Deep Learning for Wireless Community Networks (DL4WCN)

Community Networks are networks funded and established by communities where ISPs have not provided connectivity. They provide affordable internet connectivity to underserved areas but are often resource-constrained. Accurate and efficient classification is vital for maintaining Quality of Service

Objective: Develop and evaluate Deep learning models to optimise network traffic classification for resource-constrained community networks

Aim: Find an accurate and efficient deep learning model capable of classifying network traffic in real-time

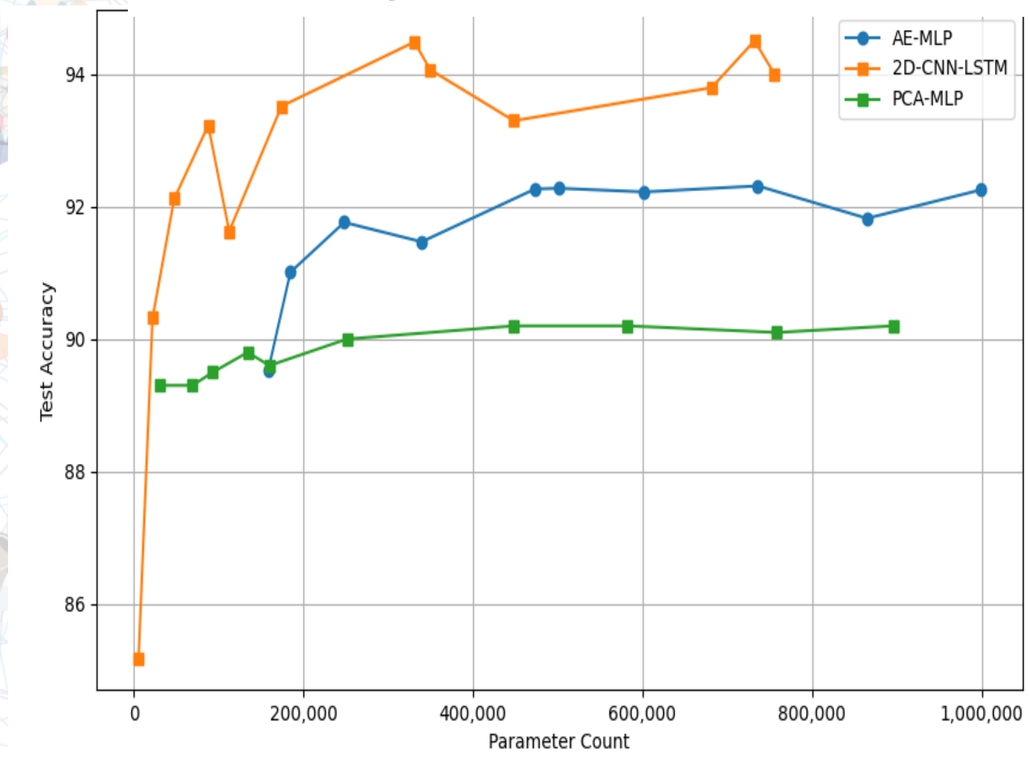
Implementation:

Data sourced from the Ocean View Community Network for model training. Preprocessing procedure included packet labelling, feature extraction and data balancing. Models classify network traffic into 10 categories (e.g. YouTube, Google Services, WhatsApp etc...)

Hybrid Model Design:

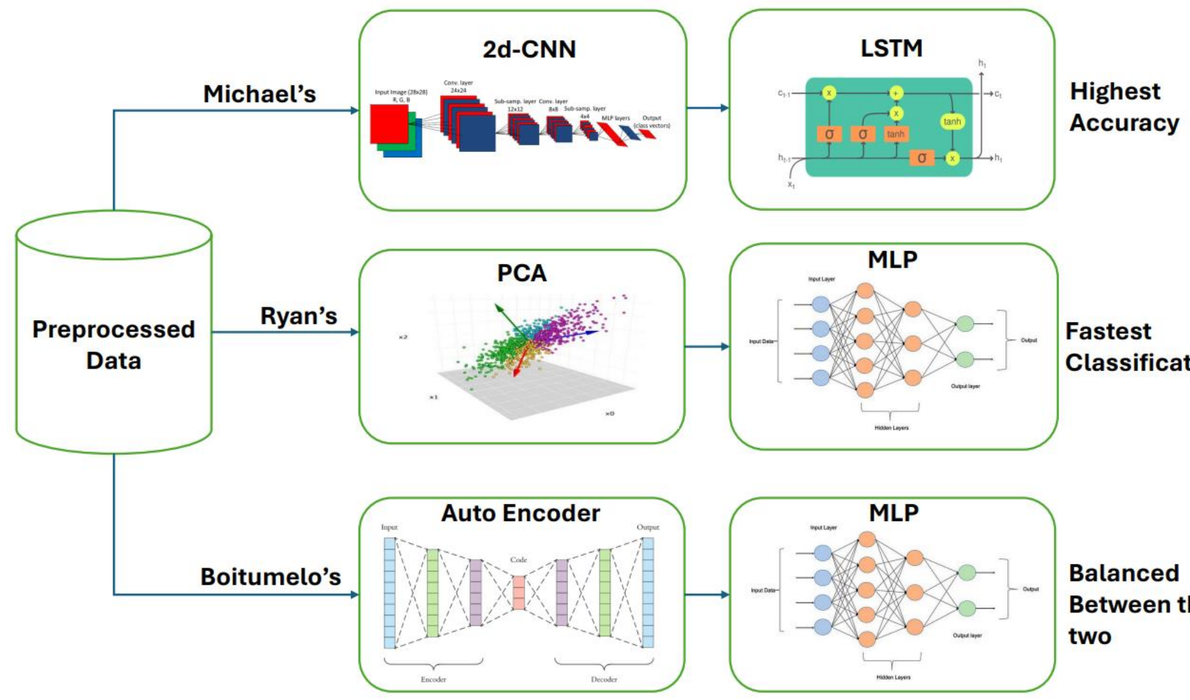
Used Principal Component Analysis (PCA) and Auto-Encoders (AEs) for unsupervised dimensionality reduction. Multi-Layer Perceptron, Convolutional Neural Networks and Long Short-Term Memory units were used for classification to maximize accuracy.

Test Accuracy vs Number of Parameters



Best Performing Models:

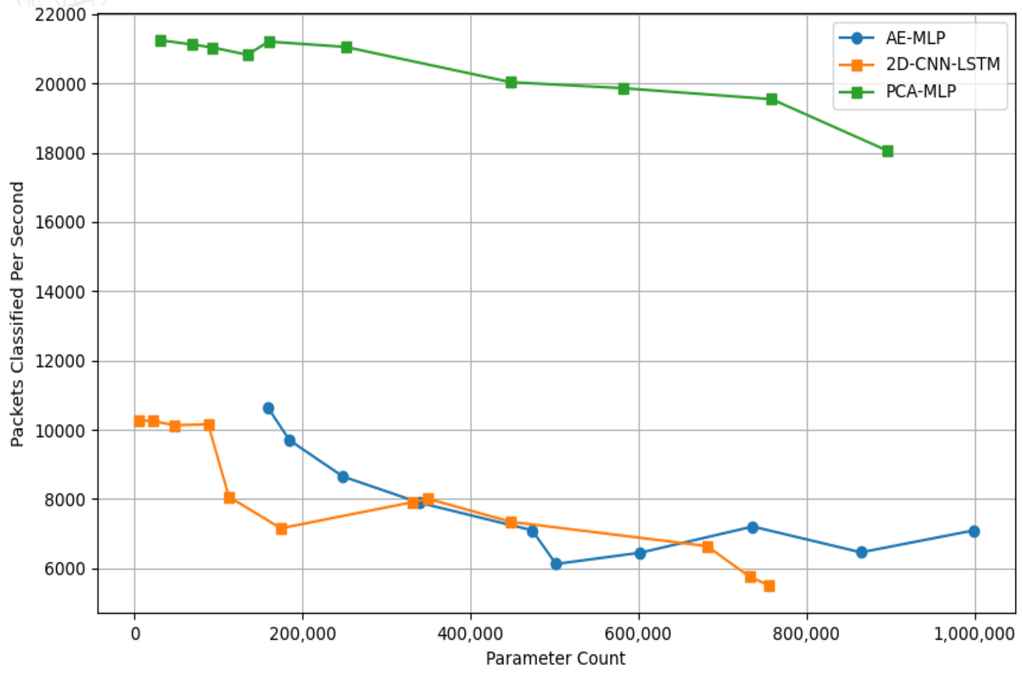
Michael: 2d-CNN LSTM hybrid model
Ryan: PCA-MLP semi-supervised model
Boitumelo: AE-MLP semi-supervised model



Findings:

The 2D-CNN-LSTM achieves the highest accuracy (~94%) but sacrifices speed, with a significant drop in packets classified per second as parameters increase. While the AE-MLP can extract patterns more accurately, reaching moderate accuracy (~92%), it experiences a decline in speed and accuracy beyond 600,000 parameters. In contrast, PCA-MLP can extract patterns much more efficiently, consistently classifying over 20,000 packets per second, though with a lower accuracy (~90%). These results highlight the trade-offs between model complexity, classification speed, and accuracy in traffic classification tasks.

Packets Classified per Second vs Parameter Count



Conclusions:

An increase in the number of parameters in a model results in the rapid decrease in classification efficiency after a certain point. A model size of ~ 500,000 – 600,000 parameters is an optimal point for an efficiency vs accuracy trade-off. Hybrid Models are effective at leveraging the strengths of different Deep Learning architectures for this task. Unsupervised Dimensionality reduction is an effective way to improve model efficiency while maintaining accuracy. 2D-CNNs are great at extracting spatial patterns in network traffic while the LSTM is effective at extracting temporal features.

Team Members:

Michael Gamsu
Ryan Murphy
Boitumelo Mokoka

University of Cape Town
Department of Computer Science Dept.
Email: dept@cs.uct.ac.za
Tel: 021 650 2663

Supervised By:
Josiah Chavula

Co-Supervised By:
Patrick Marais