## Literature Reviews:

* [Survey of intrusion detection systems: techniques, datasets and challenges](https://doi.org/10.1186/s42400-019-0038-7)
  + Taxonomy of IDS
* [Network intrusion detection system: A systematic study of machine learning and deep learning approaches](https://doi.org/10.1002/ett.4150)
  + Literature review of the field
* [A Review of the Advancement in Intrusion Detection Datasets](https://doi.org/10.1016/j.procs.2020.03.330)
* [A survey and analysis of intrusion detection models based on CSE-CIC-IDS2018 Big Data](https://doi.org/10.1186/s40537-020-00382-x)
  + Literature review of work on CSE-CIC-IDS 2018

# [Overview on Intrusion Detection Systems Design Exploiting Machine Learning for Networking Cybersecurity](https://doi.org/10.3390/app13137507)

* + literature review of the field published May 2023 (considers 3 datasets including CICIDS2018)

Possibly of interest: (meaning not yet read)

### [Towards Enhancement of Machine Learning Techniques Using CSE-CIC-IDS2018 Cybersecurity Dataset](https://www.proquest.com/openview/6223ef80659ae148bd418cd6fb01b2fd/1?pq-origsite=gscholar&cbl=18750&diss=y)

### [Deep learning approaches for detecting DDoS attacks: A systematic review](https://link.springer.com/article/10.1007/s00500-021-06608-1)

* [Machine learning approaches to IoT security: A systematic literature review](https://doi.org/10.1016/j.iot.2021.100365)

# Deep learning methods in network intrusion detection: A survey and an objective comparison

## Other Works:

* [A Data-Centric Anomaly-Based Detection System for Interactive Machine Learning Setups](https://doi.org/10.5220/0011560100003318)
  + Anomaly detection for ML IoT systems. Employs several shallow learning techniques, random forest is the top performer
* [HCRNNIDS: Hybrid Convolutional Recurrent Neural Network-Based Network Intrusion Detection System](https://doi.org/10.3390/pr9050834)
  + DL (CRNN) based IDS on CSE-CIC-IDS2018, that addresses class imbalance
* [A New Ensemble-Based Intrusion Detection System for Internet of Things](https://doi.org/10.1007/s13369-021-06086-5)
  + Intrusion detection using ensemble of 3 shallow learning techniques on CICIDS2017
* [Anomaly Detection Using Deep Neural Network for IoT Architecture](https://doi.org/10.3390/app11157050)
  + Feed forward DNN on IoT-Botnet 2020
* [Network intrusion detection using multi-architectural modular deep neural network](https://doi.org/10.1007/s11227-020-03410-y)
  + Composite approach using 4 different DNN architectures aggregated on CSE-CIC-IDS 2018
* [Towards Detecting and Classifying Network Intrusion Traffic Using Deep Learning Frameworks](https://www.researchgate.net/profile/Ram-Basnet/publication/337936440_Towards_Detecting_and_Classifying_Network_Intrusion_Traffic_Using_Deep_Learning_Frameworks/links/5df5c7b992851c83647e7dc1/Towards-Detecting-and-Classifying-Network-Intrusion-Traffic-Using-Deep-Learning-Frameworks.pdf)
  + Multilayer perceptron on CSE-CIC-IDS 2018 comparing different ML frameworks
* [A two-level anomaly detector for multiple attack classes](https://link.springer.com/chapter/10.1007/978-3-030-44038-1_63)
  + deep autoencoder on CSE-CIC-IDS 2018
* I’m not listing them all but literature review on CSE-CIC-IDS 2018 explores a lot of similar works

### [Method of analysing computer traffic based on recurrent neural networks](http://doi.org/10.1088/1742-6596/1353/1/012133)

* + Attempts to address class imbalance of CSE-CIC-IDS 2018 using recurrent neural networks and focal loss function. Highly theoretical, does not provide computing environment
* [​​CatBoost-Based Network Intrusion Detection on Imbalanced CIC-IDS-2018 Dataset](http://doi.org/10.7840/kics.2021.46.12.2191)
  + Forest ML techniques on CICIDS2018, addresses class imbalance through sampling (which improves the accuracy)
* [Handling class Imbalance problem in Intrusion Detection System based on deep learning](https://doi.org/10.15803/ijnc.12.2_467)
  + Title says it all, deep learning on 4 datasets including CIC-IDS 2018 that addresses class imbalance through oversampling and Synthetic Minority Over-Sampling

# [Long-Term Prediction of Network Security Situation Through the Use of the Transformer-Based Model](http://doi.org/10.1109/ACCESS.2022.3175516)

* + Transformer based IDS on UNSW-NB15

# [RTIDS: A Robust Transformer-Based Approach for Intrusion Detection System](http://doi.org/10.1109/ACCESS.2022.3182333)

* + transformer-based IDS on CICIDS2017 and CIC-DDoS2019, addressed class imbalance

# [Paying attention to cyber-attacks: A multi-layer perceptron with self-attention mechanism](https://doi.org/10.1016/j.cose.2023.103318)

* + Multilayer perceptron using attention mechanism on CICIDS2018 and ISCX 2016
* [Performance analysis of intrusion detection for deep learning model based on CSE‑CIC‑IDS2018 dataset](http://doi.org/10.11591/ijeecs.v26.i2.pp1165-1172)
  + LSTM on CICIDS 2018, addresses class imbalance, suitable for template
* [Towards Enhancement of Machine Learning Techniques Using CSE-CIC-IDS2018 Cybersecurity Dataset](https://www.proquest.com/docview/2512306400?pq-origsite=gscholar&fromopenview=true)
  + RF, logistic regression and KNN on CSE-CIC-IDS 2018 (discusses but does not address class imbalance)

# [Detecting web attacks using random undersampling and ensemble learners](https://doi.org/10.1186/s40537-021-00460-8)

* + Applies DT and RF to CSE-CIC-IDS 2018 (state of the art)

# [Optimization of Intrusion Detection Using Likely Point PSO and Enhanced LSTM-RNN Hybrid Technique in Communication Networks](http://doi.org/10.1109/ACCESS.2023.3240109)

* + Variation of LSTM technique applied to CSE-CIC-IDS 2018 (state of the art)
  + Reproducing this and address future work through XAI is an option

# [Evaluating Standard Feature Sets Towards Increased Generalisability and Explainability of ML-Based Network Intrusion Detection](https://doi.org/10.1016/j.bdr.2022.100359)

* + Focuses on which feature set is best for generalising across datasets (which would have different attack scenarios), includes XAI

# [A False Sense of Security?: Revisiting the State of Machine Learning-Based Industrial Intrusion Detection](https://doi.org/10.1145/3494107.3522773)

* + THE HOLY GRAIL!

General Notes:

* Use bracketing to present acronyms and then add an acronym section
* Add a related work section following introduction

The Edge (least to most relevant):

* Predict the severity of the attack
* Continuous learning model
* Multiclass prediction
* Consider speed
* Consider Snort vs ML or other non-ML techniques
* Take raw packet data and see if you can push limits with it
* Most people use ML as a black box, see if you can tinker with some of the techniques
* Explainability
* also consider unsupervised techniques
* Forget all of that, do random forest and some DL approach, and check which one generalises better to never been seen before attacks/other datasets (representing other network environments)

Branches:

HIDS or NIDS

Dataset from scratch vs adding onto existing dataset

Gaps:

* “The effects of this imbalance can be mitigated by techniques at the data level (e.g. random undersampling, feature selection) and algorithm level (e.g. cost-sensitive classification, ensemble techniques) [[9](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00382-x#ref-CR9)].”
* “None of the papers satisfactorily discuss the data cleaning of CICIDS2018. This is a significant revelation. About 60% of data scientists believe that no task is more time-consuming than data cleaning [[12](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00382-x#ref-CR12)].”
* “Finally, our survey shows that statistical analysis of performance scores appears to have been overlooked.”
* Consult the gaps in current research section whilst writing this paper
* Transfer learning may be an interesting concept to experiment with (has been explored already in 2021 and 2022)
* “Future work with the CSE-CIC-IDS2018 dataset can investigate other families of attacks, individual web attack labels (as compared to the combined web attack labels used in this study), and the effects of rarity [[50](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00460-8#ref-CR50)]. Other datasets could also be included for future work, as well as additional performance metrics, classifiers, and sampling techniques.”

Datasets:

* [CSE-CIC-IDS 2018](https://www.semanticscholar.org/paper/Toward-Generating-a-New-Intrusion-Detection-Dataset-Sharafaldin-Lashkari/a27089efabc5f4abd5ddf2be2a409bff41f31199) (top contender)
* CIC-IDS 2017 (is a simpler, smaller predecessor)
* NF-UQ-NIDS (giant amalgamation of like 5 [datasets](https://staff.itee.uq.edu.au/marius/NIDS_datasets/#RA5))
* [IoT-botnet 2020](https://doi.org/10.1109/SMC42975.2020.9283220)

## CIC-MalMem-2022 (modern but only contains memory signatures of obfuscated malware, good for HIDS)

Useful references:

* [Average cost of a cyber attack is $200,000](https://www.cnbc.com/2019/10/13/cyberattacks-cost-small-companies-200k-putting-many-out-of-business.html.)
* [Balancing Approaches towards ML for IDS: A Survey for the CSE-CIC IDS Dataset](http://doi.org/10.1109/ICCSPA49915.2021.9385742)
  + Highlights significance of class imbalance

Interesting

# [Enabling Anomaly-based Intrusion Detection Through Model Generalization](http://doi.org/10.1109/ISCC.2018.8538524)

* + Work on improving generalisability of a model to other datasets

# [Toward a reliable anomaly-based intrusion detection in real-world environments](https://doi.org/10.1016/j.comnet.2017.08.013)

* + Work on making AIDS studies more realistic and practically applicable

# [Clustering-based network intrusion detection](https://doi.org/10.1142/S0218539307002568)

**The Simulation Strategy:**

Look into the creation of CICIDS2018

DVWA is a great tool

Look into network simulation

~~Look into honeypot tools~~

Worry about user simulation

Look into something like 2016 mcafee report for choosing attacks

GNS3 can simulate virtually any network architecture

CICFlowMetre can process pcap files into features

**Whats next:**

The ideal, is to find the state of the art, reproduce it, and implement some of its future work

Figure out what research has been done on generalisability to unseen attacks

Craft a strategy to test the hypothesis, the interesting section articles may help create the kind of test data you’ll need

Look into network simulators, they may help in both generation of dataset or generation of synthetic samples (regarding this, it is certainly possible, how difficult it will be is yet to be seen, biggest obstacle will likely be resources)