NETWORK DATA ANALYSIS WITH NEURAL NETWORKS TO IMPROVE SECURITY

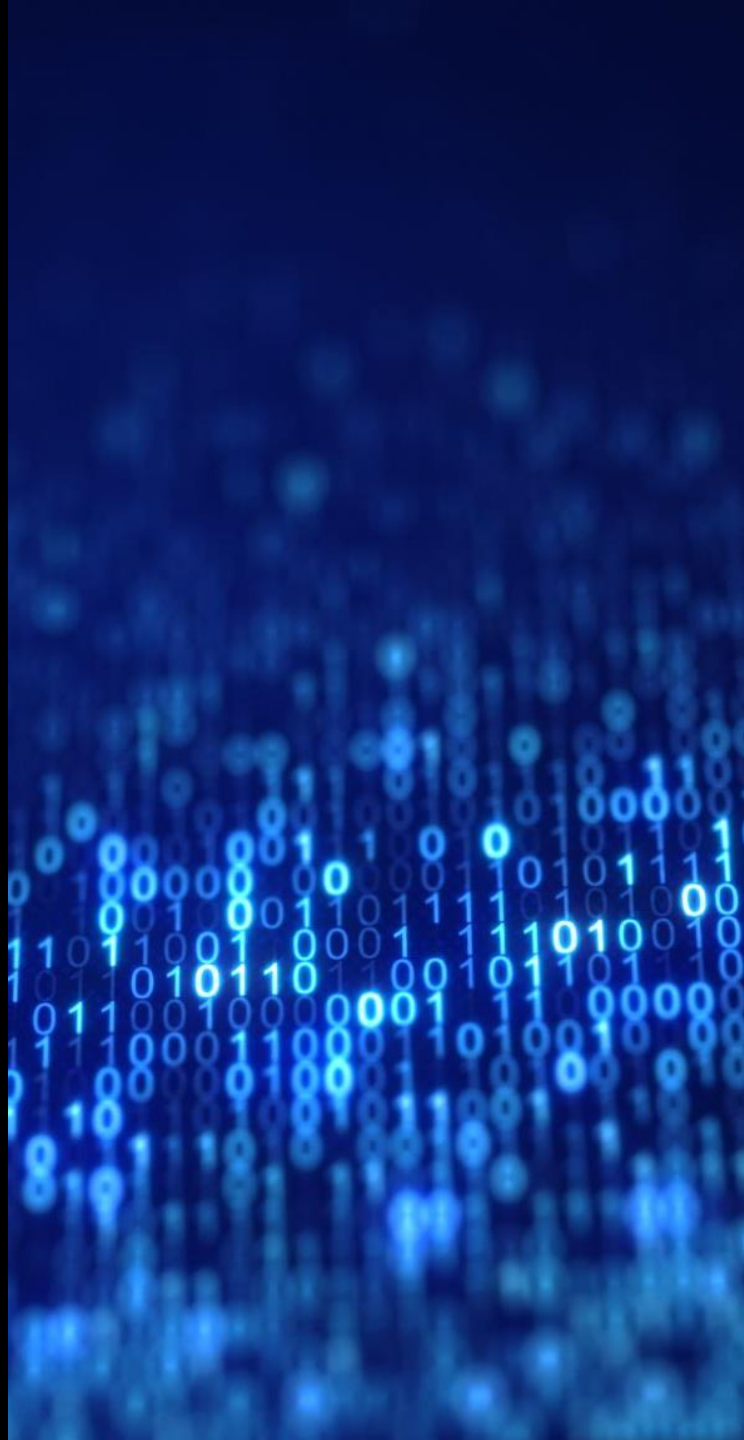
Cheemala Astha Diamond – 700746277
Chadive Varshitha Reddy – 700747558
Gnana Deepthi Pulipati – 700741831
Lingisetty Lakshmi Charan - 700756902

MOTIVATION

- The primary motivation of Malware Identification analysis with Deep Learning on Neural Networks is to detect the find out the Malware in the network dataset.
- In this work, the dataset containing the malware dataset will be taken into consideration.
- The pre-processing will be applied in to the dataset and the noisy and null value data will be removed from the dataset.
- After the data will be analysed and visualized for further processing. The Convolutional Neural Networks algorithm will be chosen to implementation process.
- The project evaluation can be tested with the deep learning algorithm prediction results
- Since the Convolutional Neural Networks algorithm will be used to predict the Malware, the accuracy of the algorithm result will be helpful to evaluate the results.

OBJECTIVE

- The objective of Malware identification with deep learning is to detect the Malware in the network with the available attributes.
- In this work, the dataset containing the malware dataset will be taken into consideration.
- The primary contribution is to apply the deep learning to detect the Malware.
- The pre-processing will be applied in to the dataset and the noisy and null value data will be removed from the dataset.
- After the data will be analysed and visualized for further processing.
- The Deep Learning neural network algorithm will be chosen to make the good accuracy prediction.
- It will be helpful in all the distributed network records to detect the Malware



RELATED WORK

Liu et al. in 2017 provided the system with three components: data processing, decision making, and new malicious software detection.

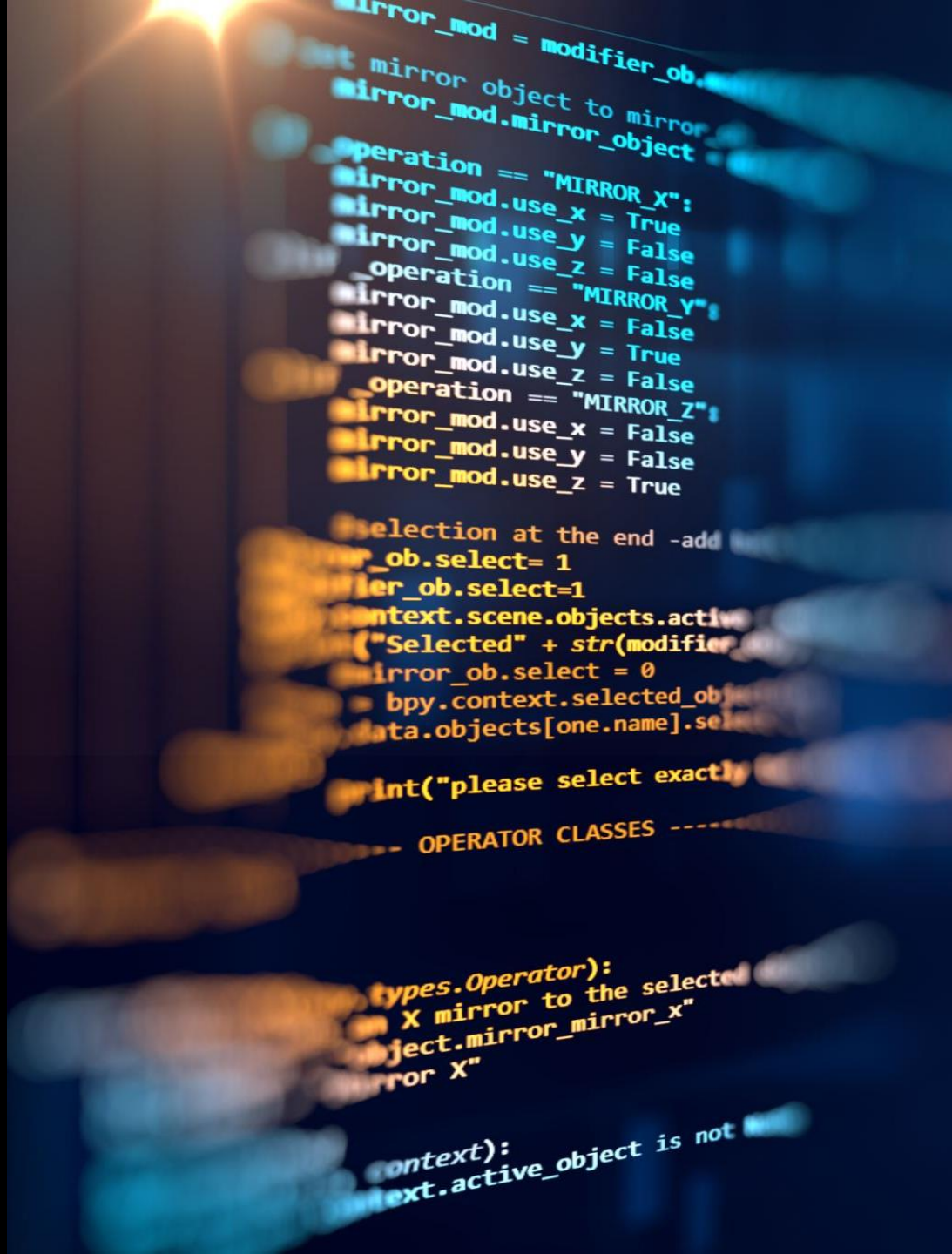
The data processing part extracts the characteristics of malicious software using various techniques, such as gay-scale records and import functions.

The decision-making part uses these features to classify malicious software and identify suspicious samples.

Finally, the detection module employs an algorithm to discover new families of malicious software

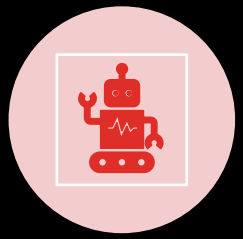
RELATED WORK

- Kediri et al. in 2019 focused on identifying malware in mobile applications using machine learning methods and reverse engineering of Android Java code.
- They found the algorithms that provide the highest malware detection rate and identified the application features with the highest utility in classifying malware.
- The results showed that two classification algorithms, Random Forest and K Nearest Neighbors, performed best in terms of correctly classifying instances of malware.
- Overall, these studies show the effectiveness of deep learning in detecting and classifying malicious software and can help enhance our digital security



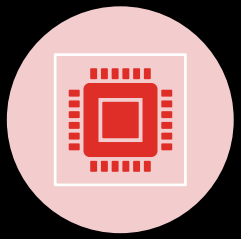
RELATED WORK

MALWARE ATTACKS ARE BECOMING MORE COMMON EVERY DAY, AND IT'S IMPORTANT TO HAVE EFFECTIVE SYSTEMS IN PLACE TO DETECT AND PREVENT THEM.



MACHINE LEARNING ALGORITHMS ARE BEING USED TO DEVELOP SUCH SYSTEMS, BUT THERE ARE SOME CHALLENGES.

FOR EXAMPLE, BIASED DATA CAN LIMIT THE PERFORMANCE OF THESE ALGORITHMS IN REAL-WORLD SCENARIOS.



SEVERAL RESEARCH STUDIES HAVE COMPARED TRADITIONAL MACHINE LEARNING ALGORITHMS WITH MORE ADVANCED DEEP LEARNING ALGORITHMS FOR MALWARE DETECTION, CLASSIFICATION, AND CATEGORIZATION USING VARIOUS DATASETS

PROBLEM STATEMENT

- Given the magnitude of losses and the frequency of attacks, accurate and timely detection methods are imperative.
- However, current static and dynamic methods do not provide efficient detection, particularly when dealing with zero-day attacks.
- The existing work provides the main points and concerns of learning-based malware detection, as well as explores the best feature representation and classification methods.
- It's crucial to understand the fundamentals of two malware analysis approaches - static and dynamic malware analysis.

LIMITATION OF EXISTING SYSTEM

- Less Performance
- Malware detection accuracy is less.

PROPOSED SYSTEM

- The proposed methods aim to find the Malware with higher standard. The accuracy levels of the identification of the Malware will be improved with the proposed system.

- The deep learning on neural network will provide the better solution to solve the problem of identification of the Malware in the real world intrusion data.

- The Convolutional Neural Network algorithm will check the data in more compact with training and testing the data. It will provide more accuracy as compared with the other type of techniques.

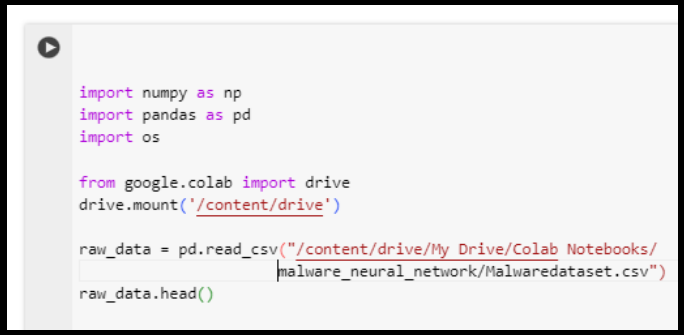
- The malware dataset will be taken as the input to the application and the dataset will be passed into the Convolutional Neural Network algorithm and the data will be analysed with the different visual graphs.

ADVANTAGES OF PROPOSED SYSTEM

- HIGH Performance
- Good Accuracy level.

EXPERIMENTAL RESULTS

- Exploratory examination is a cycle to investigate and comprehend the information and information relationship in a total profundity with the goal that it makes highlight designing and deep learning demonstrating steps smooth and smoothed out for expectation. .



```
import numpy as np
import pandas as pd
import os

from google.colab import drive
drive.mount('/content/drive')

raw_data = pd.read_csv("/content/drive/My Drive/Colab Notebooks/
                        malware_neural_network/Malwaredataset.csv")
raw_data.head()
```

EXPERIMENTAL RESULTS

- The Malware attributes contains 34 unique columns with each column denoting the attribute which identifies the malware in the network. The noisy data present inside the record is also removed and improves the record quality which will be more helpful in the application of the prediction of the Malware in the dataset.

```
[ ]  raw_data.columns

     Index(['hash', 'millisecond', 'classification', 'state', 'usage_counter',
            'prio', 'static_prio', 'normal_prio', 'policy', 'vm_pgoff',
            'vm_truncate_count', 'task_size', 'cached_hole_size', 'free_area_cache',
            'mm_users', 'map_count', 'hiwater_rss', 'total_vm', 'shared_vm',
            'exec_vm', 'reserved_vm', 'nr_ptes', 'end_data', 'last_interval',
            'nvcsw', 'nivcsw', 'min_flt', 'maj_flt', 'fs_excl_counter', 'lock',
            'utime', 'stime', 'gtime', 'cgtime', 'signal_nvcsw'],
           dtype='object')
```

EXPERIMENTAL RESULTS

- The classification part of the malware dataset is analysed to find out number of malwares types in the dataset.

```
[ ] data["classification"].value_counts()

    classification
    malware    50000
    benign     50000
    Name: count, dtype: int64

[ ] data['classification'] = data.classification.map({'benign':0, 'malware':1})
    data.head()
```

	hash	millisecond	classification
0	42fb5e2ec009a05ff5143227297074f1e9c6c3ebb9c914...	0	1
1	42fb5e2ec009a05ff5143227297074f1e9c6c3ebb9c914...	1	1
2	42fb5e2ec009a05ff5143227297074f1e9c6c3ebb9c914...	2	1
3	42fb5e2ec009a05ff5143227297074f1e9c6c3ebb9c914...	3	1
4	42fb5e2ec009a05ff5143227297074f1e9c6c3ebb9c914...	4	1

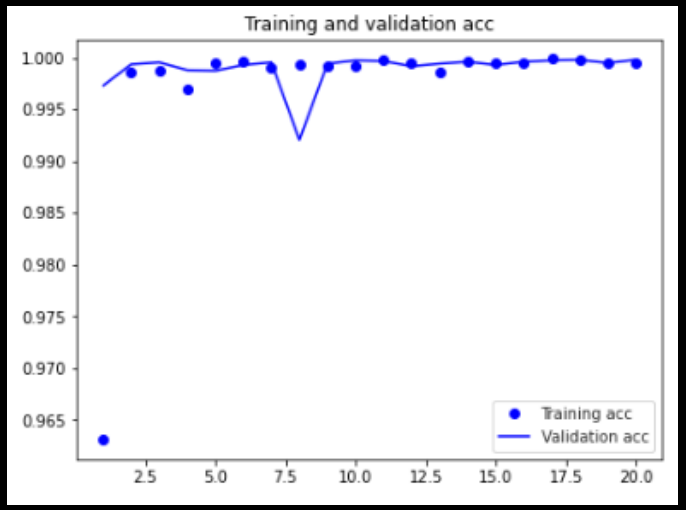
EXPERIMENTAL RESULTS

- The Convolutional Neural Networks algorithm is applied with creating the sequential model. The output of the sequential model with layers is displayed.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 50)                1400
_________________________________________________________________
dense_1 (Dense)              (None, 50)                2550
_________________________________________________________________
dense_2 (Dense)              (None, 50)                2550
_________________________________________________________________
dense_3 (Dense)              (None, 50)                2550
_________________________________________________________________
dense_4 (Dense)              (None, 50)                2550
_________________________________________________________________
dense_5 (Dense)              (None, 50)                2550
_________________________________________________________________
dense_6 (Dense)              (None, 2)                 102
=================================================================
Total params: 14,252
Trainable params: 14,252
Non-trainable params: 0
```

EXPERIMENTAL RESULTS

- The training and the validation accuracy graph shows the results with graphical format. The training accuracy is getting in the increase ratio and it reaches the good saturation point.



CONCLUSION

- A cutting-edge framework for detecting Malware has been developed using deep neural networks and diverse distributed network data.
- The framework employs all records with malware information for model training and data classification.
- By constructing functional intellectual networks based on the correlation, the neural network formation is optimized using correlation coefficient information.
- This methodology greatly enhances diagnostic accuracy compared to traditional approaches, demonstrating that integrating advanced deep learning.

REFERENCE

[1] Dong, G., Liao, G., Liu, H., Kiang, G., 2018. Are view of the auto-encoder and its variants: A comparative perspective from target recognition in synthetic caperture radar images. IEEE Geoscience and Remote Sensing Magazine 6, 44–68.
[2] Kedziora, M., Gawin, P., Szczepanik, M., & Jozwiak, I. (2019). Malware detection using machine learning algorithms and reverse engineering of android java code. International Journal of Network Security & Its Applications (IJNSA) Vol, 11.
[3] Alzaylaee, M. K., Yerima, S. Y., & Sezer, S. (2020). DL-Droid: Deep learning based android malware detection using real devices. Computers & Security, 89, 101663.
[4] Singh, J., & Singh, J. (2021). A survey on machine learning-based malware detection in executable files. Journal of Systems Architecture, 112, 101861.
[5] Usman, N., Usman, S., Khan, F., Jan, M. A., Sajid, A., Alazab, M., & Watters, P. (2021). Intelligent dynamic malware detection using machine learning in IP reputation for forensics data analytics. Future Generation Computer Systems, 118, 124-141.

REFERENCE

[6] Vasan, D., Alazab, M., Wassan, S., Naeem, H., Safaei, B., & Zheng, Q. (2020). IMCFN: Image-based malware classification using fine-tuned convolutional neural network architecture. Computer Networks, 171, 107138.

[7] Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., & Venkatraman, S. (2019). Robust intelligent malware detection using deep learning. IEEE Access, 7, 46717-46738.

[8] Wang, Z. J., Turko, R., Shaikh, O., Park, H., Das, N., Hohman, F., ... & Chau, D. H. P. (2020). CNN explainer: learning convolutional neural networks with interactive visualization. IEEE Transactions on Visualization and Computer Graphics, 27(2), 1396-1406.

[9] Soofi, A. A., & Awan, A. (2017). Classification techniques in machine learning: applications and issues. J. Basic Appl. Sci, 13, 459-465.

[10] Talan, T. ve Aktürk, C. (2021) Bilgisayar Biliminde teorik ve uygulamalı araştırmalar, Efe akademi yayıncılık, Istanbul, ISBN: 978-625-8065-42-8.

THANK YOU !!!