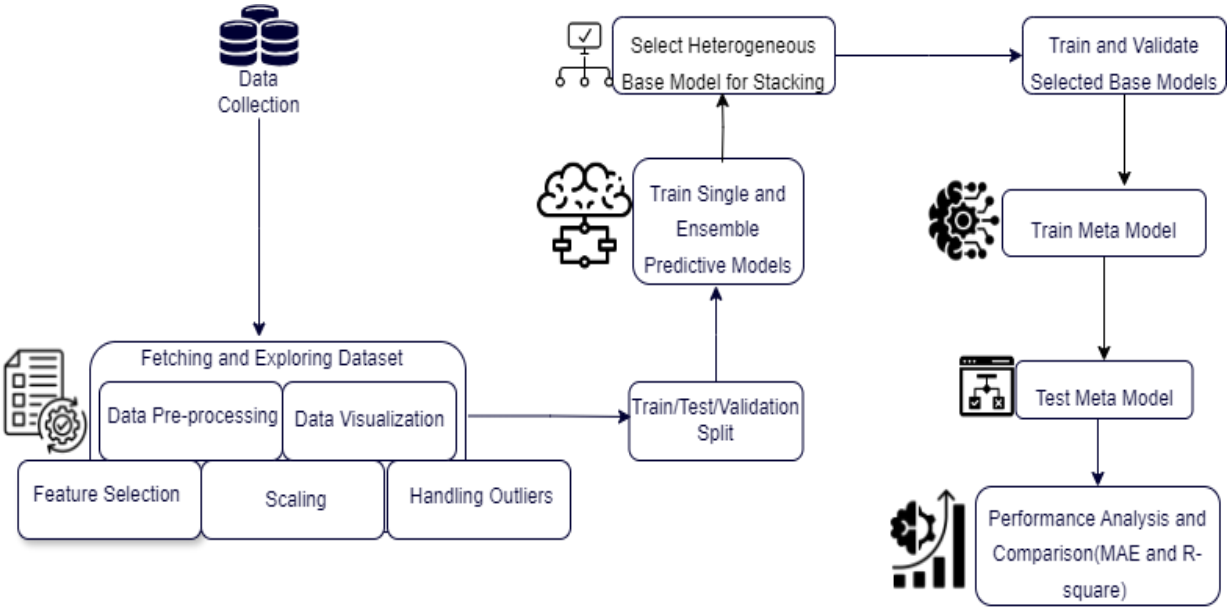


Figure 1: Block diagram of Research Methodology

Dataset

Kaggle made two publicly available datasets of LTE to train and test predictive models of Channel Quality Indicator (CQI). Netflix_v0 Dataset has 34,709 records of network measurements of video streaming. It includes such features as Timestamp, Location (longitude and latitude), RSRP (dBm), RSRQ (dB), SNR (dB), CQI (target), RSSI (dBm), and Downlink (DL) and Uplink (UL) bitrates (*Combine-All-the-Data-from-Lte-Dataset-into-One*, 2023).
LTE Combined Dataset consists of 102, 557 records and sixteen features that are Timestamp, Longitude, Latitude, Speed, Network Mode, RSRP, RSRQ, SNR, RSSI, DL and UL Bitrates, Serving Cell Information, and Path (transport mode)(*5G Dataset*, 2020).

Data Preparation

Preprocessing of the data was done to make sure that it was of quality and fit that was necessary to machine learning. The duplicate records and missing values were eliminated. Domain specific filtering used selected the measurements that fell within reasonable RF ranges (see Table 1), trimming the aggregate data set by more than 31,000 records. Quantile-based procedures identified and eliminated the outliers to ensure that the model was not trained on extreme values.

Histograms and Q-Q plots to evaluate the distribution of features were part of the exploratory data analysis, and box plots were used to visualize and verify the removal of outliers (Kang & Tian, 2018; Moltchanov, 2019)

Table 1: RSRP, RSRQ and SNR values standard range

RSRP (dBm)	**RSRQ (dB)**	**SNR (dB)**	**CQI**
≥ -80	≥ -10	≥ 20	15
-90 to -80	-15 to -10	13 to < 20	12 - 14
-100 to -90	-20 to -15	0 to < 13	5 - 11
< -100	< -20	≤ 0	≤ 4

The characteristics with a substantial correlation with CQI were prioritized to be incorporated into the modeling procedure to make sure that the analysis uses the most significant predictors (Denis, 2021).

Such pretreatment ensured that the data was clean and comparable across all instances and acceptable in terms of modeling. The data were thoroughly prepared through the usage of powerful statistical tools and the knowledge of the field in order to develop accurate and reliable models of prediction of CQI. Data protection in machine learning is the core aspect because it directly influences the extent to which the model can learn on the data and give realistic predictions(Kang & Tian, 2018).

Machine Learning Models

Three baseline models were chosen, Linear Regression (LR), Gradient Boosting Regressor (GBR) and Feed-forward Neural Network (FNN), to compare CQI prediction on a variety of learning paradigms. It was specifically these models that were selected since they form a set of complementary classes of hypothesis, all capable of explaining different aspects of the behavior of LTE signals. Given that CQI values were 1-15, the initial inspection revealed that various models can yield more predictable results within various CQI sub-ranges, implying that the integration of both could give more predictable results across the entire range.

Linear Regression offers an easily understandable base-line and works well in cases where the correlation between radio-signal indicators and CQI is roughly linear. Although was less effective in the models of non-linear variations of channels, whereas LR sets a baseline against which better models can be evaluated (Su et al., 2012).

Gradient Boosting Regressor was chosen because it can be used in modeling non-linear patterns with sequential tree boosting. GBR eliminates residual errors progressively and has been very effective in wireless prediction problems where the relationships between RSRP, RSRQ, RSSI and SNR are non-linear (Natekin & Knoll, 2013).

Feedforward Neural Network represents complex interactions of features with many hidden layers. FNN in contrast to LR and tree-based methods learns hierarchical relation-ships hence being appropriate in areas where channel behavior changes fast. This array of learning behavior inspired the move towards ensemble approach. (*Effiecient Learning Machines*, 2017.; *Ensemble Learning in Machine Learning: Stacking, Bagging and Boosting*, 2023; Wolpert, 1992).

Stacking Ensemble

A stacking ensemble was created to enhance the strength and combine the strengths of single models. LR, GBR, and FNN are used as base learners and another FNN is used as a meta-learner. The selection of this architecture was made on two notes:

1. There were only specific CQI value ranges in which individual models gave reliable results.
2. Theoretical discrepancies in the generalization of errors by the models indicated that a combination of the models could fit a wider range of performance.

The piling process had a two-phase training process.

In Stage-1, the training was independently trained on each base learner on training set. Out of-fold (OOF) predictions were created in order to prevent information leakage with the help of k-fold cross-validation. These OOF predictions create a new intermediate feature space that is a reflection of the interpretation of the input by each model.

Stage-2: the stacked OOF output was used as input features to train the meta-learner (FNN). By training on learning to make generalizations between base model predictors as opposed to

raw features, the meta-learner is able to learn optimum weighting and interaction relationships, leading to better generalization.

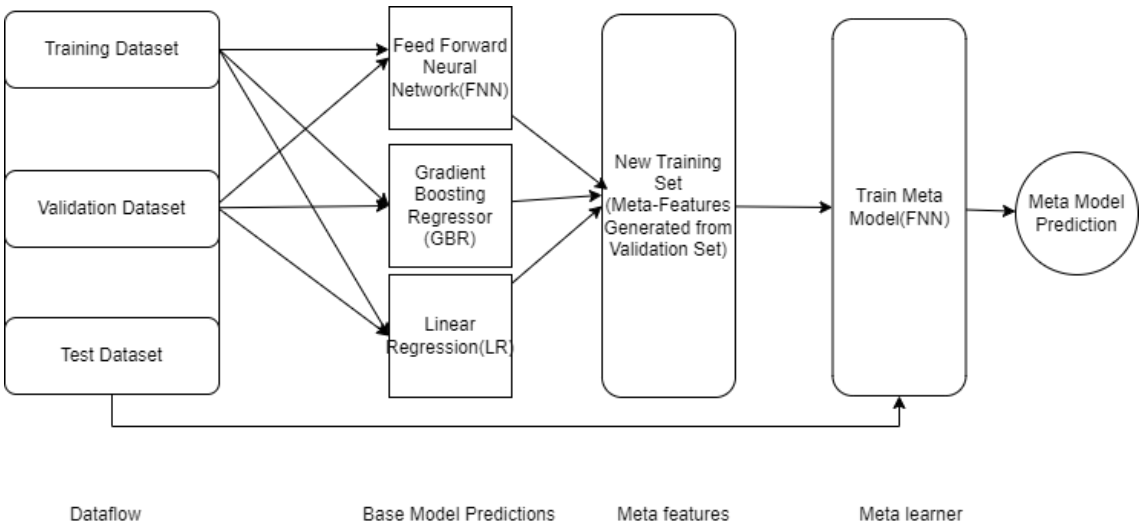


Figure 2: Proposed Stacking Model

Data Splitting

Table 2 reveals data splitting strategy. This study used 13, 268 LTE samples. In the case of training the base models, 80/20 split was implemented with 10,614 samples being used to train LR, GBR and FNN, and 2,654 samples being used to create out-of-fold predictions to use in the stacking process. The base models generated a prediction on each sample to give a meta-feature matrix of 2,654 by 3. This meta-dataset was again di-vided into 60 training, 20 validation and 20 test of the meta-learner (FNN) to provide unbiased learning and sound evaluation.

Table 2: Dataset split used for baseline and stacking experiments

Experiment Stage	**Train (%)**	**Validation (%)**	**Test (%)**	**Purpose**
Base Model Evaluation	80	—	20	Benchmark performance
Stacking Ensemble	60	20	20	Meta-learner training + final evaluation

Hyperparameters tuning

Hyperparameters of model tuning are presented in Table 3. The grid search of the 5-fold cross-validation was applied to optimize the learning rate, depth, and the width of the layers. Weights of the neural networks were initialized with Glorot Uniform (seed=42) and trained with the Adam optimizer with MAE loss.

Table 3: Hyperparameter settings for all learning models.

Model	Key Hyperparameters
Linear Regression	Polynomial degree=2 (via pipeline)
Gradient Boosting Regressor	n_estimators=200, learning_rate=0.05, max_depth=4, subsample=0.8
FNN (Base Learner)	Layers: 64→32→1, activation=ReLU, epochs=100, batch_size=32
FNN (Meta-Learner)	Layers: 64→32→1, activation=ReLU, optimizer=Adam, epochs=100

Overfitting Prevention

A number of steps were undertaken to minimize overfitting. The FNN models included dropout layers and L2 regularization of dense layers. Validation loss was used to monitor meta-learner training and early stopping was used to avoid unjustified epochs. Splitting K-folds in the generation of meta-features also provided robust learning.

Evaluation Metrics

A model performance was considered as a set of quantitative parameters and this considers both the predicted accuracy and the computational efficiency. Mean Absolute Error (MAE) and coefficient of determination (R^2) were used to measure prediction performance. MAE is calculated using the absolute differences between predicted and actual CQI values, and is used to form an indication of the average magnitude of the prediction errors. Lower values of MAE indicate more precise estimation and it is particularly effective in regression problems whose targets are discrete as in CQI. R^2 is the proportion of the change in the target variable that the model explains, which gives a more in-depth view of predictive fit.

In addition to accuracy, computational performance was assessed through training duration and prediction delay. The training period is the overall time necessary to fit a model to the training dataset, including preparation and parameter tuning. Prediction latency is the time it takes to develop CQI estimations for previously unseen data and is a critical concern for real-time or near-real-time LTE applications. These measures, when combined, allow for a comprehensive model comparison that balances accuracy with practical practicality(Willmott & Matsuura, 2005).

Result and Discussion

The predictive models on estimating CQI were tested with both single algorithms LR, Ridge, Lasso, SVR, FNN, LSTM, and GRU and ensemble algorithms, such as Bagging Regressor, GBR and the suggested stacking ensemble. The measures of accuracy of the model were MAE and R^2, whereas the computational efficiency was determined as the memory consumption, training time and time to predict a sample.

Table 4: Comparison of MAE and R-square of different algorithms

Model	**MAE**	**R-Square**
Linear Regression	0.92	0.92
Ridge Regression	0.87	0.93
Lasso Regression	1.2	0.88
LSTM	0.74	0.91
GRU	0.72	0.91
Feedforward Neural Network	0.76	0.94
Support Vector Regressor	0.76	0.94
Bagging Regressor	0.77	0.89
Gradient Boosting Regressor	0.73	0.91
Stacking (Proposed)	0.66	0.93

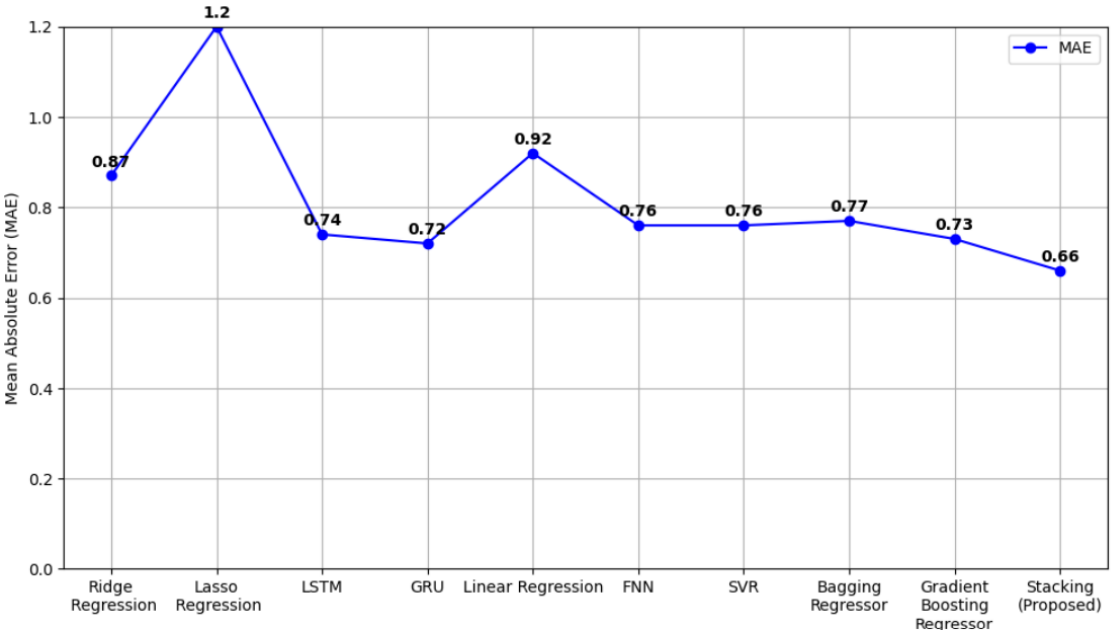


Figure 3: Mean absolute error comparison of different machine learning algorithms.

Table 2 provides an overview of the best predictive performance of the stacking ensemble as it has the lowest MAE (0.66) and high R^2 (0.93), which shows that the ensemble has better predictive performance. Base models that were used in the stacking ensemble (LR, GBR, and FNN) were selected based on their complementary strengths of capturing linear, non-linear, and complex interaction. The predicted CQI values (Table 3) are close to the actual values, especially in terms of mid to high range CQI. The lower CQI levels of 1-2 demonstrated more deviation, probably because of changing network conditions and data imbalance because there were less low CQI samples.

Table 5: Actual and Predicted CQI of proposed meta model

Actual CQI	**Avg. Predicted CQI (meta model)**
1	3.32
2	3.12
3	3.55
4	3.83
5	6.86
6	7.34
7	7.3
8	7.45

Actual CQI	Avg. Predicted CQI (meta model)
9	9.82
10	10.9
11	11.08
12	11.64
13	14.71
14	14.78
15	14.86

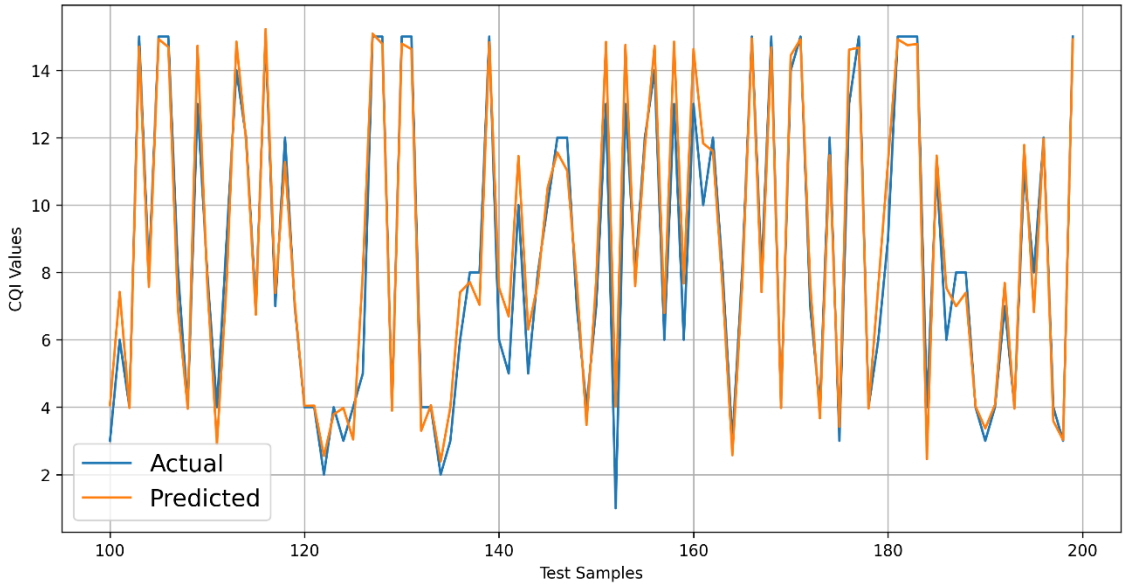


Figure 4: Actual vs Predicted value of Proposed Stacking Model

The stacking ensemble had a longer training time (64.3 s) and memory (0.43 MB) requirement than single models, but the prediction time (0.052 ms/sample/sample) was still viable in order to deploy the model in real-time (Table 4). Single models (LR and Ridge) were quicker to train but with greater MAE whereas GBR and FNN had superior accuracy with small computing expense. These findings indicate the trade-off between predictive accuracy and computational efficiency and affirm that the stacking method is an effective way to combine the bests of different models to provide effective CQI prediction in a wide range of network scenarios.

Table 6: Performance Comparison of Regression, Neural and Ensemble Models

Model	**Memory Usage (MB)**	**Prediction time (ms/Sample)**	**Training Time (s)**	**Laptop specs**
Linear Regression	0.02	0.00037	0.00228	11th Gen Intel Core i5-4 cores, 8GB RAM
Ridge Regression	0.02	0.00029	0.00267	11th Gen Intel Core i5-4 cores, 8GB RAM
Lasso Regression	0.02	0.00043	0.00277	11th Gen Intel Core i5-4 cores, 8GB RAM
LSTM	0.41	0.13	9.6	11th Gen Intel Core i5-4 cores, 8GB RAM
GRU	0.38	0.12	7.2	11th Gen Intel Core i5-4 cores, 8GB RAM
Feedforward Neural Network	0.33	0.09	5.62	11th Gen Intel Core i5-4 cores, 8GB RAM
Support Vector Regressor	0.04	0.031	0.48	11th Gen Intel Core i5-4 cores, 8GB RAM
Bagging Regressor	0.048	0.019	0.48	11th Gen Intel Core i5-4 cores, 8GB RAM
Gradient Boosting Regressor	0.02	0.0015	0.17	11th Gen Intel Core i5-4 cores, 8GB RAM
Stacking (Proposed)	0.43	0.052	64.3	11th Gen Intel Core i5-4 cores, 8GB RAM

Altogether, the suggested stacking ensemble has great potentials in the practical LTE system implementation, offering adequate CQI estimation with reasonable computation needs. Low CQI predictions deviation hints at the future research paths, including the possibility of increasing the sample of low-CQI or attaching other context-related capabilities to the models to improve the reliability even more.

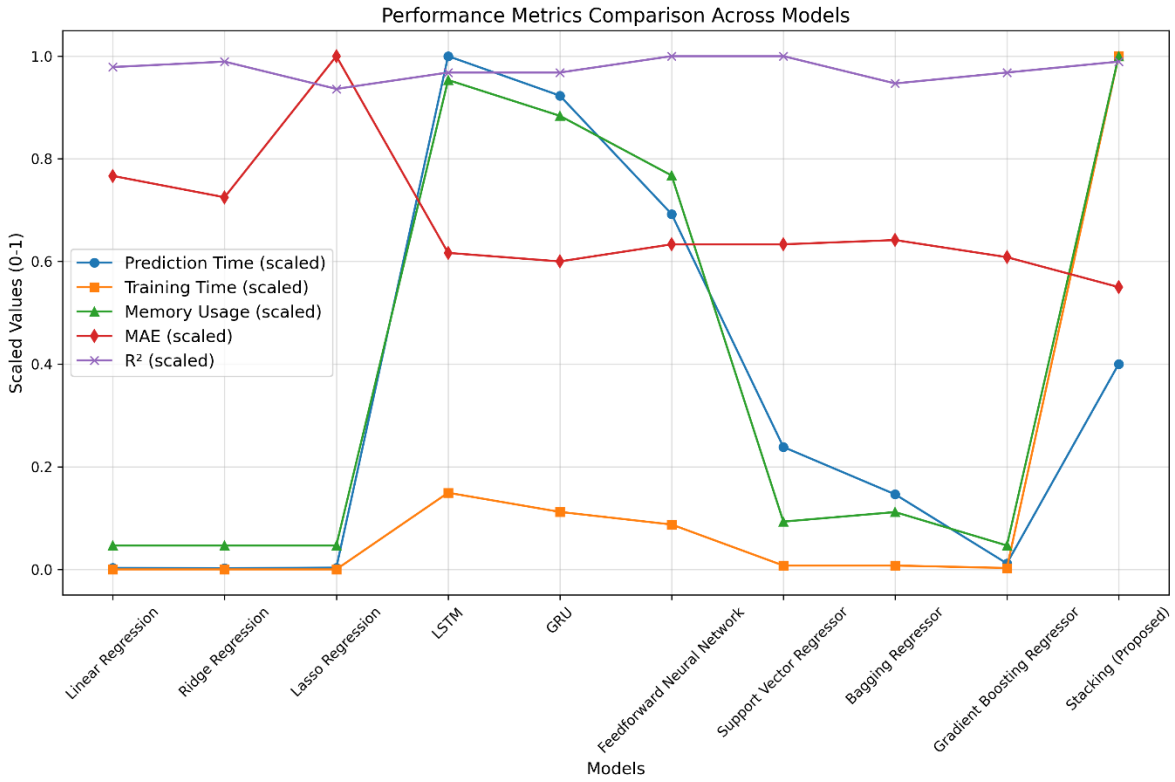


Figure 5: Performance metrics comparison of different machine learning model

Training vs Validation Loss

Figure 6 shows the training and validation curves of stacking ensemble model. As both curves approach each other with a smooth decline and converged without much variation, there is a stable process of learning the curves depict. This trend suggests that the ensemble has high generalization to novel data and achieves success in modeling the underlying associations of the LTE signal features. Additional indicators of weak overfit-ting are the absence of a large distance between the curves indicating that the meta-learner is successful in adopting the predictive capabilities of the base models. In general, the trends of training-validation confirm that the stacking framework provides a reliable and balanced learning behavior that suits CQI prediction activities.

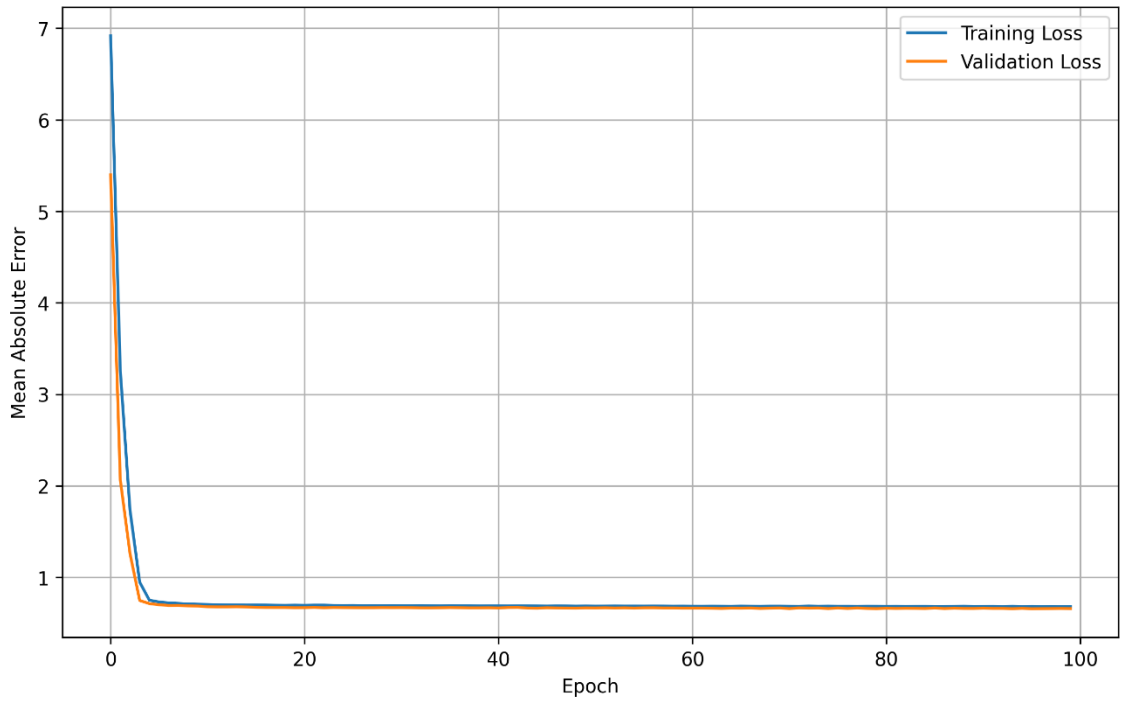


Figure 6: Showing Validation Loss Vs Training Loss of Proposed Stacking Model

Conclusion

This paper examined predictive modelling of the Channel Quality Indicator (CQI) in LTE networks, first individually with machine learning algorithms, and then in ensembles. The empirical findings prove that the suggested stacking ensemble, consisting of Linear Regression, Gradient Boosting Regressor, and Feedforward Neural Network as a base model together with an FNN meta-learner, is always more accurate and robust in its performance compared to a single model. The stacking model produced the least Mean Absolute Error (0.66), high R-squared (0.93) with realistic prediction times to be deployed in real-time.

The results indicate the benefits of ensemble methods when predicting CQI especially in handling of complex non-linear and linear associations among network attributes. Although the predictions of low CQI values were also more varied because of the imbalance between the datasets and the changing nature of the network conditions, the model still proved to be highly useful in the mid- to high-range CQI values.

Further improvements of low-CQI prediction deviations can be made in future work either by boosting the underrepresented samples or adding other context-sensitive features. Additional

studies may also be done on deployment in live LTE systems to confirm real-time functionality and test adaptive functionality in dynamic network environments. Altogether, the suggested ensemble method offers a strong framework of precise and effective CQI prediction and enhances the process of distributing resources and optimizing networks in LTE settings.

References

Yang, Y., Zhang, H., Wang, W., & Li, Z. (2022). Deep learning-based channel quality indicator prediction for improved wireless performance. Electronics, 11(4), Article 626. https://doi.org/10.3390/electronics11040626

Pavan, M., & Reddy, B. R. (2022). Machine learning based prediction for channel stability in a 5G network. In Proceedings of the 2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon). IEEE. https://doi.org/10.1109/NKCon56289.2022.10126726

Sharma, S., Sharma, O. P., & Shekhawat, S. (2025). *Channel quality prediction,ehancement & optimization of 5G network using machine learning. IET Conference Proceedings*, Article 0902. The Institution of Engineering and Technology. https://doi.org/10.1049/icp.2025.0902
Morrison Okiemute Osiezagha, B. K. Mishra, & T. M. Fagbola. (2025). Ensemble supervised learning-based approaches for mobile network coverage and quality predictions in a university setting. In *Proceedings of [Conference Name]* (pp. 3–26)

Kim, J., & Han, D. S. (2023). Deep learning-based channel quality indicators prediction for vehicular communication. *ICT Express*, *9*(1), 116–121. https://doi.org/10.1016/j.icte.2022.05.002

Ahmad, M. N. (2025). *CQI prediction method for 5G NR cellular systems* (Master's thesis, Tampere University). Trepo. https://urn.fi/URN:NBN:fi:tuni-202506056777

5G dataset. (2020). *Dataset*. GitHub. Retrieved Month Day, 2025, from https://github.com/uccmisl/5Gdataset

Combine-all-the-data-from-LTE-dataset-into-one. (2023). *Dataset*. Kaggle. Retrieved Month Day, 2025, from https://www.kaggle.com/code/liamghogan/combine-all-the-data-from-lte-dataset-into-one

Comprehensive guide for ensemble models. (2018). Analytics Vidhya. Retrieved Month Day, 2025, from https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models

Cordina, M., & Debono, C. J. (2017). A support vector machine based sub-band CQI feedback compression scheme for 3GPP LTE systems. *2017 International Symposium on Wireless Communication Systems (ISWCS)*, 325–330. https://doi.org/10.1109/ISWCS.2017.8108134

Denis, D. J. (Ed.). (2021). *Applied Univariate, Bivariate, and Multivariate Statistics Using Python: A Beginner's Guide to Advanced Data Analysis* (1st ed.). Wiley. https://doi.org/10.1002/9781119578208

Kurniawan, B., & Idris, M. (Eds.). (2017). *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Apress. https://doi.org/10.1007/978-1-4302-5990-9

Elsherbiny, H., Abbas, H. M., Abou-zeid, H., Hassanein, H. S., & Noureldin, A. (2020). 4G LTE Network Throughput Modelling and Prediction. *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 1–6. https://doi.org/10.1109/GLOBECOM42002.2020.9322410

Ensemble learning in machine learning: Stacking, bagging and boosting. (2023). Analytics Vidhya. Retrieved Month Day, 2025, from https://www.analyticsvidhya.com/blog/2023/01/ensemble-learning-methods-bagging-boosting-and-stacking/

Kang, M., & Tian, J. (2018). Machine Learning: Data Pre-processing. In M. G. Pecht & M. Kang (Eds.), *Prognostics and Health Management of Electronics* (1st ed., pp. 111–130). Wiley. https://doi.org/10.1002/9781119515326.ch5

Moltchanov, D. (2019). CQI-MCS-and-SNR-mapping-for-3GPP. *CQI-MCS-and-SNR-Mapping-for-3GPP*. https://www.researchgate.net/figure/CQI-MCS-and-SNR-mapping-for-3GPP-NR_tbl1_335767868

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*. https://doi.org/10.3389/fnbot.2013.00021

Rappaport, T. S. (2013). *Wireless Communications—Principles and Practice, Second Edition. (The Book End)*.

Soret, B., Pedersen, K. I., Jørgensen, N. T. K., & Fernández-López, V. (2015). Interference coordination for dense wireless networks. *IEEE Communications Magazine*, *53*(1), 102–109. https://doi.org/10.1109/MCOM.2015.7010522

Vankayala, S. K., & Shenoy, K. G. (2020). A neural network for estimating CQI in 5G communication systems. In *2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)* (pp. 1–5). https://doi.org/10.1109/WCNCW48565.2020.9124744

Vankayala, S. K., & Shenoy, K. G. (2020). A Neural Network for Estimating CQI in 5G Communication Systems. *2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 1–5. https://doi.org/10.1109/WCNCW48565.2020.9124744

Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*, 79–82. https://doi.org/10.3354/cr030079

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1