Deep Learning-Driven Classification of 5G and LTE Signals for Next-Generation Wireless Networks

1st Rajendra Paudyal
George Mason University, VA, USA
rpaudyal@gmu.edu

2nd Duminda Wijesekera
George Mason University, VA, USA
dwijesek@gmu.edu

***Abstract*—Cellular Radio Frequency (RF) spectrum monitoring and analysis are crucial for identifying and mitigating interfering and disruptive RF signals. In this study, we use Deep Learning (DL) models, including Convolutional Neural Networks (CNNs) and Transformer Networks (TNs), to classify and identify these signals. The primary objective is to classify cellular RF signals to aid in mitigating interference, thereby enhancing security and quality of service (QoS). Two models, ResNet50 [1] and ViT [2], are evaluated for their accuracy in classifying 5th generation New Radio (5G NR), 4th generation long-term evolution (4G LTE), and combined LTE-NR cellular signals. Both models demonstrate high true positive rates, exceeding 95% across all classes. However, ViT consistently outperforms ResNet50, showcasing superior capability in capturing distinguishing features and more accurately classifying RF signals. Notably, ViT's perfect true positive rates for the LTE and NR classes underscore its robustness and potential in classification and signal identification tasks, however ViT is much more resource intensive.**

***Index Terms*—RF Signal Classification, RF Spectrum, Deep Learning, ResNet50, ViT, NR, LTE, 5G, RF, Cellular**

I. Introduction

Radio Frequency (RF) is a vital electromagnetic resource that requires constant monitoring to maintain high-quality service (QoS) in wireless communications. The increasing demand for spectrum, particularly with the rollout of 5G and beyond (5G+), significantly heightens the risk of signal interference, which threatens both network performance and reliability [3]. Classification of 5G and LTE signals is essential in modern wireless communication research, enabling network operators to optimize resources, enhance performance, and improve security. Accurate signal classification helps detect anomalies, prevent unauthorized access, and provides insights into the characteristics of these technologies, supporting the development of advanced communication systems. Traditional RF spectrum monitoring methods, however, are often cumbersome and inadequate for the complexity of today's networks. To address these limitations, we propose deep learning (DL) methods for 4G and 5G RF signal classification, aligning with the U.S. shift towards these technologies [4], [5]. Unlike conventional approaches, which rely on manual feature extraction and demand domain expertise, DL methods offer automated, adaptable solutions suited for today's dynamic RF environments [6]–[8]. Deep neural networks (DNNs), especially Convolutional Neural Networks (CNNs), have shown higher accuracy and efficiency in RF signal classification than traditional techniques [9]–[11]. The primary objective of this paper is to evaluate and determine which of the two DL model ResNet50 or ViT is more suitable for the classification of cellular RF signals. Additionally, the study aims to assess if the outputs from these models can be effectively utilized to identify and mitigate sources of signal interference in 5G $5G_NR$, LTE, and $NR - LTE$ networks. The primary contribution and novelty of this papers are:

1) This work is the first to utilize spectrogram images generated from 5G NR signals, incorporating all numerology features (eg. subcarrier spacing, symbol duration). This is in contrast to previous research which focused primarily on classifying cellular RF signals using I/Q symbols, modulation, and constellation types. Our work allows for more precise detection and differentiation of 5G signals under diverse numerology configuration, and provides a richer and more detailed time-frequency representation of 5G NR signals, thereby enhancing the classification accuracy.
2) Trained models validation is done with real world cellular RF signals.

The remaining sections of this paper are organized as follows. Section II describes related work of applying of DL in RF signal identification. Section III describes the system model and scientific method applied in this effort. Section IV shows our simulation outputs and Section V compares the output of our DL models. Section VI summarizes this study, and Lastly Section VII suggests a potential future work.

II. Related Work

DL has become a promising approach in the field of wireless communications by demonstrating its effectiveness in tasks like signal identification, channel estimation, channel coding, and resource allocation [12]. Modulation recognition has an extensive exploration of neural network architectures, particularly CNNs and Recurrent Neural Networks (RNNs). CNNs excel in extracting local features from time-frequency representations like spectrograms, leveraging their spatial feature extraction capabilities. O'Shea et. al [13] conducted a comprehensive study demonstrating the potential of DL for modulation classification. This research, along with other studies [12], [14] utilized *In-phase/quadrature (I/Q)* radio signal samples to train CNNs and residual neural networks (ResNet). Similarly, the method proposed in [15] also trains a CNN on I/Q samples and enhances the signal representation with constellation diagrams, allowing the model to identify higher-order digital modulations.

Further exploration of constellation diagrams for RF signal classification was conducted in [16], where AlexNet and GoogLeNet were trained using tailored image processing techniques. A unified framework for modulation recognition and interference detection was proposed in [17], employing CNN-based classifiers trained on amplitude, phase, frequency transformation, and I/Q samples to condition RF signals. Similarly, a CNN-based classifier capable of separating simultaneous Wireless Local Area Network (WLAN) and Long-Term Evolution (LTE) transmissions from radar signals was introduced in [9]. Another study [10], utilized spectrograms and the *You Only Look Once* [18] object detector to identify Internet of Things (IoT) signals. This study uses a spectrum sensing technique that leverages DL to concurrently detect 5G NR and LTE signals within a wideband spectrogram. This approach harnesses the pattern recognition and feature extraction capabilities of deep learning. Instead of processing received signals using complex envelope data, we apply the Short-Time Fourier Transform (STFT) to convert them into spectrogram images. This transformation creates a richer data representation by capturing the radio properties of the signals in the time-frequency domain. Inspired by the semantic image segmentation techniques in computer vision, we employ ResNet50 and the Vision Transformer (ViT) model to identify and locate spectral contents of 5G NR and LTE signals within the wideband spectrogram images.

III. System Architecture & Methodology

A. The System Model

The system focuses on classifying RF signals based on their spectrogram representations.

- RF Signal Generation: LTE and 5G NR signals are transmitted with distinct characteristics.
- Signal Reception: The signals are captured using a receiver and converted into a digital form.
- Spectrogram Generation: The captured signals are processed into spectrograms using the Short-Time Fourier Transform (STFT).
- Classification: The spectrograms are used as inputs to a machine learning model (e.g., ResNet50, Vision Transformer) for classification.

The signal $s(t)$ is transformed into the time-frequency domain using the Short Time Fourier Transform (STFT) to generate a spectrogram. The STFT of the signal s(t) is defined in Equation 1, where $h(t-\tau)$ is the window function that control the timing of the fourier transform.

$$S(t,f) = \int_{-\infty}^{\infty} s(t) \cdot h(t-\tau) \cdot e^{-j2\pi f \tau} d\tau \tag{1}$$

$$Spectrogram(t,f) = |S(t,f)|^2 \tag{2}$$

Spectrogram, of a frequency f and time t dependent function $S(t,f)$ is defined in Equation 2 which captures the power distribution of the signal over time and frequency and is represented as a matrix $X(t,f)$. This spectrogram represents the energy content of the signal as a function of time t and frequency f and serves as the input to the deep learning model for classification.

Deep Learning Model Input: The spectrogram of a signal be represented as as a matrix $X \in \mathbb{R}^{T\times F}$ where T represents the time segments and F represents the frequency bins. Given a dataset $\{X_i, y_i\}_{i=1}^{N}$, X_i is the spectrogram of the i-th signal & $y_i \in \{0,1,2\}$ is the corresponding class label (0:LTE, 1:5G NR, 2:Combined LTE-NR), the goal is to train a model to predict the class label from the spectrogram. Now the classification task can be mathematically formulated as finding a function $f_\theta(X)$ parameterized by θ that maps the input spectrogram X to a probability distribution over the classes:

$$f_\theta(X) = \hat{y} = Softmax(W \cdot g_\theta(X)) \tag{3}$$

where $f_\theta(X)$ represents a function with parameter θ, input X and $\hat{y}$ is the predicted class label. In the equation, Softmax is the Softmax function is applied to normalize the output into probabilities, W is the weight matrix of the final classification layer, $g_\theta(X)$ represents feature extraction process of the deep learning models and (y) are the true labels y.

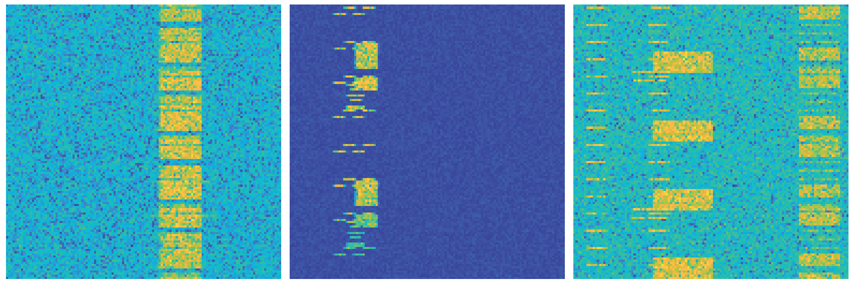
$$L(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{3} y_i(k)\log(f_\theta(X_i)(k)) \tag{4}$$

In Equation 4 $L(\theta)$ is the cross-entropy loss, which the model tries to minimize during training, $y_i(k)$ is the true label of sample i for class k, $log(f_\theta(X_i)(k))$ is the logarithm of the probability of the ith sample being in class k. Logarithm is used to penalize incorrect classification more heavily, N is the total number of samples in the dataset, and k indexes the three possible classes.

Figure 2 shows the overall methods used in this work. We use an Universal Software Radio Peripheral (USRP) [19] software-defined radio (SDR) configured to adapt to different RF signals. It harvests 5G and LTE signals and processes their wave forms to have time-frequency spectrogram images. Resulting images are tested and validated using the DL networks for the classification works to identify what signal exists in the environment.

B. RF Dataset Generation and Representation

The dataset was created using MATLAB 5G and LTE toolboxes [20]. Spectrogram images were generated based on standard 5G NR and LTE signal parameters to create comprehensive dataset for deep learning model training, while



(a) LTE spectrogram (b) 5G NR spectrogram (c) Combined LTE and NR

Fig. 1: Time (ms) Vs Frequency (MHz) spectrogram sample images with different SNR

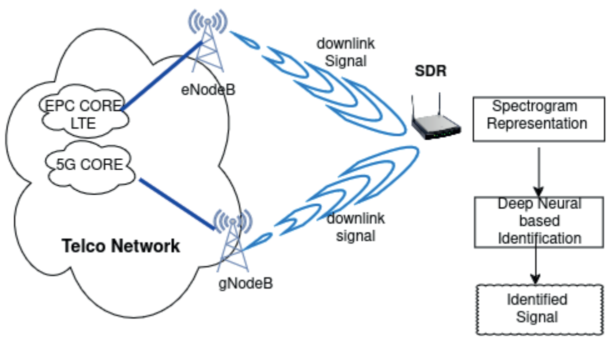


Fig. 2: Overall RF signal identification diagram

synthetic data offers flexibility in testing a wide range of signal configurations, we acknowledge that it may introduce biases or simplifications not present in real-world scenarios. Various artificial noise and channel effects, such as additive white Gaussian noise (AWGN), phase noise, frequency-selective fading, path loss, and nonlinear distortion, were introduced to the synthetic signals to emulate real-world network conditions. A total of approximately 3,000 spectrogram images were generated, with 1,000 images per class. Figure 1 presents sample spectrogram images of the three classes, where the images represent the time (ms) versus frequency (MHz) characteristics. The design of the deep learning networks necessitates a labeled RF dataset, and the learning process enables the networks to recognize and classify radio frequencies based on the dataset. For this purpose, spectrograms were chosen as they capture the time-varying spectrum content of RF signals on a time-frequency plane. To ensure a diverse and rich dataset, signal images were generated by varying key 5G and 4G parameters, such as channel bandwidth, sub-carrier spacing, synchronization signal block (SSB) patterns, SSB periods, reference channels, duplex modes, and channel models, with different center frequencies ranging from 700 MHz to below 6 GHz.

Dataset: To process the most recent standard signals and a variety of low-cost Software Defined Radio systems, the sampling rate is set at 100MHz, which provides up to 50 MHz of usable bandwidth. Bandwidth of $[10, 15, 20, 25, 30, 40, 50]$ MHz, sub-carrier spacing of $[15, 30, 60]$ KHz, synchronization signal block (SSB) duration (20 ms), and SSB Case A and Case B patterns [21] are a few typical 5G NR signal parameters. We adjusted the following parameters for LTE signals: bandwidth $[5, 10, 15, 20]$, the frequency-division duplexing transmission method [22], and the reference channel of $[R.2, R.6, R.8, R.9]$. The training dataset comprises signal images, where individual pixels are classified into one of three classes: NR (5G-NR), LTE, and NR-LTE.

The dataset is partitioned into training, and testing sets, with 80%, and 20% of signal images allocated to each set, respectively. To optimize efficiency, transfer learning was applied with ResNet50 chosen as the foundational network for image segmentation. Model's transfer learning capabilities, pre-trained models on large datasets, provide a valuable resource for adapting knowledge from one domain to another, particularly in scenarios with limited annotated data. ResNet uses a residual connection to address the vanishing gradient problem, enabling efficient training of very deep networks. Also, it exhibits parameter efficiency, allowing for more effective use of model capacity without a significant increase in parameters. Dataset and code is available on https://github.com/rajendra1124/RFsignalClassification

C. Deep Neural Network

The research methodology centered on employing a deep neural network for the semantic segmentation of wireless signal images. CNN-based ResNet50 model architecture and Transformer-based ViT models were used.

ResNet50: Residual Network (ResNet50) [1], a powerful deep CNN widely used in image classification due to its residual connections, which allow efficient learning in deep architectures. By enabling information to bypass certain layers, these connections focus the network's learning on residual mappings—differences between layer inputs and outputs. This design enhances the model's ability to capture complex patterns, resulting in higher classification accuracy. In this study, ResNet50's residual features are applied to distinguish RF signals.

ViT: Dosovitskiy et al. [2] introduced the Vision Transformer, a deep learning architecture that applies transformers, rather than conventional CNNs, for image recognition. ViT's attention mechanism enables it to capture global dependencies within images, making it highly effective for tasks like segmentation and classification. For RF signal classification, we use a ViT model pre-trained on ImageNet, adjusting the final fully connected layer to match the specific number of classes required. This model, optimized through hyperparameter tuning and validated to ensure generalization, leverages ViT's scalability and generalization strengths for enhanced classification performance.

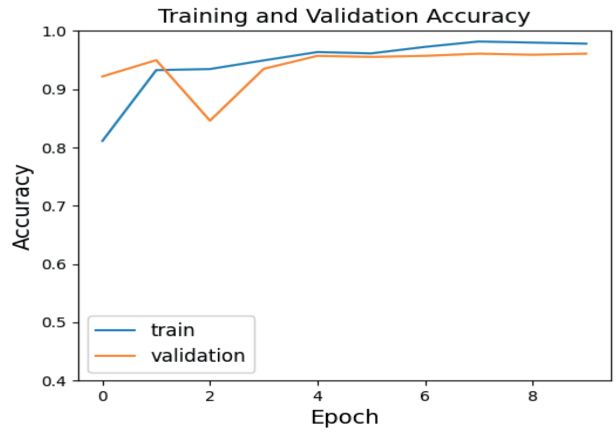
IV. Performance Evaluation

The ResNet50 model is set up with settings that do not include top layers, making it easier to customize for the particular classification assignment. The model architecture has an input shape (180, 180, 3), the size of the RF signal image, and the spatial dimensions are reduced by average pooling. Weights are initialized using pre-trained ImageNet weights for the three classes (LTE, NR, and NR-LTE). The model's output is flattened into a one-dimensional feature vector, facilitating advanced pattern learning. A dense layer with 512 units and ReLU activation enhances this capability, followed by a second dense layer with 3 units and softmax activation to generate classification probabilities. These additional layers, tailored for the RF signal dataset, enable precise classification into predefined categories, ensuring the model's adaptation for accurate categorization.

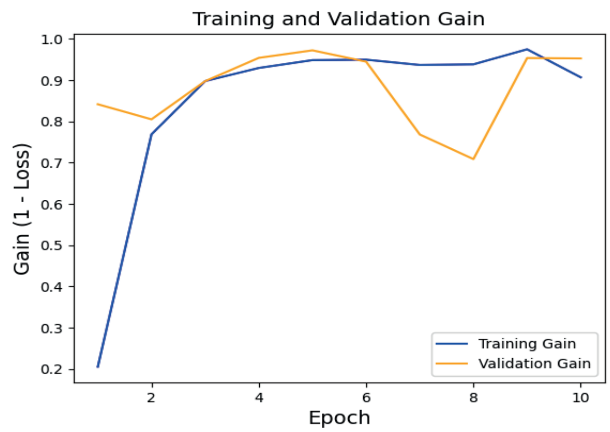
A transformer-based ViT model is used for classification using PyTorch and the Timm library. $ViT - base - patch16 - 224A$ architecture is used to load datasets from a directory. Reads images from subdirectories, assigns labels based on subdirectory names, and applies transformations using $torchvision.transforms.Compose$. The entropy loss function and the Adam optimizer with a learning rate of 0.0001 have been used. It iterates through the training dataset, calculates the loss, back-propagates, and updates the weights. The training loop runs for 10 epochs, and during each epoch, it iterates through 32 batches of data, computes loss, and updates the model parameters.

The accuracy reported for each epoch provides valuable insights into the training progress and performance of the ResNet50 image segmentation model over time. As shown in Figure 3a, the training accuracy increases after the first epochs. This indicates the model is learning and improving its ability to correctly classify images over time. Validation accuracy also increases, indicating that model generalization to unseen data is well done. Validation accuracy is consistently high and close to the accuracy on the training dataset, and it is not overfitting to the training data.

The Figure 3b depicts the training and validation accuracy of a Vision Transformer (ViT) model over 10 epochs, highlighting the model's learning and generalization performance. A training accuracy rises sharply in the initial epochs and stabilizes around 0.9 and validation accuracy fluctuating between 0.95 and 0.7. A dip in validation accuracy between epochs 7 and 9 suggests potential overfitting, though the performance stabilizes in later epochs. The small gap between training and validation accuracy indicates slight overfitting, though convergence is achieved after epoch 9. This suggests that the



(a) ResNet50 Model Training and Validation accuracy



(b) ViT Model Training and Validation accuracy

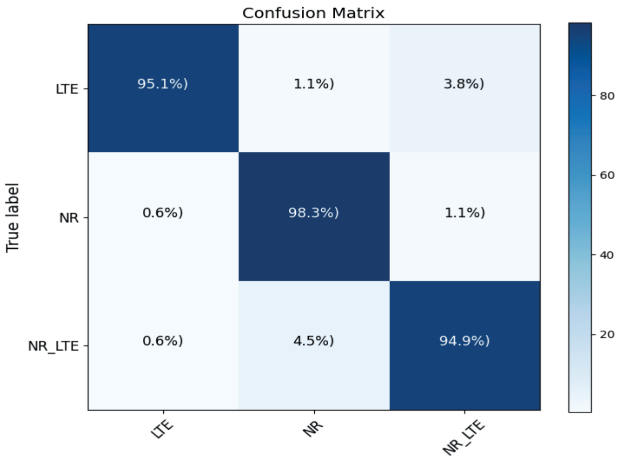
Fig. 3: Training and Validation accuracy of ResNet50 and ViT network architecture

model has effectively learned from the data, though further techniques such as regularization or learning rate adjustment could improve generalization. Despite minor fluctuations, the model shows strong overall performance with minimal divergence between training and validation.

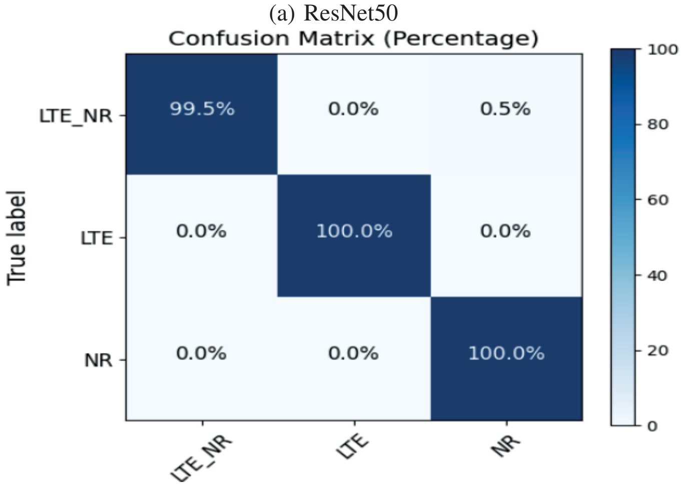
A. Results and Validation

Comparing the Vision Transformer (ViT) and ResNet50 models' training and validation accuracy graphs reveals distinct patterns in their learning and generalization behaviors. Both models demonstrate rapid improvement in the initial epochs, followed by stabilization. In the ViT model, there is a noticeable dip in validation accuracy, suggesting some overfitting, though it stabilizes at 85% accuracy, closely trailing the training accuracy of 92%. In contrast, the ResNet50 model shows more consistent and smoother performance, with both training and validation accuracies converging near 95% without significant fluctuations.

Figures 3 (a) and (b) show both models exhibit fluctuations. However, the ResNet50 accuracy graphs show steady



(a) ResNet50



(b) Vision Transformer

Fig. 4: Confusion matrix for ResNet50 and ViT network model

improvement over epochs, while the validation loss graphs of the ViT model demonstrate more erratic behavior. This could be insufficient training data for the ViT model, a suboptimal learning rate, or inadequate model capacity. A ViT model graph provides scope for optimization and improvement; I will put this task into future work with larger number of dataset.

Figure 4a shows the confusion matrix of the ResNet50 model. The true positive rate for the NR class is 98.3%. This shows the model correctly predicts the NR class with high accuracy. The true positive rate for the LTE class is 95.1% which is slightly less than NR, but the model still demonstrates good performance in correctly predicting. For the combined images, the *true positive* is 94.9%. The model performs well in correctly predicting the class. The *false positive* classification is not greater than 5.1%. The high true positive rates for all three classes indicate that the model effectively captures the distinguishing features of each class. Figure 4b shows the confusion matrix of the ViT model that reveals nearly perfect positive rates for all three classes. The ViT model demonstrates exceptional performance in correctly predicting LTE and NR classes on the given dataset. This shows the model's ability to

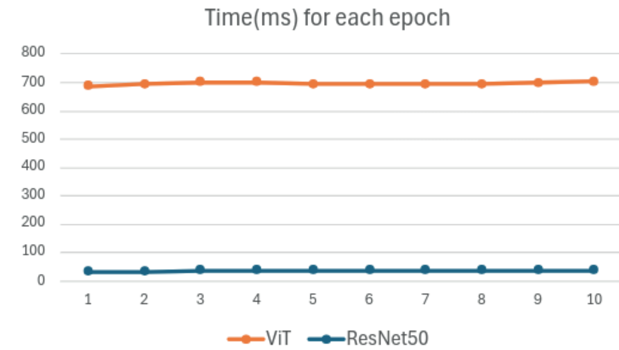


Fig. 5: Computational time comparision

accurately capture the unique characteristics of LTE samples, leading to flawless predictions.

B. Identification of real world Signals

We utilized the USRP B200mini Software Defined Radio (SDR) to capture 5G and LTE signals for signal identification experiments conducted at the Cyber Innovation Lab, George Mason University, Arlington, VA. An iPhone operating in 5G-only and LTE-only network modes was used to retrieve serving cell information by dialing the code *3001#12345#* in field test mode. This provided essential details such as Absolute Radio Frequency Channel Number (ARFCN) and bandwidth, enabling the USRP B200mini to be tuned to the correct frequency and bandwidth for signal measurement. The captured RF signals were processed into spectrogram images using MATLAB, which were then classified using machine learning models trained on synthetic datasets. While the models were able to identify real RF signals, the classification accuracy on real-world datasets was approximately 65%, notably lower than the performance achieved with synthetic data. The reduced accuracy on real datasets compared to synthetic datasets is attributed to several factors. Real-world RF signals are affected by environmental complexities such as multipath propagation, interference, and noise, introducing distortions that challenge the model's ability to generalize effectively. In contrast, synthetic datasets are generated under controlled conditions, resulting in cleaner and more predictable data.

C. Computation comparision

The figure 5 compares the computational time per epoch for the Vision Transformer (ViT) and ResNet50 models across 10 epochs. The ViT model consistently requires approximately 700 ms per epoch, showing minimal variation. In contrast, the ResNet50 model maintains a significantly lower computational time, remaining under 50 ms per epoch. This indicates that ViT is around fourteen times more computationally expensive than ResNet50 for each epoch. The higher cost of ViT is attributed to its transformer-based architecture, which relies on self-attention mechanisms. ResNet50, being a convolutional neural network, is computationally efficient but lacks the ability to capture global dependencies. This comparision highlights the trade-off between computational efficiency and computational cost in deep learning.

V. Discussion

The ResNet50 and ViT models demonstrate high true positive rates across all classes, indicating their strong performance in classifying NR, LTE, and $LTE - NR$ samples. The ViT model consistently achieves higher true positive rates compared to the ResNet50 model within our data set, suggesting superior performance in capturing the intricate features and patterns present in the data. Unlike traditional convolutional neural networks like ResNet50, which operate on local image patches, ViT treats the entire image as a sequence of tokens and processes them through self-attention mechanisms. This enables ViT to capture global context and long-range dependencies within the images, resulting in enhanced performance in capturing the distinguishing features of the classes. Furthermore, the consistently higher true positive rates across all classes in the ViT model underscore its effectiveness in accurately classifying samples. The ViT model's ability to achieve perfect true positive rates for the LTE and NR classes indicates its robustness in capturing the unique characteristics of these classes, leading to flawless predictions. On the flip side ViT model is much higher resource intensive than ResNet50 model.

VI. Conclusions

The Vision Transformer (ViT) model's near-perfect accuracy within our dataset across all classes highlights its exceptional ability to recognize intricate patterns and pixel correlations. This performance underscores its potential as a highly effective model for classification tasks in various image segmentation domains.

The ViT model excels at generalizing from known to unknown data and accurately classifying samples, emphasizing its pivotal role in advancing computer vision and its applicability to real-world challenges across diverse fields. Although its training and validation gain shows greater fluctuations compared to the ResNet model, the ViT achieves superior classification accuracy. However, this comes at a significantly higher computational nearly 14 times compared to ResNet50.

VII. Future Work

Multi RF signals classification: Enriching the training dataset with a wider range of RF signals (WiFi, Zigbee, Bluetooth, etc.) will enhance the model's ability to classify different signal types accurately and generalize well to various RF environments

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[2] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations*, 2021.

[3] NTIA, "Spectrum Sharing Framework," accessed: June 29, 2024. [Online]. Available: https://www.ntia.gov/press-release/2024/ntia-fcc-navy-work-expand-innovative-35-ghz-spectrum-sharing-framework

[4] FCC, "3G Service Shutdown," accessed: June 29, 2024. [Online]. Available: https://www.fcc.gov/consumers/guides/plan-ahead-phase-out-3g-cellular-networks-and-service

[5] S. Segan, "3g service shutdown," accessed: June 29, 2024. [Online]. Available: https://www.pcmag.com/how-to/the-3g-shutdown-how-will-it-affect-your-phone

[6] A. Vagollari, V. Schram, W. Wicke, M. Hirschbeck, and W. Gerstacker, "Joint detection and classification of rf signals using deep learning," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, 2021, pp. 1–7.

[7] D. Ariananda, M. Lakshmanan, and H. Nikookar, "A survey on spectrum sensing techniques for cognitive radio," in *2009 Second International Workshop on Cognitive Radio and Advanced Spectrum Management*, 2009, pp. 74–79.

[8] O. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," *IET Communications*, vol. 1, pp. 137–156, 2007.

[9] A. Selim, F. Paisana, J. A. Arokkiam, Y. Zhang, L. Doyle, and L. A. DaSilva, "Spectrum monitoring for radar bands using deep convolutional neural networks," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–6.

[10] E. Lo and J. Kohl, "Internet of things (iot) discovery using deep neural networks," in *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020, pp. 795–803.

[11] K. N. R. S. V. Prasad, K. B. D'souza, and V. K. Bhargava, "A downscaled faster-rcnn framework for signal detection and time-frequency localization in wideband rf systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4847–4862, 2020.

[12] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[13] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 168–179, 2018.

[14] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Engineering Applications of Neural Networks*. Springer International Publishing, 2016, pp. 213–226.

[15] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 4074–4077, 2019.

[16] S. Peng, H. Jiang, H. Wang, H. Alwageed, Y. Zhou, M. M. Sebdani, and Y.-D. Yao, "Modulation classification based on signal constellation diagrams and deep learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 3, pp. 718–727, 2019.

[17] M. Kulin, T. Kazaz, I. Moerman, and E. De Poorter, "End-to-end learning from spectrum data: A deep learning approach for wireless signal identification in spectrum monitoring applications," *IEEE Access*, vol. 6, pp. 18 484–18 501, 2018.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[19] E. Research, "Software defined radio," accessed: June 29, 2024. [Online]. Available: https://www.ettus.com/

[20] MathWorks, "5g toolbox," accessed: March 10, 2024. [Online]. Available: https://www.mathworks.com/products/5g.html

[21] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "3GPP TS 38.101-1: User Equipment (UE) Radio Transmission and Reception; Part 1: Range 1 Standalone," accessed: March 6, 2024.

[22] 3rd Generation Partnership Project Technical Specification Group Radio Access Network, "3GPP TS 36.101: Evolved Universal Terrestrial Radio Access (E-UTRA) User Equipment (UE) Radio Transmission and Reception," accessed: March 5, 2024.