Slide 1: Title Slide

"Today, we'll be discussing a solution to the ITU AI Challenge PS-003."

Slide 2: Introduction

"First lets recall the challenge. The Berlin V2X dataset provides extensive wireless measurements from Berlin. Our central challenge is Can we effectively train a model on cellular communication data from one operator and adapt it to accurately predict the datarate on another operator's data."

"Now, you might be wondering, why should we even care about this challenge?"

Slide 3: Why is important

"Every day, more vehicles are becoming interconnected, creating a massive web of communication."

"And we want to ensure that data transmission between vehicles and infrastructure is consistent and dependable."

"However, a significant challenge arises when the data itself comes from varied distributions."

Slide 4: Dataset exploration

"Our primary data source is the Berlin V2X dataset, offering cellular network metrics."

And from the figure 2 in Berlin V2X paper, we can see similar downlink datarate distributions between the two operators like in avenue and residential, while there are different distributions in park and highway. So, how do we tackle this? The solution lies in Domain Adaptation."

Slide 5: Domain Adaptation

"It's about taking what a model has learned in one domain and making it applicable to another."

"It's crucial because, in the real world, data is messy and diverse. Our model needs to be versatile."

Slide6:

Here is our solution to solve this challenge. First we set different operators as source and target domains. And do the data exploration and preprocessing to ensures that our data is in the right format and quality for modeling.

In this approach, our novelty is to use the combination of models that have different functions.

For Feature Selection, "We chose to use the Random Forest Regressor for feature selection because not all features are equally informative. Using this method allows us to identify the most significant predictors for our continuous target variable. It also aids in reducing overfitting and computational costs."

Slide 7: Random Forest Regressor

"Random Forest is a collection of decision trees that 'aggregate' their predictions to provide a more accurate estimate of the continuous target variable."

As shown in the table, we can see the features that most significantly influence the downlink datarate.

Slide 8: Domain Adaptation with DANN

"With our features ready, we harnessed the power of the Domain-Adversarial Neural Network, or DANN."

"DANN is a multi-task learner."

"It consists of three main components: The Feature Extractor, the Main Task Regression, and the Domain Classifier."

"The Feature Extractor takes in the data and transforms it into a format that's easier for the model to understand."

"The Main Task Regressor tries to predict the data rate."

"The Domain Classifier tries to distinguish if the data comes from the source or the target domain."

"Enter the Gradient Reversal Layer. During the training, this layer essentially reverses the gradients for the domain classifier. The goal of this layer is to help the model to learn features that are useful for the main task while being invariant to domain shifts. So, the domain classifier gets confused, ensuring our features remain domain-agnostic."

However, in our task, it's essential to note that while both domains' data were fed into the model, the primary datarate regression task was trained only using the source domain's data, while the domain classification task used the domain labels.

"Now, let’s see the results."

Slide 11: Results

In the results table, the DANN model's performance is tracked across multiple epochs. Notably, By the final epoch, the DANN Loss has refined to about 0.1056. Additionally, we used r2 score to evaluate the model, which quantifies how effectively our model can predict datarate in the target domain based on its training in the source domain. And we achieved a good score of about 0.8352, which suggests the model’s capability across different data distributions.

Slide 12: conclusion

In conclusion, our efforts in this research illuminate effectiveness of the combination of two algorithms and the promise of Domain Adversarial Neural Networks in bridging data distribution discrepancies. Our model managed to attain an impressive r2 score on the target domain. This signifies that although there are differences in data distribution between the two operators, our model effectively generalized and made reliable predictions.

"Thank you for your time and attention. Let’s dive into your questions!"

However, a notable departure from typical DANN implementations is the exclusive training of the Domain Classifier on the source domain. While this does limit our model's robustness in achieving full domain invariance, it provides us with a unique opportunity to study how well such a model can generalize on an entirely unseen target domain. By comparing the model's performance on the target domain before and after integrating it into the training process, we can gauge the inherent adaptability of features extracted purely from the source domain.