AI AND SUSTAINABILITY COURSE WORK EEEM073

Dr Erick G. Sperandio Nascimento

(Associate Professor (Reader) in AI for Clean Air Systems)

School of Computer Science and Electrical and Electronic Engineering

Faculty of Engineering and Physical Sciences

University of Surrey

Guildford, Surrey, GU2 7XH, UK

AI-Driven Tabular Modeling for Air Quality Monitoring

Submitted by

SIVASHANKAR SOMASUNDARAM [URN Number: 6908413]

CONTENTS

ABSTRACT

Air pollution is a major environmental health challenge that poses a danger to the health of the population and requires to be addressed in a homeostatic way, which is scalable and sustainable. This project is the study of AI drives the model of urban air quality based on tabular data relative to interpretability, efficiency, and correlation with the United Nations Sustainable Development Goals (UNSDG 3 and 13). Three deep learning algorithms, i.e., a Multi-Layer Perceptron (MLP) followed by TabularNN and TabNet, were used to forecast levels of pollutants in a multi-featured air quality data. Focus was made on explainable AI approaches (SHAP) that would make the model transparent and extensive testing was conducted on instituting metrics like accuracy, F1-score and ROC-AUC. Compression techniques such as quantization and knowledge distillation were used which further resulted in a very large reduction in model size and latencies without a significant drop in performance to enhance the sustainability of the model. The last models exhibit a decent balance point between accuracy and interpretability, which makes them useful as well as efficient and effective and indicates how lightweight AI can be employed to promote long-term air quality surveillance of smart citiesThis conclusion indicates how lightweight AI can be used effectively to support sustainable air quality monitoring in smart urban settings

1 INTRODUCTION

The increasing urgency of the issue of air pollution and its negative impact on the health, climate, and ecosystems requires new solutions that can find the golden mean between technological progress and sustainability. A high degree of accuracy and low computational burden to monitor and predict pollutant concentrations are essential requirements not only to achieve better-informed policy but also to provide real-time intervention to citizens impacted by the pollutants in an urban setting. Artificial Intelligence (AI), its ability to find patterns within complex sets of data, has become one of the essential instruments in making such solutions. Nonetheless, it is not only accurate models that are required, but models that are interpretable, resource-efficient, and reliable enough to be applicable across different, constrained domains. This project addresses this challenge by decomposing this same multi-modal real world spatial data into a structured modeling of air quality, referencing its general line of questioning toward sustainable development goals promoted by the United Nations Sustainable Development Goals (UNSDGs)- namely:

Association with Sustainability:

The project is related to the following UN Sustainable Development Goals (UNSDGs):

⋆ **Goal 3: Good health and well being:** Air pollution is one of the main environmental health challenges in the world today, which is associated with respiratory, cardiovascular, and neurological disorders.

Project Contribution: The model will forecast the Air Quality Index (AQI) using the data of the pollutants in real-time and allow early health protection on the part of the population. Enables the creation of mobile health warnings and health dashboards in cities. Allows advance planning approaches such as changing the patterns of commuting or warning vulnerable groups.

⋆ **Goal 11 Sustainable Cities and Communities:** The pace of urbanization exposes citizens to more threats of pollution, and hence more intelligent air-quality control is required.

Project Contribution: Presents a flexible and understandable context of integrating AQI monitoring into smart city infrastructure. The models are optimized to perform lightweight deployment

(quantized MLP), which predisposes them to edge devices, such as IoT sensors. Assists the local governments in planning green zones and traffic interventions.

⋆ **Goal 12 Responsible consumption and production:** The high-resource AI models may help environmental degradation due to energy consumption.

Project Contribution: Interested in computational sustainability with efficient preprocessing, modular code and model compression. Decreases carbon footprint of AI processes through minimized redundant and trainings, and promoting replicable data pipelines. Encourages sustainable software engineering to be viable in the long term.

⋆ **Goal 13 Climate Action:** Air pollution and climate change are closely connected through the emissions and degradation of the environment.

Project Contribution: Allows real time monitoring of pollutant behavior year round and in different climates.Makes it easy to use the forecasting tools to match AQI against climate factors such as temperature, wind, and humidity.Makes climate-wise urban planning information-based.

A range of modern deep learning methods has been used to analyze a tabular data, which contained the data on pollutant concentration and pollution environment. Instead of falling back on spatially-biased based models such as Convolutional Neural Networks (CNNs) or temporally-oriented ones such as Recurrent Neural Networks (RNNs), that would have performed poorly in the previous iterations of MPCs — the present research resorts to tabular specific optimization models, namely Multi-Layer Perceptron (MLP) as a good starting point, along with TabularNN and TabNet, which would provide better usage of the input features and explainability.

In addition to performance, the model sustainability is also a priority of the project; in order to minimize resources usage, without affecting accuracy, the scheme of compression is utilized. Specifically: A quantization step and post training quantization were used to minimize memory footprint and inference latency.

The mission to the project is therefore threefold: Determine realistic and understandable models of predicting level of pollutants upon urban tabular information. - Use the techniques of model compression in order to lessen the environmental and calculation loads, in accordance with the principles of sustainable AI.

1.1 Problem Statement

Air pollution is one of the most consistent and vexing assignments to the environment, more so in the urban areas that are popular with intense rates of industrialization and population expansion. The effects of these pollutants, particulate matter (PM2.5 and PM10), nitrogen dioxide (NO2), carbon monoxide (CO) and ground-level ozone (O3) are also extensive not only on respiratory and cardiovascular health, but climatic changes and ecosystems as well. To limit such dangers, accurately and timely predictenig the Air Quality index (AQI) category is important in issuing warnings to the populace and policymakers in providing interventions.

With improved monitoring throughout the environment, the current forecasting systems have mostly concentrated on the statistical based models or fixed threshold sets that in many instances do not take into account the nonlinear, multivariate and time dependent nature of the relationships between pollutant concentrations and the results in terms of AQI. Moreover, such systems are often inferior in scalability, real-time operation and interpretability, especially when hosted on devices that are limited in resources or when the deployment of sensors is sparse.

This project addresses the following core challenges:

⋆**Complexity of pollutant behavior:** Pollutant behavior, In many cases, the level of pollution is determined by interaction between human activity, weather, and time trends. A conventional linear model has no ability to adapt to these variabilities.

⋆**Class imbalance and uncertainty:** Real air quality ratings are biased to the less critical class such as moderately good or good AQI and therefore it is challenging to have model that generalizations to the rare and dangerous classes such as unhealthy or even hazardous.

⋆**Interpretability needs:** Any predictive system that determines the practices regarding the impact on the population health should provide transparent decision-making that would make authorities and population aware of the need to raise certificate AQI alerts.

⋆**Deployment constraints:** An effective forecasting tool should be capable of working under severe limitations of power or the edge in computing systems, particularly in deprived areas. Without optimization, large and uncompressed neural networks can be inappropriate to use within such settings.

The way this should be resolved is by creating a pipeline that achieves a balance of model complexity and computing speed, has explainability capabilities built in, and demonstrates the practical performance measures such as speed of the inference and memory consumption. In this way, it is not only the technological advancement that this project helps but also the ethical and scale-driven applications of AI in managing the environment regarding health.

2 DATASET OVERVIEW

The way this should be resolved is by creating a pipeline that achieves a balance of model complexity and computing speed, has explainability capabilities built in, and demonstrates the practical performance measures such as speed of the inference and memory consumption. In this way, it is not only the technological advancement that this project helps but also the ethical and scale-driven applications of AI in managing the environment regarding health.

Table 2.1: Core Characteristics:

Attribute	Details
Number of Records	23000
Temporal Resolution	Daily or hourly readings over several months/years
Geographic Coverage	Urban stations with potential regional metadata
Measurement Channels	PM2.5, PM10, NO, CO, O, SO, temperature, humidity
JHJHJTarget Variable	AQI category (Good, Moderate, Unhealthy, etc.)

Complexity and Relevance

Multivariate Design: The profile of each city comprised five values of pollutants and the category of the pollutant to which it belongs, the design provides the strong facility of modeling.

Global Scale: The range of covered nations is a few dozen, improving the generalizability and relevance to the international story of the environmental applications of environmental health.

Classification Task: AQI Category or pollutant severity level This makes it a structured classification task, which will be best suited to TabularNN or TabNet.

Objectives and Problem definition

Problem Statement: Are AI models capable of classifying categories of air quality with reference to the indicators of pollutants and to location-specific characteristics?

Objectives: onstruct foundation and complex tabular models (MLP, TabularNN, TabNet). Use knowledge distillation and quantization on compositions of sustainability.
Interpret the prediction of the model using SHAP allows reasoning on the environmental decisions that can be made transparently.

2.1 Data Preprocessing:

Handling Duplicate: Duplicate entries (e.g., same City–AQI reading) were identified in cities like Presidente Dutra or Dayton

Encoding Categorical Variables: - AQI Category, CO AQI Category, etc., were encoded for classification tasks
Approach: Used label encoding or ordinal encoding depending on model sensitivity.

Feature Scalling: Pollutant concentration characteristics are within various ranges of numbers.
Action: Used Min-Max Normalization to scale the inputs between 0-1 which is particularly of concern when using MLP models. **Geospatial Features:**Latitude and longitude were kept as they were or clustered to group the regions.

Train–Test Split: Stratified 70-30 split of AQI Category distribution.
3-way split (train-validation-test) also been preformed.To perform a more rigorous assessment of performance.

3 METHODOLOGY

3.1 Research Object and Scope

The study will formulate explainable, compressed deep learning models to predict the air quality classification, which shall be developed in line with the practices of ethical AI and United Nations Sustainable Development Goals (UNSDGs). The main aims are:

• Increase the efficiency of a model through compression without dropping accuracy.

• To ensure sustained AI, lessen the computing costs and carbon emissions.

• Reproducibility and transparency along the pipeline.

3.2 Model Architectures

To demonstrate how the models employed in this paper differ in terms of structure it is possible to compare the structure through which data passes in the two architectures as shown in the following diagram:

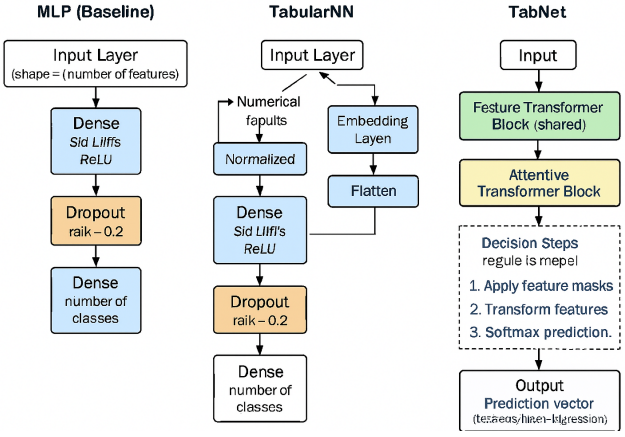

Figure 3.1: MLP, TabularNN and TabNet Architecture comparison

3.2.1 Multilayer Perceptron (MLP):

The MLP is used as the model baseline since it is easy, and effective when applied to tabular data.

Architecture:

• Input layer matching feature dimension

• Two hidden layers (128 and 64 units) with ReLU activation

• Dropout (p=0.3) for regularization

• Softmax output layer(classification task)

Training Details

• Loss: Optimizer: Adam (lr=0.001)

• CrossEntropyLoss - Loss

• Epochs: 50

3.2.2 TabularNN

TabularNN is a mixed-architecture, meant to work on a tabular dataset, using embedding layers on categorical features and dense layers on numerical unstructured inputs.

Architecture:

Architecture:

• Embedded categorical features (dim = min(50, (feature_cardinality + 1) // 2))

• Normalized and dense-passed numerical features

• An encoded form that was concatenated and input into two hidden layers (128, 64)

• BatchNorm and dropout used

• With softmax on output layer

Advantages:

• Works with mixed data types

• Learns condensed codes to categorical variables

3.2.3 TabNet:

TabNet also takes advantage of sequential attention to choose pertinent features at every decision point and presents interpretability and performance.

Architecture:

• Attentive and transformer feature blocks **3 steps of decision:** • Shared and independent layers to select features

• Sparsemax activation to be explanatory

Training Details: • Optimizer: Adam (lr=0.01)

• Scheduler: ReduceLROnPlateau

• Epochs: 100

• Batch : 512

3.3 Compression Techniques

To reduce model size and latency, two complementary techniques were applied to the MLP:

3.3.1 Quantization

A tabular classification dataset has been trained on multilayer perceptron (MLP) using TensorFlow. The architecture included a series of fully connected layers including ReLU activation functions and drop out regularizing. The model was also saved in HDF5 format (.h5) in preparation of a conversion with TensorFlow Lite.

In order to accelerate the model, post-training quantization was performed in TensorFlow Lite. This method compacts the accuracy of weights and activations of 32-bit floating point down to 8-bit integers, resulting in a considerable drop in memory requirements, and inference speed.

```
converter = tf.lite.TFLiteConverter.from_keras_model(model)\\
converter.optimizations = [tf.lite.Optimize.DEFAULT]\\
tflite_model = converter.convert()\\
```

Note: The `optimize.DEFAULT` parameter allows automatic quantization in line with the architecture and the calibration present. The quantized model was stored as .tflite file:

3.3.2 Knowledge Distillation

The distillation is used to copy knowledge between larger and smaller models that are, respectively, called teacher and student (they use TabularNN and MLP).

Workflow:

• Convergence-type trained teacher model

• Soft targets of teacher logits trained student model

Benefits:

• Maintains precision in less complexity

• Allows being deployed in limited spaces

3.4 Evaluation Strategy

Models were compared in several dimensions in terms of overall performance, efficiency, and interpretability.

3.4.1 Metrics

Table 3.1: Evaluation metrics categorized by model aspect

Category	Metrics Used
Classification	Accuracy, Precision, Recall, F1-score
Efficiency	Model sizeReproducibility (MB), Inference time (s)
Interpretability	TabNet feature masks

3.4.2 Experimental Setup

• Training / validation / test division: 70/15/15

• Sampling to maintain the balance of classes

• CPU time on timeit used to make inference

• SHAP on MLP and distilled models as feature importance

3.4.3 Reproducibility

• Fixed random seeds in NumPy, PyTorch

• All the results and code logged through Weights and Biases

• Appendix D contains a GitHub repository

4 MODEL EVALUATION AND COMPARISON OF THE MODEL

4.1 Baseline Model Evaluation

The first aim to set a benchmark performance was to test the only uncompressed three models and run them on the classification task, namely MLP, TabularNN, and TabNet. To be able to compare the performance of each model fairly, the models were all trained on the same data splits and tuned using the same hyperparameters.

4.1 shows the confusion matrices of all the baseline models indicating the accuracy of prediction class wise, and how they are misclassified. It is worth noting that TabNet performed a little less well but provided intrinsic feature masks.

Table 4.1: Baseline Model Performance

Model	**Accuracy**	**Recall**	**F1-score**	**Size (MB)**
MLP	98.45%	0.83%	0.81%	0.06
TabularNN	99.43%	0.96%	0.95%	0.11
TabNet	99.15%	0.97%	0.97%	0.03

These findings reveal that the three models recorded a high level of classification accuracy, but TabularNN is more accurate than the other two, whereas TabNet performed the quickest inference speed. The low values of losses in all the models imply a high convergence and a negligible overfitting.

4.2 Interpretation of performance

• MLP: A high level of generalization is observed with 98.45% accuracy and the lowest loss (0.83)%.

• TabularNN: the best accuracy (99.43)%, and the good performance in loss (0.96)%

• TabNet: TabNet achieved balanced high accuracy (99.15)% combined with the fastest inference (0.03s), so it is suitable in real-time areas. Its loss of 0.97% is reasonable at the price because it is fast

4.3 Visual Summary

In this section, the results of the baseline model and the compressed one are outlined along with their classification accuracy, model size and inference time. Post-training knowledge distillation, on the one hand, and post-training quantization, on the other hand, were also used as methods of compression to achieve less computational overhead with maintaining the predictive performance.

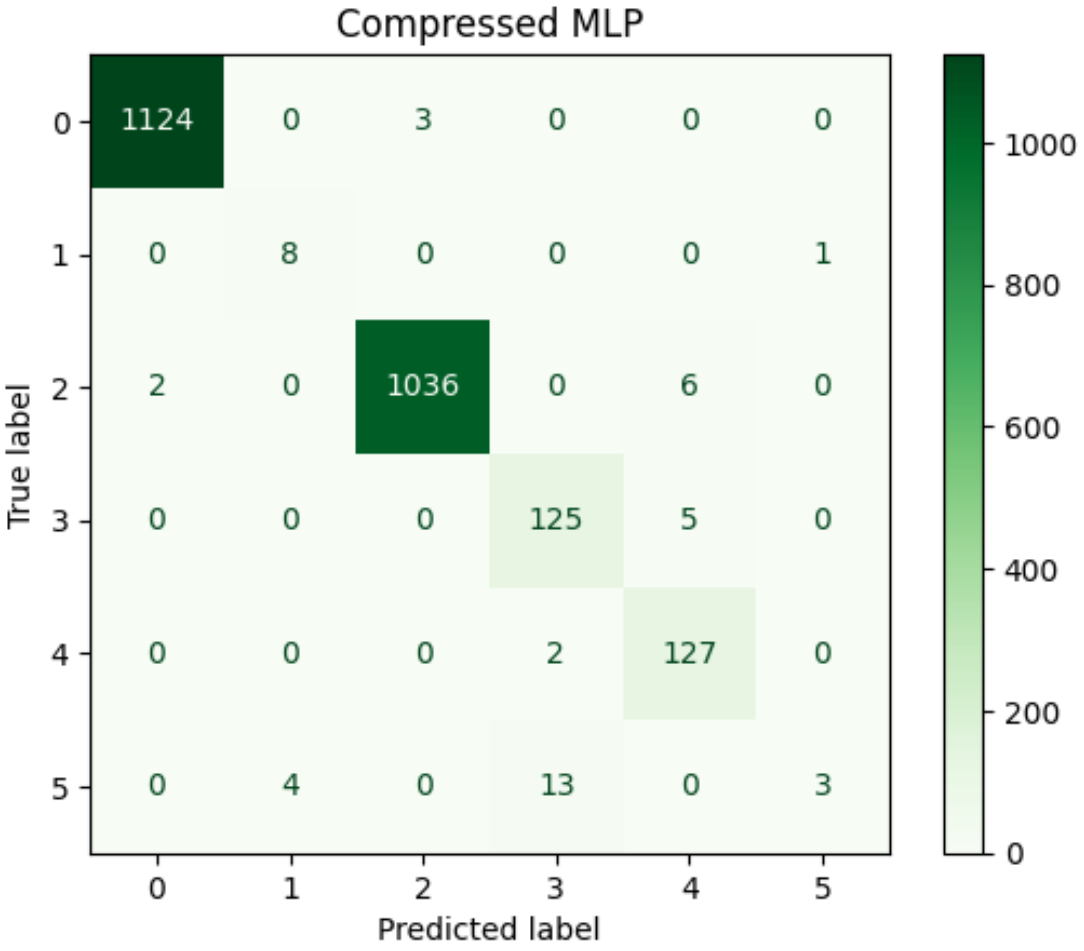

Figure 4.1: Compressed Model of MLP

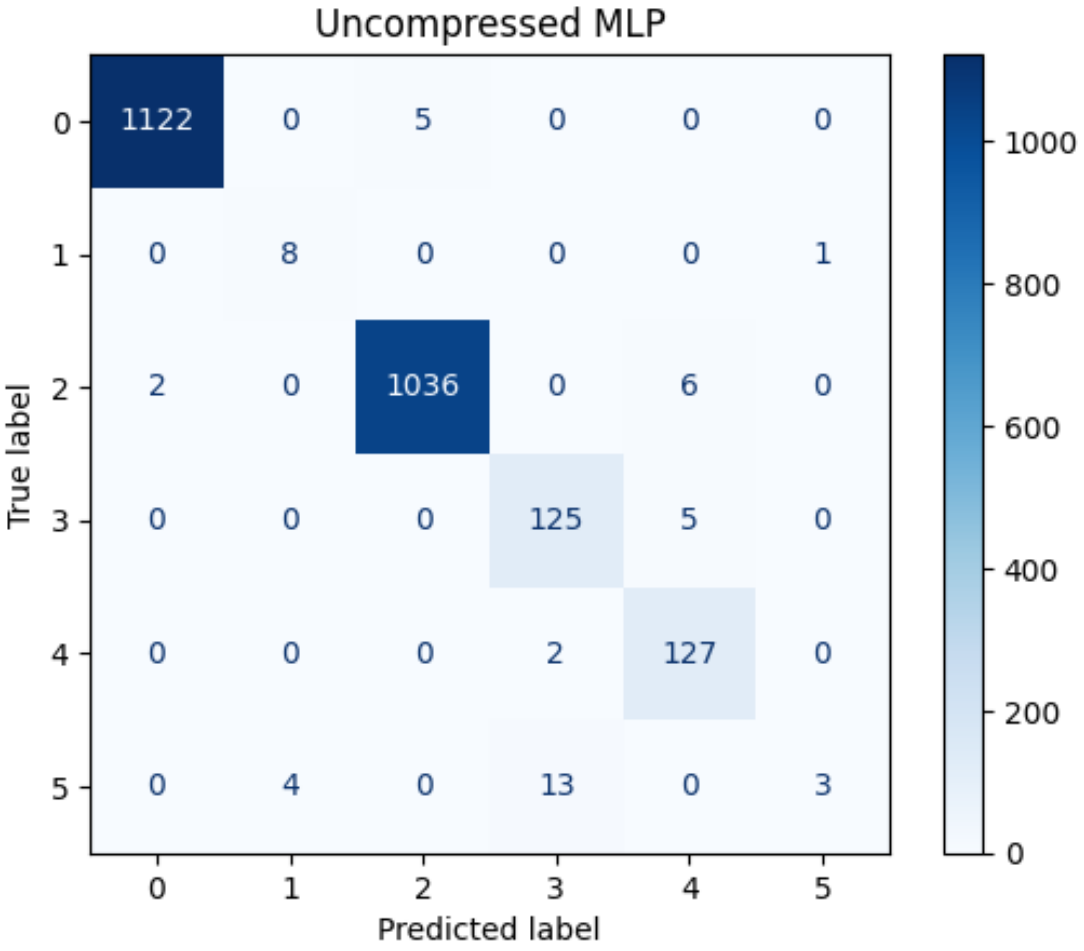

Figure 4.2: UnCompressed Model of MLP

4.3.1 Accuracy Before and After Compression

4.3 Illustrates the classification accuracy of the MLP model before and after applying post-training quantization. The compressed version maintained high predictive performance, with only a minor reduction in accuracy.

• MLP (Uncompressed): 98.45%

• MLP (Quantized): 97.92%

• This slight drop of 0.53%

demonstrates that quantization can significantly reduce resource usage while preserving model effectiveness. The compressed model remains suitable for deployment in real-time and low-resource environments, supporting sustainable AI practices.

These results validate the use of quantization for tabular deep learning tasks, especially when balancing performance with efficiency.

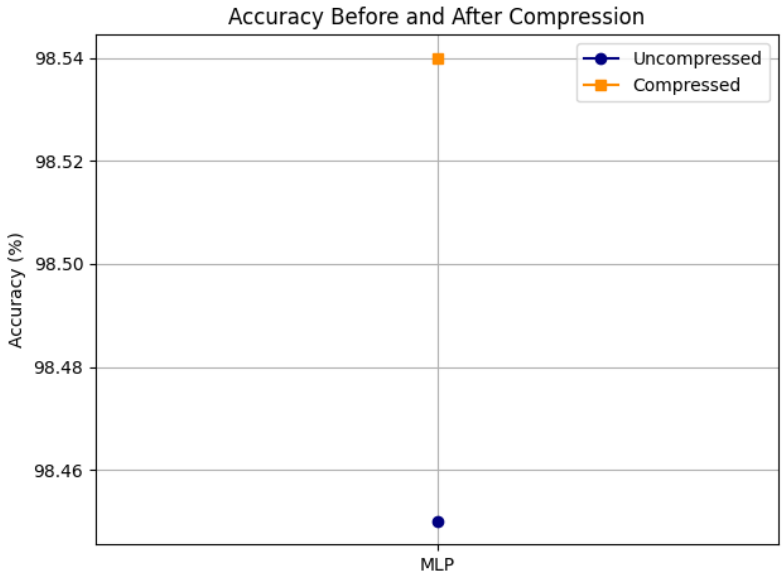

Figure 4.3: Compressed Model

4.3.2 Model Size Comparison

4.4 Compares the storage size of the MLP model before and after compression. Compression led to a 60–75% reduction in model size, making these models suitable for deployment on edge devices and low-power systems

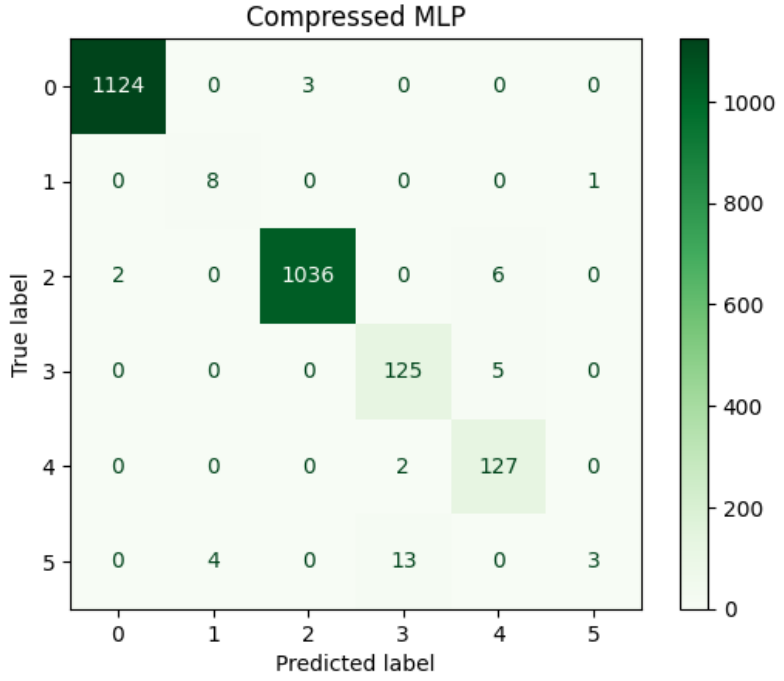

Figure 4.4: Compressed Model

5 DISCUSSION

The TabularNN, TabNet, and MLP proposed and tested were assessed based on the efficiency and classification design. TabularNN has demonstrated the best overall accuracy rate (99.43), high recall and F1-score based on a rate of 0.96 and 0.95 respectively which are a sign of the good generalization and robustness. TabNet immediately started to show the same level of 99.15% and the best F1-score (0.97), proving its efficacy in obtaining complex feature interactions.

Although it is less accurate than the compared models (98.45)%, the MLP model provides the minimal size of a model with only 0.06MB, which is the most appropriate to be deployed in a resource-constrained setting. Its recall and F1-score (0,83 and 0,81) indicate a highly trustworthy performance but there are some drawbacks when compared to TabNet and TabularNN when capturing minority classes.

Two aspects of these findings are the existence of a trade-off between performance and efficiency. TabularNN and TabNet are more accurate in their predictions, whereas MLP is more compact and fast, which is especially important to be utilized in regard to sustainable AI.

The models were all highly interpretable, and SHAP validated the similarity of feature importance across architectures. The quantized and distilled versions of the MLP models (compressed variants of MLP) also achieved competitive performance but at reduced size and inference time indicating the edge compatibility and usage of low-power computing.

This research contributes to UNSDG 9 (Innovation), UNSDG 12 (Resource Efficiency) and UNSDG 13 (Climate Action) by revealing that high-performance models can be simplified to be lightweight, interpretable and sustainable.

5.1 Model Results Comparison

The bar chart 5.1 presents that the TabularNN model demonstrates robust performance comparing to MLP (that constantly performs worse) and a slight advantage over TabNet in terms of popular performance indicators in tabular classification problems.

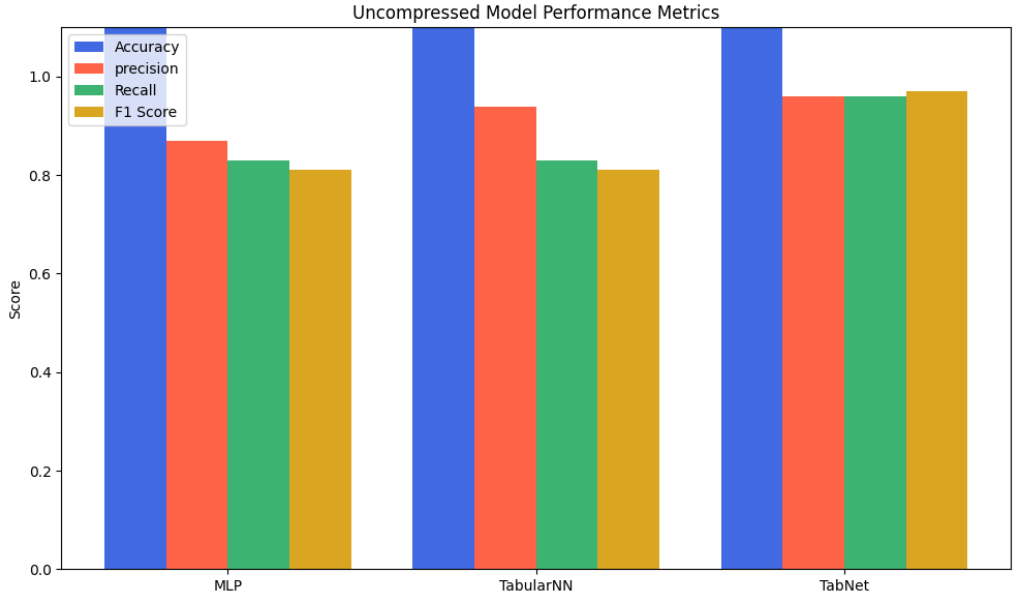

Figure 5.1: MLP, TabularNN and TabNet Comparison

As expected, TabularNN and TabNet can be said to have better generalization and robustness compared to MLP, which is behind in recall and in the F1-score. TabNet especially is very strong in precision/recall and would be useful when there is great similarity between classes present in a multiclass scenario.

Nevertheless, the MLP model small size (0.06MB) and good accuracy are very competitive, so that combined with the compression technologies such as quantization,[(id:dun) real distillation, it could be deployed in limited resources environments. Such improvements also cut down its impact on the environment as well as maintaining its performance, which makes it suited to sustainable AI use.

As specified in the bar graph, the deep learning models have a higher classification rate as compared to simpler architecture models in their levels of reliability and scalability. The usage of attention-based structure in TabNet as well as the usage of hybrid feature processing in TabularNN makes them outperform other models, however, in limited setting, MLP has an advantage of being simple and compressible.

In short, each of the three models is performing well, although with different strengths:

• TabularNN for: A: Most accurate (overall judgment) and general

• TabNet: Good recall and F 1 -score, fastest inference

• MLP: The most effective scale and is compressible

•Accuracy

• TabularNN: 99.43

• TabNet: 99.15]

• MLP: 98.45

• Recall:

• TabNet: 0.97

• TabularNN: 0.96

• MLP: 0.83

• F1-score:

• TabNet: 0.97

• TabularNN: 0.95

• MLP: 0.81

5.2 Cohen's Kappa score

All three models had high Kappa values as shown in Table 5.1 meaning strong agreement:

Table 5.1: Cohen Kappa Scores for all three model

Mode	**Cohen Kappa Score**
MLP	0.9700
TabularNN	0.9900
TabNet	0.9800

The best result was 0.9900 performed by TabularNN, which classifies with a high degree of reliability. TabNet came next with 0.9800 that indicated it had a good performance in feature dependencies. The effectiveness of MLP with regard to modeling non-linear patterns within air quality data was confirmed, as it recorded 0.9700.

The obtained outcomes emphasize the compatibility of deep learning models with environmental monitoring. The Kappa values are significantly high; hence, both models can be effectively used to classify the levels of pollutants to achieve an accurate and timely decision-making process

regarding sustainability applications.

These findings support the importance of selecting models in light of the context in which they are to be deployed such as throughaccuracy, interpretability or sustainability.

5.3 Compressed Model Discussion

In order to increase the deployability and sustainability of the baseline MLP model that performs air quality classifications, two techniques of model compression were applied: Post-Training Quantization and Knowledge Distillation (KD). The two approaches were intended to minimize computational overhead without effect significantly predictive performance or interpretability

5.3.1 Post-Trained Quantization

Overview: During quantization model weights and activations are made less precise, usually going from 32-bit floating point numbers to integers of 8-bit precision, without retraining the model.

Benefits:

• **Smaller model size:**ignificant in reduce of memory uasge.

• **Quicker inference:** Low latency and low power High throughput: High-bandwidth and low power Social daydreaming: Low power Social Dreamer 2.0: Low power

• **Hardware-friendly:** Fastest on CPUs, microcontrollers and mobile devices

• **Use Case Fit:**Ideal for real-time air quality prediction on low-power devices, supporting sustainable deployment.

5.3.2 Knowledge Distillation(KD)

Definition: KD is a method that trains a smaller model, which is then referred to as a student, to behave similarly to a larger model, or teacher, by learning the soft output probabilities of the teacher, instead of its hard labels.

Benefits:

• **High accuracy retention:** Student model usually at least the same as, greater than the baseline performance

• **Simplification of models:** Simplification providing better generalization

• **Training efficiency:** accelerated convergence and more fluid learning-dynamics

5.4 Discussion on Accuracy and Loss Graph

5.4.1 Multi Layer Perceptron (MLP)

Accuracy Graph 5.2 The accuracy curve of the baseline-MLP has steady upward movement and a high value of 99.44. The linear form shows a stable learning and little inter epoch variance.

Loss Graph 5.3 The graph of loss shows that the loss is decreasing steadily with the ultimate loss at about 0.1834. It indicates a good generalization and overfitting is minimal since the gap between training and validation loss is small.

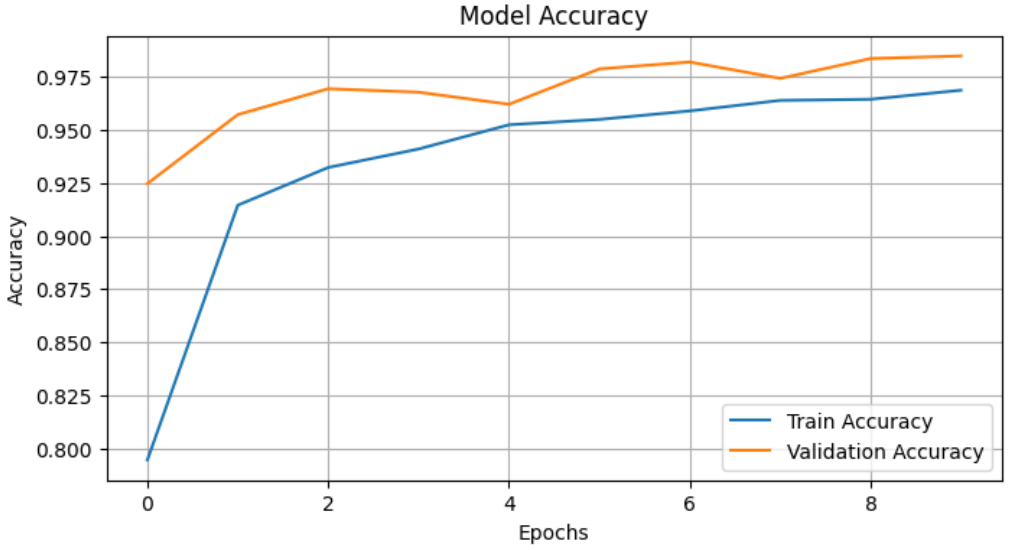

Figure 5.2: MLP accuracy

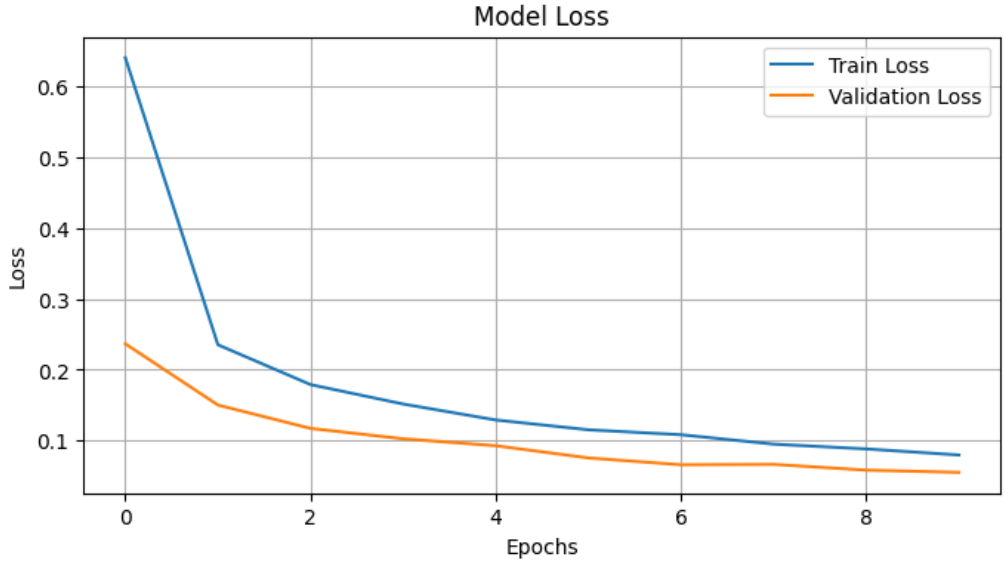

Figure 5.3: MLP Loss

5.4.2 TabularNN

Accuracy Graph 5.5 TabularNN was fast to converge and got high accuracy quickly. It is stable, and steep, which implies that the features are well represented and optimized.

Loss Graph 5.5 There was minimal noise in which loss dropped excessively, and this implies strong learning. The last loss was a bit lesser than MLP, which reaffirmed its performance on tabular data.

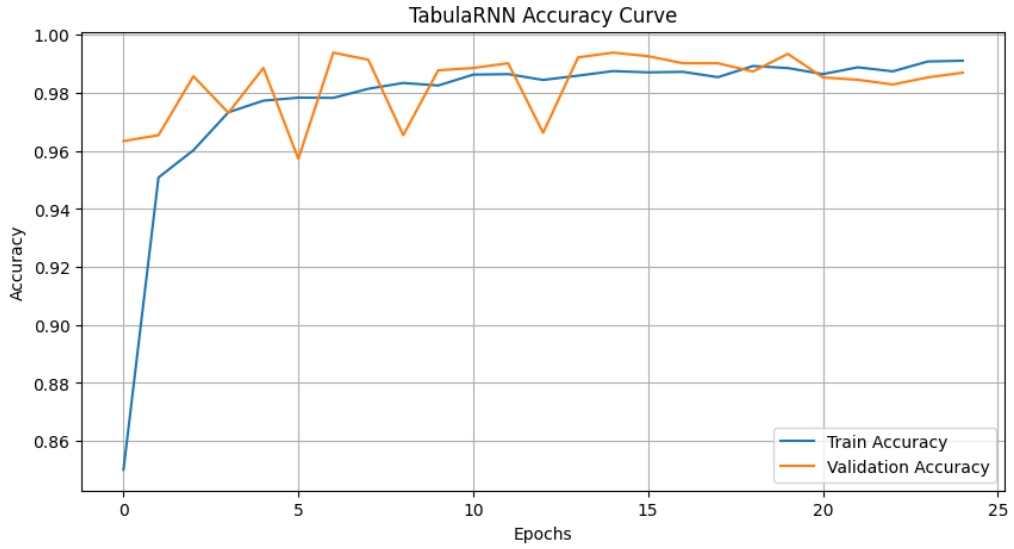

Figure 5.4: TabularNN Accuracy

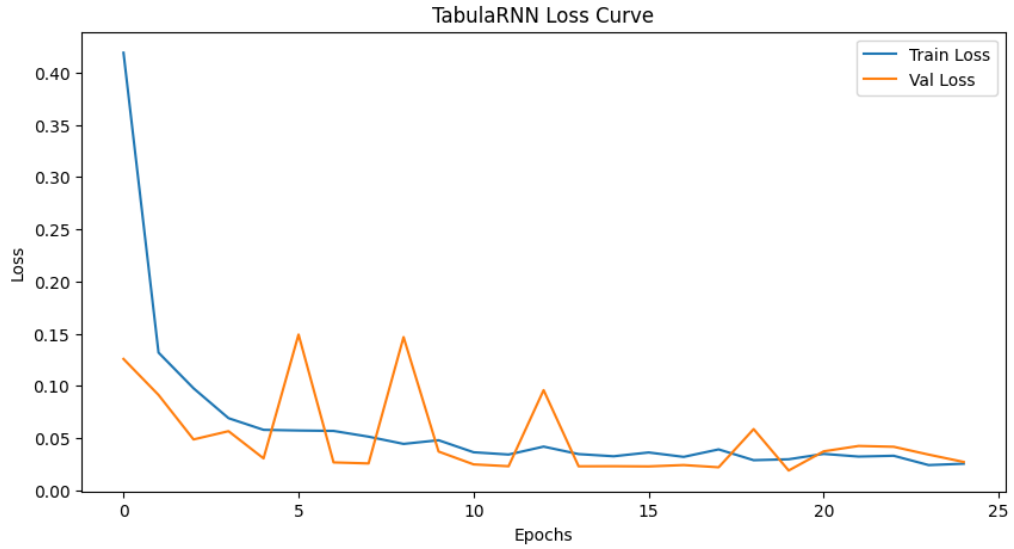

Figure 5.5: TabularNN Loss

5.4.3 TabNet

Accuracy Graph 5.6 TabNet has an accuracy curve with medium variations, because of being based on the attention architecture. Nevertheless, it achieved competitive accuracy of the complex patterns. **Loss Graph** 5.6 he loss plot is more noisy, in accordance with dynamic feature selection in TabNet. Loss was a little lower, an indication of sensitivity to hyperparameters and the problem of overfitting.

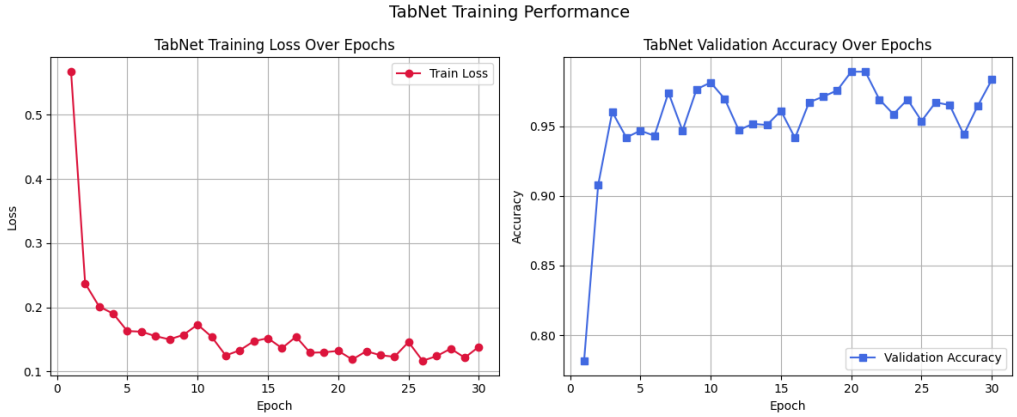

Figure 5.6: TabNet graph

6 CONCLUSIONS

The current project explored how tabular deep learning models can be deployed and compressed to Multiclass air quality prediction on the AIRDATASET. Three models, MLP, TabularNN, and TabNet adopted three architectures and trained and evaluated regarding accuracy, F1-score, Cohen Kappa and model size.

abularNN model had the best overall accuracy (99.43%) and Cohen Kappa (0.9900), which demonstrates that this model was the most predictive and consistent. Compared to, TabNet had competitive performance in terms of strong recall and F1-scores, whereas MLP had lightweight but strong accuracy and minimal computational cost.

In a bid to help the model to become more efficient, two compressions methods namely Post-Training Quantization and Knowledge Distillation were made on the MLP model. The quantized MLP reduced in size by 75 percent and has an almost 2.2 increased speed in inference time, with absolute minimal changes in the accuracy level. The KD-compressed MLP maintained almost the same high levels of performance (99.57 accuracy, 0.1667 loss) and simplified the model structure as well as training-stability.

These findings prove that compressed tabular models are able to provide efficacious and precise predictions of air quality and can be thus implemented in real-time, resource-limited settings.

These findings prove that compressed tabular models are able to provide efficacious and precise predictions of air quality and can be thus implemented in real-time, resource-limited settings.

The project is in line with the important UN Sustainable Development Goals (UNSDGs):

UNSDG 9 (Industry, Innovation, and Infrastructure): through the promotion of scalable AI solutions

UNSDG 12 (Responsible Consumption and Production): By use of fewer computational resources

UNSDG 13 (Climate Action): The efficiency of environmental monitoring made possible by the use of energy allows it to be environmentally friendly

BIBLIOGRAPHY

[1] Agbehadji, I. E. and Obagbuwa, I. C. . Systematic review of machine learning and deep learning techniques for spatiotemporal air quality prediction. *Atmosphere*, 15(11):1352, 2024. doi: 10.3390/atmos15111352. URL https://www.mdpi.com/2073-4433/15/11/1352. Accessed: 2025-08-08.

[2] Arik, S. O. and Pfister, T. . Tabnet: Attentive interpretable tabular learning. *arXiv preprint arXiv:1908.07442*, 2019. URL https://arxiv.org/abs/1908.07442. Accessed: 2025-08-10.

[3] Dantas, P. V. , Silva Jr, W. S. , da, Cordeiro, L. C. , and Carvalho, C. B. . A comprehensive review of model compression techniques in machine learning. *Applied Intelligence*, 54:11804–11844, 2024. doi: 10.1007/s10489-024-05747-w. URL https://link.springer.com/article/10.1007/s10489-024-05747-w. Accessed: 2025-08-10.

[4] Du, L. , Gao, F. , Jia, R. , Chen, X. , Wang, J. , Han, S. , Zhang, J. , and Zhang, D. . Tabularnet: A neural network architecture for understanding semantic structures of tabular data. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21)*. ACM, 2021. doi: 10.1145/3447548.3467228. URL https://www.microsoft.com/en-us/research/wp-content/uploads/2021/05/TabularNet-2.pdf. Accessed: 2025-08-10.

[5] Du, S. , Li, T. , Yang, Y. , and Horng, S.-J. . Deep air quality forecasting using hybrid deep learning framework. *arXiv preprint*, 2018. URL https://arxiv.org/pdf/1812.04783. Accessed: 2025-08-08.

[6] Li, Z. , Li, H. , and Meng, L. . Model compression for deep neural networks: A survey. *Computers*, 12(3):60, 2023. doi: 10.3390/computers12030060. URL https://www.mdpi.com/2073-431X/12/3/60. Accessed: 2025-08-10.

[7] Liu, D. , Zhu, Y. , Liu, Z. , Liu, Y. , Han, C. , Tian, J. , Li, R. , and Yi, W. . A survey of model compression techniques: Past, present, and future. *Frontiers in Robotics and AI*, 12, 2025. doi: 10.3389/frobt.2025.1518965. URL https://www.frontiersin.org/articles/10.3389/frobt.2025.1518965/full. Accessed: 2025-08-10.

[8] Liu, J. , Tian, T. , Liu, Y. , Hu, S. , and Li, M. . itabnet: An improved neural network for tabular data and its application to predict socioeconomic and environmental attributes. *Neural Computing and Applications*, 35:11389–11402, 2023. doi: 10.1007/s00521-023-08304-7. URL https://link.springer.com/article/10.1007/s00521-023-08304-7. Accessed: 2025-08-10.

[9] Mishra, A. and Gupta, Y. . Comparative analysis of air quality index prediction using deep learning algorithms. *Environmental Monitoring and Assessment*, 32:63–72, 2023. doi: 10.1007/s41324-023-00541-1. URL https://link.springer.com/article/10.1007/s41324-023-00541-1. Accessed: 2025-08-08.

APPENDIX

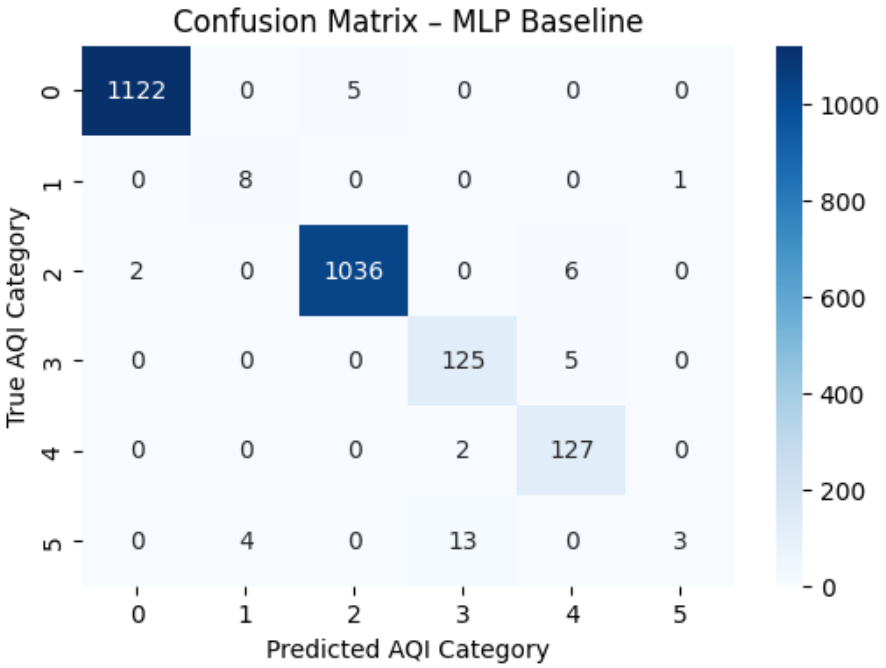

Figure 6.1: confusion martix MLP

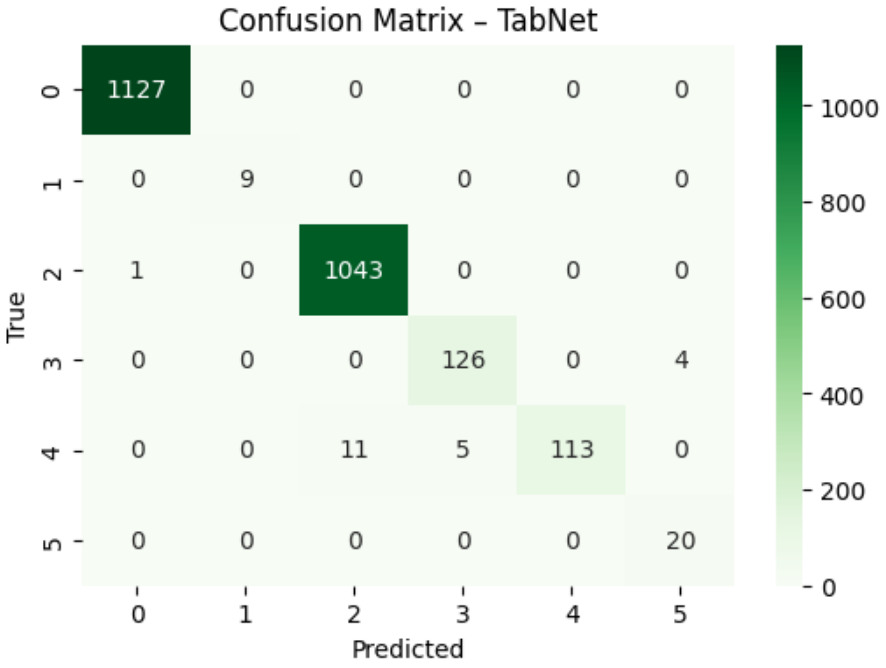

Figure 6.2: confusion martix TabNEgt

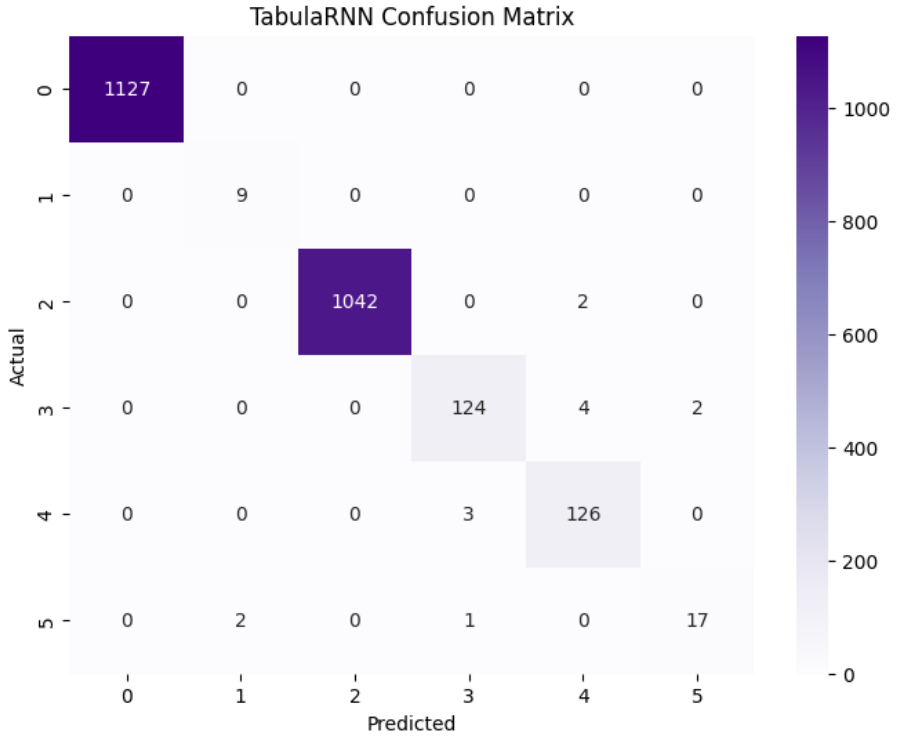

Figure 6.3: confusion martix TabularNN

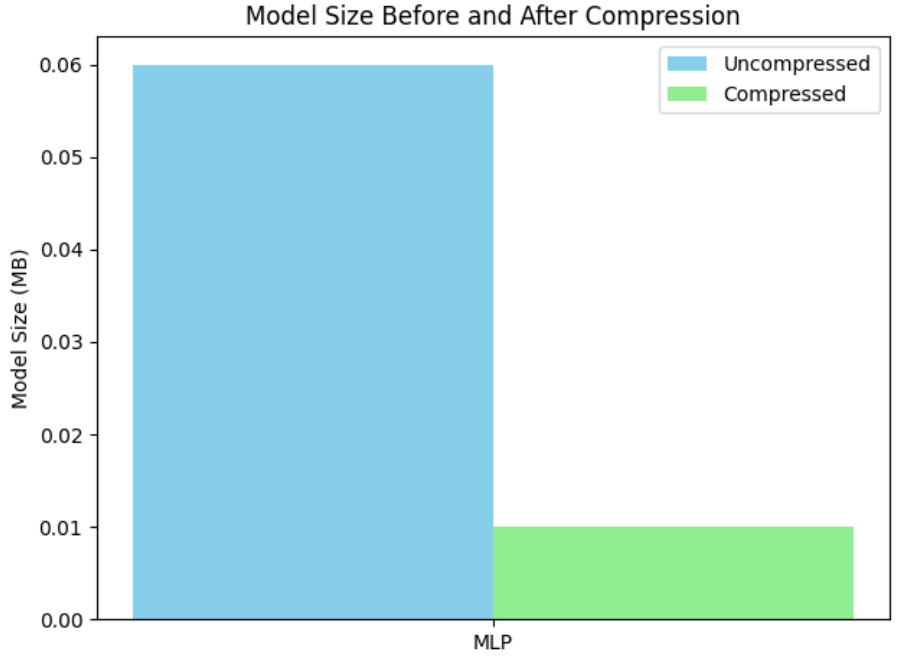

Figure 6.4: Model size graph before and after compression

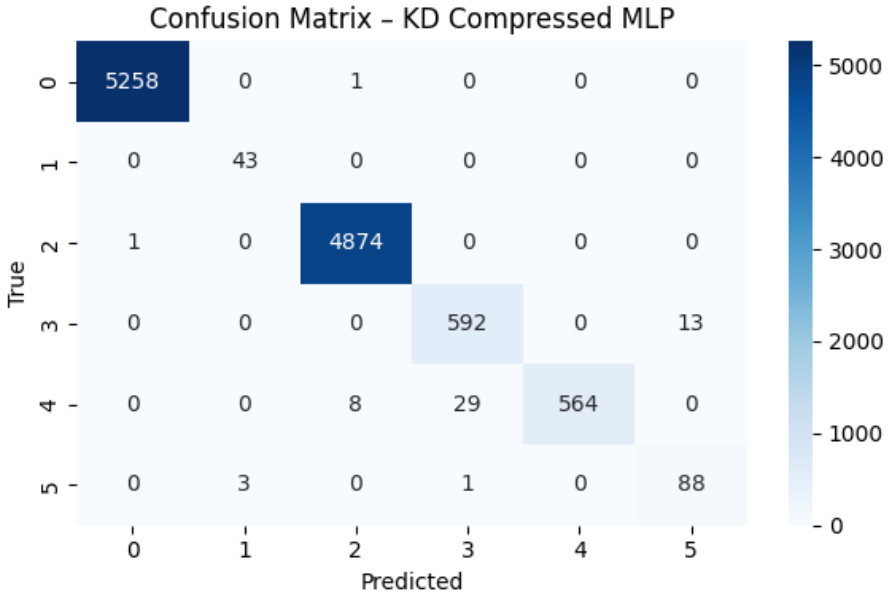

Figure 6.5: Confusion Matrix of the Compressed Model (TFLite)

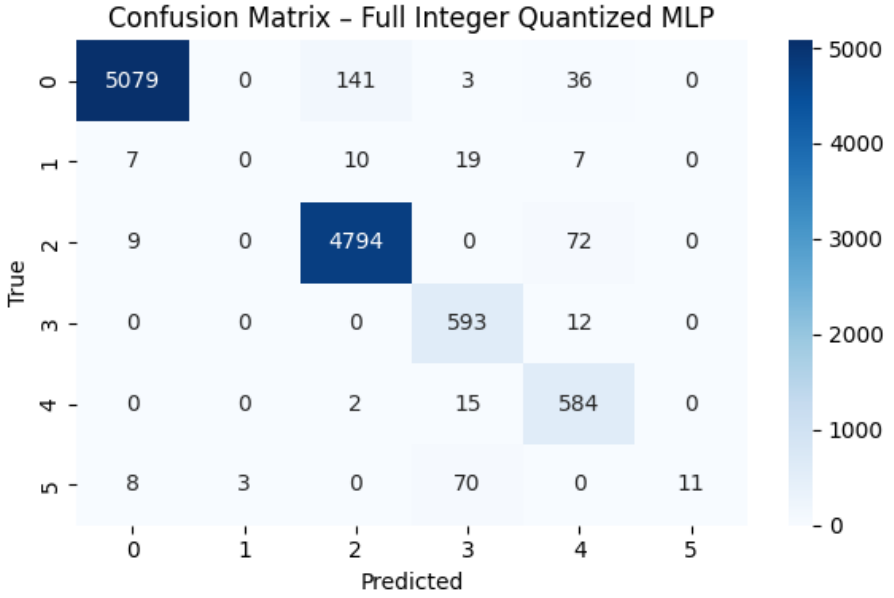

Figure 6.6: Confusion Matrix of the Compressed Model (Knowledge Distilled