DL-Models report

Maxime Bossant

June 2025

Introduction

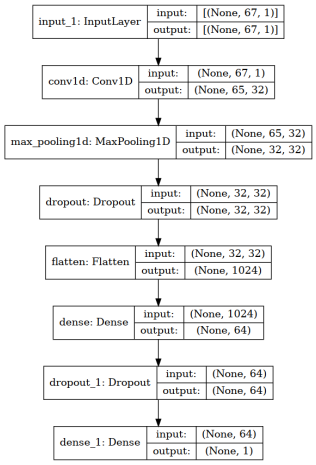
- CNN is more efficient to capture local patterns in data. It takes less time to train. It is used more frequently for image processing but can be used for datasets like ours.
- LSTM is supposed to be more adapted to capture temporal dependencies in data.
- AE-LSTM is supposed to be more robust than classic models like LSTM. More robust to overfitting and more generalizable.

We finally expect that AE-LSTM will outperform both CNN and LSTM in terms of generalization and robustness, especially when handling complex traffic patterns. The LSTM model is expected to achieve better results than CNN due to its ability to capture temporal dynamics within network flows. Therefore, we anticipate the following hierarchy in performance:

$$\text{AE-LSTM} \geq \text{LSTM} > \text{CNN} \quad \text{(particularly in terms of AUC and recall)}$$

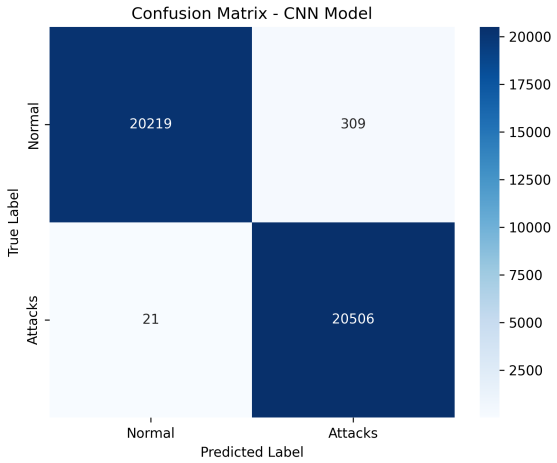
InSDN

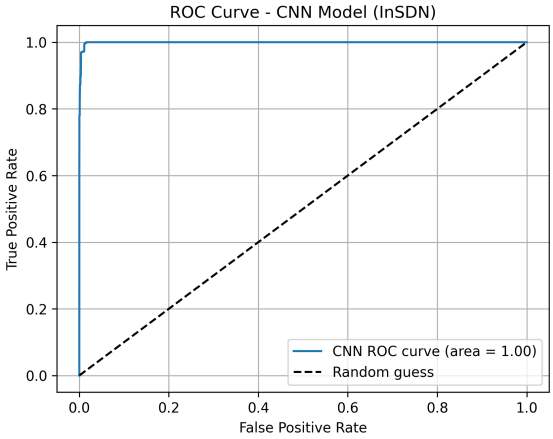
CNN



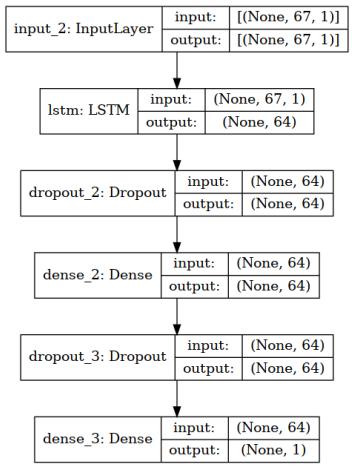
Test Loss: 0.0273
Test Accuracy: 0.9920

	Precision	Recall	F1-score
Normal	1.00	0.98	0.99
Attacks	0.99	1.00	0.99



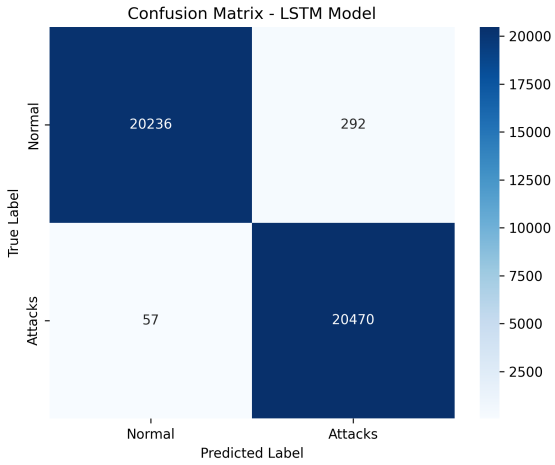


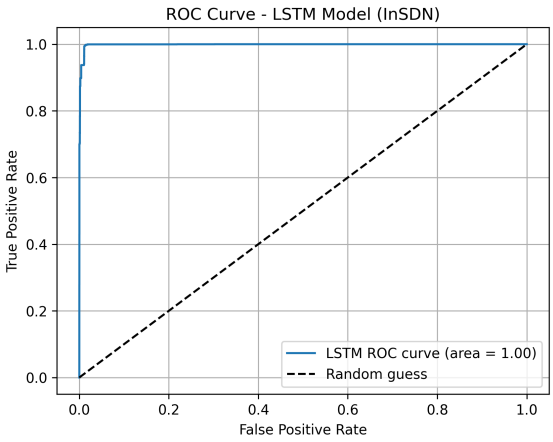
LSTM



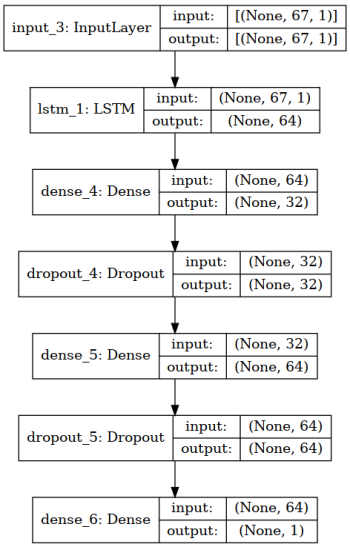
Test Loss: 0.0323
Test Accuracy: 0.9915

	Precision	Recall	F1-score
Normal	1.00	0.99	0.99
Attacks	0.99	1.00	0.99



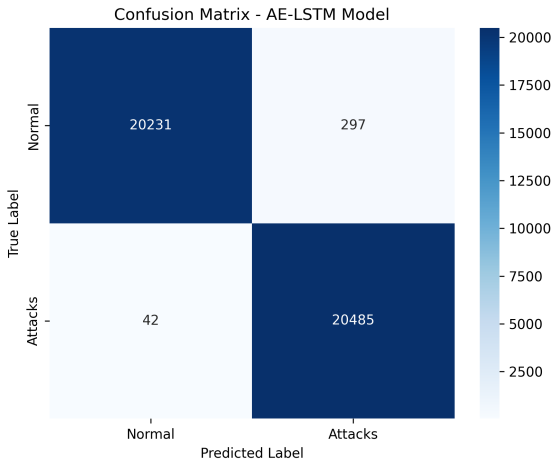


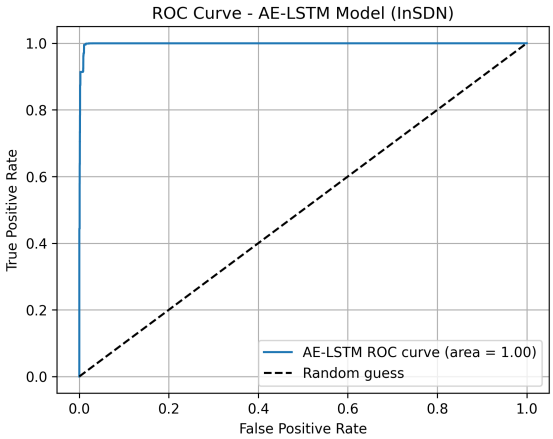
AE-LSTM



Test Loss: 0.0361
Test Accuracy: 0.9917

	Precision	Recall	F1-score
Normal	1.00	0.99	0.99
Attacks	0.99	1.00	0.99





Discussion of Results

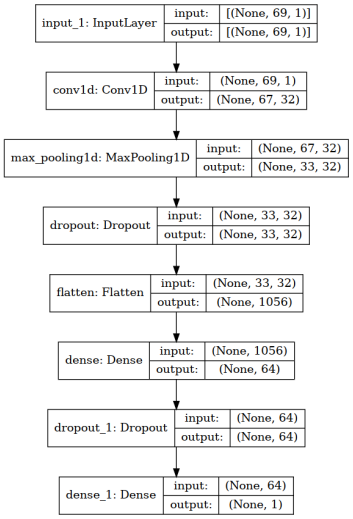
All three models (CNN, LSTM, AE-LSTM) perform very well on this dataset. They all reach almost perfect accuracy and F1-scores for both *Normal* and *Attack* classes. The confusion matrices show very few errors, and the ROC curves are close to ideal.

Conclusion: The InSDN dataset seems to be not very difficult, and even a simple CNN model works well. There is no strong sign of overfitting : the test set also have good accuracies. Using a more complex model like AE-LSTM is not necessary in this case.

CICIDS2017

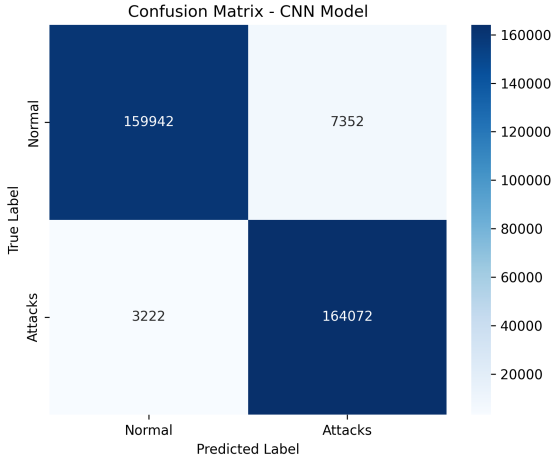
CNN

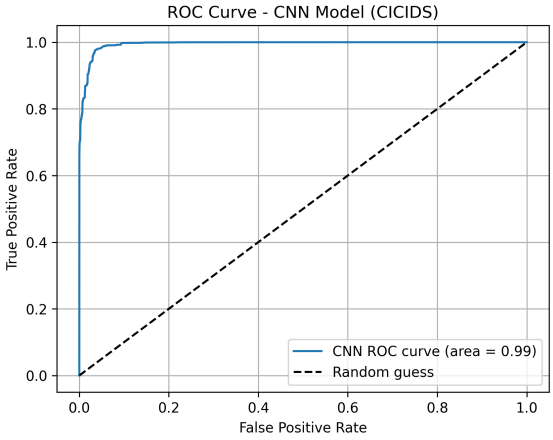
5 epochs



Test Loss: 0.0894
Test Accuracy: 0.9684

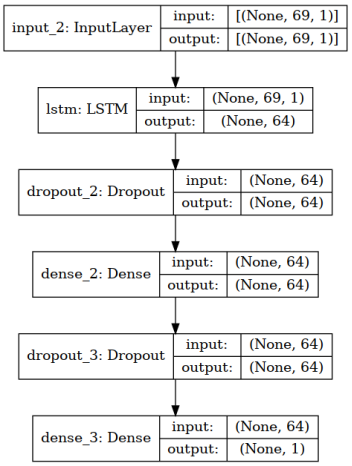
	Precision	Recall	F1-score
Normal	0.98	0.96	0.97
Attacks	0.96	0.98	0.97





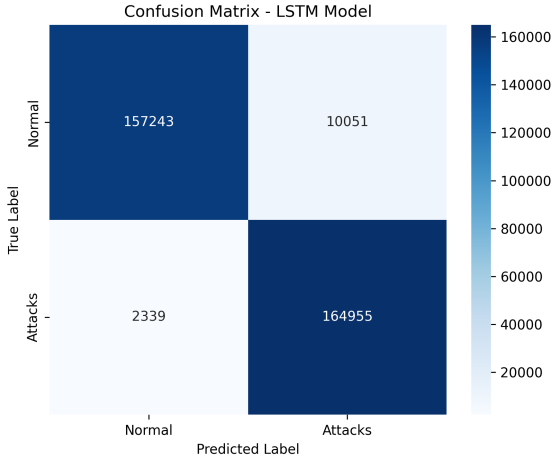
LSTM

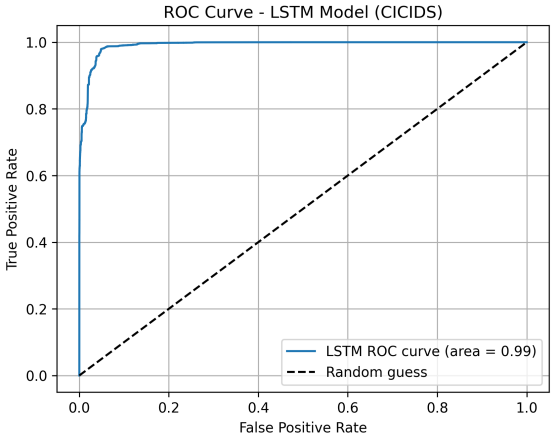
5 epochs



Test Loss: 0.1084
Test Accuracy: 0.9630

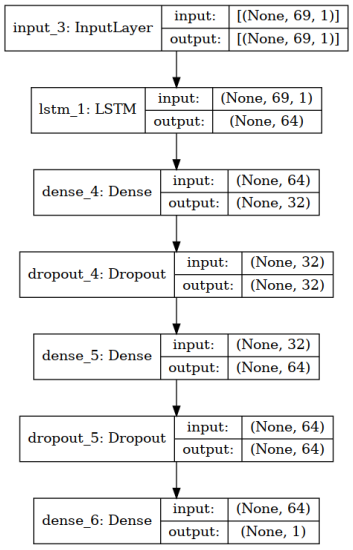
	Precision	Recall	F1-score
Normal	0.99	0.94	0.96
Attacks	0.94	0.99	0.96





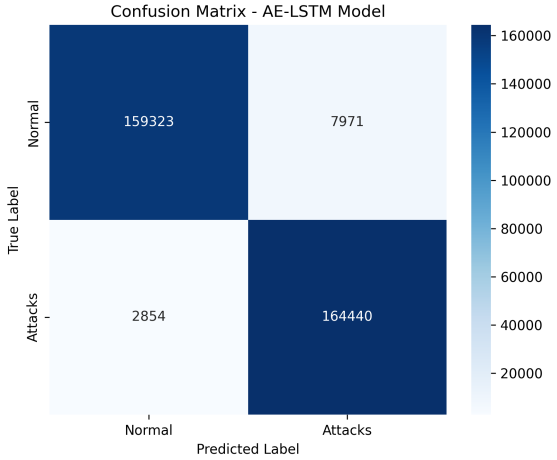
AE-LSTM

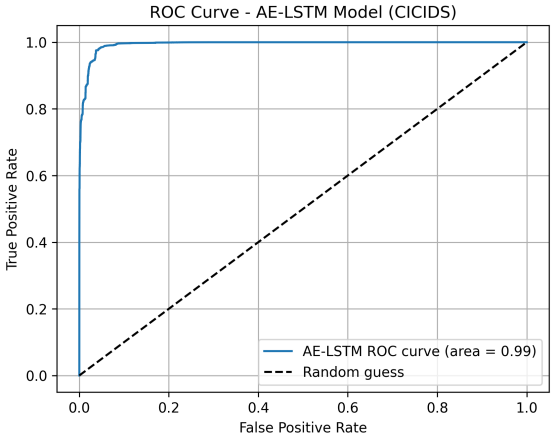
5 epochs



Test Loss: 0.0947
Test Accuracy: 0.9676

	Precision	Recall	F1-score
Normal	0.98	0.95	0.97
Attacks	0.95	0.98	0.97





Discussion of Results

All three models achieve very similar performance on the CICIDS2017 dataset, with an accuracy around 96.9%. LSTM and AE-LSTM slightly outperform CNN on some metrics, especially for detecting attacks.

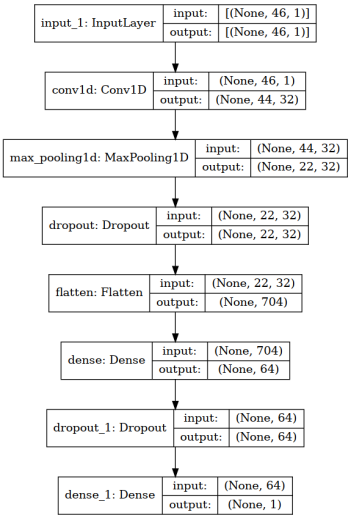
The confusion matrices and ROC curves confirm that all models are good at separating normal and malicious traffic. The differences between models remain small, even if LSTM-based models tend to generalize slightly better.

Conclusion: This dataset is more complex than InSDN, and all models perform well despite this. AE-LSTM still does not bring a major improvement over the classic LSTM. However, both LSTM and AE-LSTM detect attacks more reliably than CNN, which may be important in security contexts.

CICIOT2023

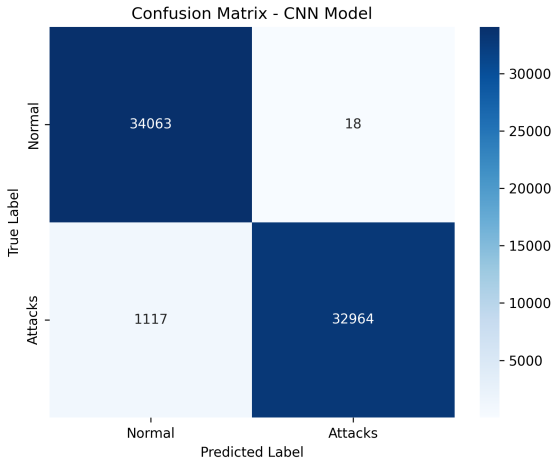
CNN

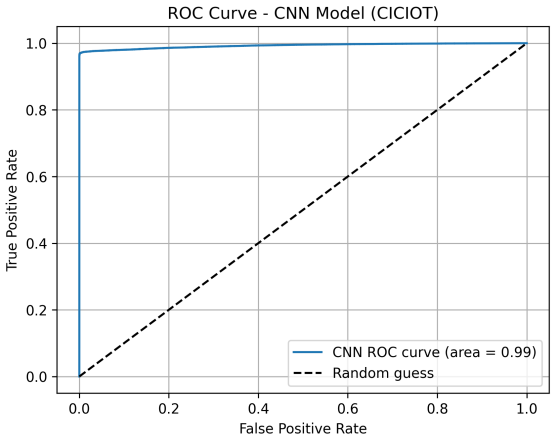
5 epochs



Test Loss: 0.0649
Test Accuracy: 0.9833

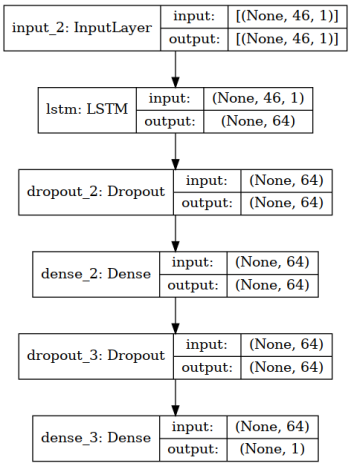
	Precision	Recall	F1-score
Normal	0.97	1.00	0.98
Attacks	1.00	0.97	0.98





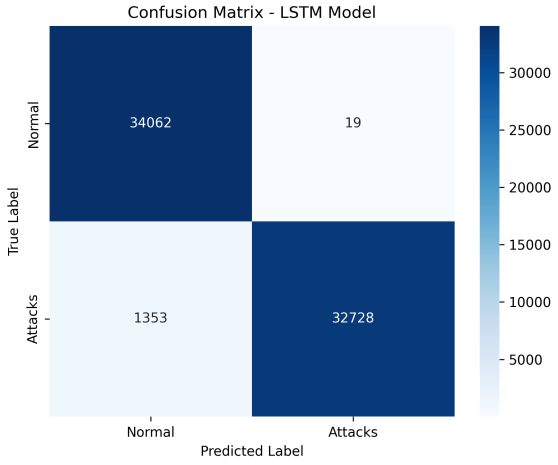
LSTM

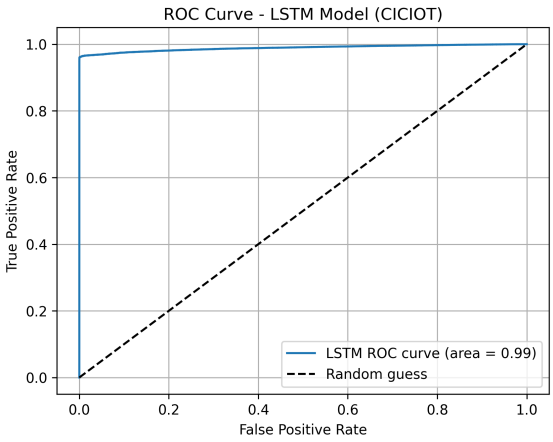
5 epochs



Test Loss: 0.0649
Test Accuracy: 0.9833

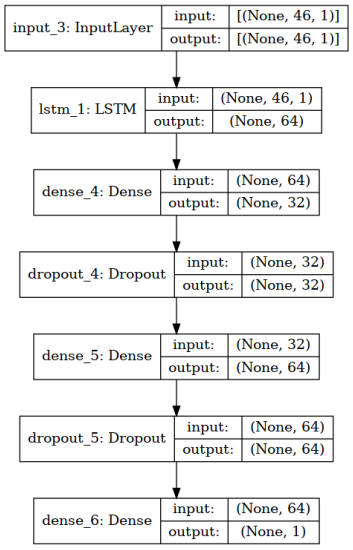
	Precision	Recall	F1-score
Normal	0.96	1.00	0.98
Attacks	1.00	0.96	0.98





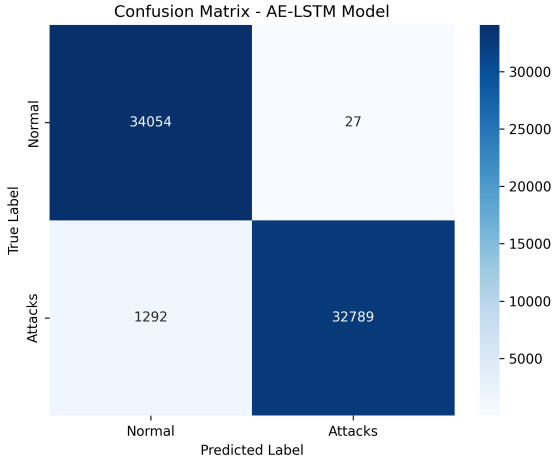
AE-LSTM

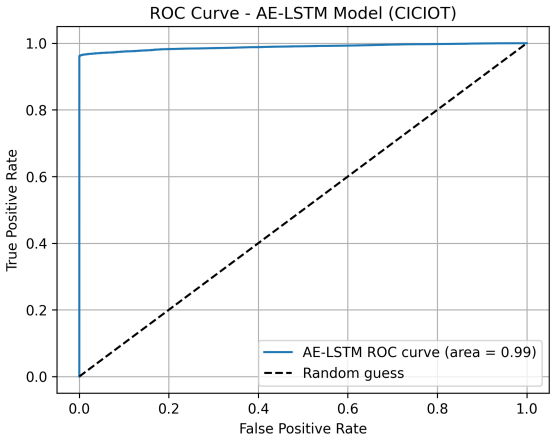
5 epochs



Test Loss: 0.0649
Test Accuracy: 0.9833

	Precision	Recall	F1-score
Normal	0.96	1.00	0.98
Attacks	1.00	0.96	0.98





Discussion of Results

After balancing the dataset, all models (CNN, LSTM, AE-LSTM) show very similar performance. The accuracy is high (around 98.5%) and the F1-scores are nearly identical for both classes.
The confusion matrices confirm this: *Normal* and *Attack* samples are detected with good precision and recall. The ROC curves also show strong classification performance.
Conclusion: Balancing the dataset clearly helped the models to better detect both classes equally. The AE-LSTM model does not offer a clear advantage in this case. All models work well, and a simple CNN may be enough in this situation.

Conclusion

In this report, we compared three deep learning models (CNN, LSTM, AE-LSTM) across three different datasets. The results show that all models can reach high accuracy, especially when the datasets are well-preprocessed and balanced. I don't know if these high accuracies are realistic, maybe it's because of my

models or because of the dataset. We also observed that CNN, even though simpler, often performs nearly as well as more complex models.

The AE-LSTM model, which combines encoding and classification, did not clearly outperform the classic LSTM. In future work, the autoencoder part of the AE-LSTM should be improved or redesigned to better capture useful features.

Overall, model performance depends more on data quality and balance than on complexity alone.