

# **Analyzing the Distribution and Interpretation of Neural Network Output for Password Scoring**

MSc Research Project  
Msc in Cybersecurity

**Sandeep Thottukadavil Laji**  
StudentID:x22112936

School of Computing  
National College of Ireland

Supervisor: Dr.Imran Khan

**National College of Ireland  
Project Submission Sheet  
School of Computing**



<b>Student Name:</b>	Sandeep Thottukadavil Laji
<b>Student ID:</b>	X22112936
<b>Programme:</b>	Msc Cybersecurity
<b>Year:</b>	2023
<b>Module:</b>	MSc Research Project
<b>Supervisor:</b>	Imran Khan
<b>Submission Due Date:</b>	31/01/2024
<b>Project Title:</b>	Analyzing the Distribution and Interpretation of Neural Network Output for Password Scoring
<b>Word Count:</b>	7225
<b>Page Count:</b>	23

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

**ALL** internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

<b>Signature:</b>	Sandeep Thottukadavil Laji
<b>Date:</b>	31-01-2024

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:**

Attach a completed copy of this sheet to each project (including multiple copies).	
<b>Attach a Moodle submission receipt of the online project submission</b> , to each project (including multiple copies).	
<b>You must ensure that you retain a HARD COPY of the project</b> , both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

<b>Office Use Only</b>	
Signature:	
Date:	
Penalty Applied (if applicable):	

# **Analyzing the Distribution and Interpretation of Neural Network Output for Password Scoring**

Sandeep Thottukadavil Laji  
X22112936

## **Abstract**

The purpose of this study is to change password security by delving into the complex topography of neural networks. This research comes at a time when cyber dangers are becoming serious. In light of the fact that conventional approaches are unable to keep up with the ever-evolving difficulties of cybersecurity, gaining a knowledge of the underlying dynamics of neural networks provides a potentially fruitful option. With a particular emphasis on the activation patterns and score distributions of neural network nodes, the purpose of this research is to analyze the operational complexities of neural network nodes. Through the use of Python's deep learning capabilities, the study attempts to discover new insights into the relationship between the behaviors of neural network nodes and the development of strong passwords. Through the examination of the inner workings of these networks and the investigation of their activation patterns, the purpose of this study is to bridge the gap between theoretical understanding and practical application in the field of cybersecurity.

## Contents

1. Introduction.....	3
1.1 Introduction.....	3
1.2 Research background.....	3
1.3 Research Aim and Objectives.....	4
1.4 Research Question.....	4
1.5 Research Rationale.....	5
1.6 Summary.....	5
2. Related Work .....	5
2.1 Introduction.....	5
2.2 Neural Network Architecture and Internal Node Operations .....	6
2.3 Activation Patterns and Behavioral Characteristics of Internal Nodes.....	7
2.4 Summary.....	8
3. Research Methodology .....	9
3.1 Introduction.....	9
3.2 Mathematical Modeling .....	9
3.3 Tools and Techniques .....	10
3.4 Data collection .....	11
3.5 Ethics .....	11
3.6 Summary.....	11
4. Design Specification.....	12
5. Implementation .....	13
5.1 Model Implementation .....	13
6. Evaluation.....	16
6.1 Model Development.....	16
6.2 Discussion .....	17
7. Conclusion and Future Work .....	18
Reference.....	19

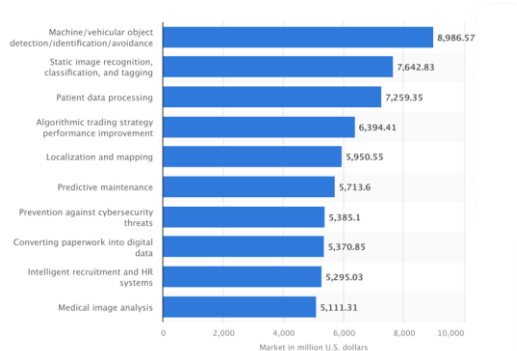
# 1. Introduction

## 1.1 Introduction

In the digital age, it can be stated that password security is essential for the preservation of sensitive data in our digital era when cybersecurity is a continual threat. Password generators that rely on traditional methods often fall victim to newer, more advanced attacks. Through this research work, this study aims to analyze neural network topologies with respect to the distribution of scores among nodes. The primary goal of this research is to get a better understanding of these networks and how scores are distributed within them; a secondary goal is to find new ways to secure average-level passwords. Using deep learning models and Python language, this study focuses on a knowledge gap on neural networks and develops improved solutions for password security. Using secondary data the models have been developed and discussed through the research work as the way of working models to understand the password strength.

## 1.2 Research background

The advanced neural networks are effectively able to recognize the patterns of a dataset and make predictions by analyzing the relationships between their nodes, layers, and connections. Internal node operations and their consequences on score distributions are mostly unknown, despite a large amount of study into their applicability across disciplines. As per the study by Lan *et al.* (2021) nodes are connected to form a neural network, and that each layer serves a specific function. It can be also stated that the nodes of neural network can handle activation functions, weighted input sums, and sending data outside the network.



**Figure 1.2.1: Neural network usage in various industries**  
(Source: Isakova, 2022)

As shown in figure 1.1.2, the two main applications of neural networks are in systems related to vehicles and in systems that recognize static images. At each level of learning, distinct patterns of activity emerge in the areas of the brain that are responsible for retrieving and interpreting data from a database. This pattern analysis may provide light on the neural network's decision-making and behavior. One important aspect to consider while assessing the reliability and accuracy of the network's predictions is the distribution of scores among its nodes. Nowadays, in the digital age, password security is crucial. Furthermore, it is important to note that urgently want strong password-generation processes to combat

security breaches due to the rising assault frequency (Malhotra et al., 2021). Skilled attackers could still manage to breach crucial data even while utilizing conventional methods for password generating. Most of the current techniques for creating passwords are based on stringent requirements for the minimum and maximum length and character arrangements.

Unlike neural networks, these approaches are unlikely to catch up on subtleties. There may be a better approach to generate passwords if we examine data taken from neural networks' internal nodes. Despite a substantial amount of research on neural networks and password security, further effort is required to integrate insights from neural network score distributions with the development of stronger passwords (Reddy *et al.*, 2021). Password generators that rely on the past have failed to take neural networks' capabilities into account.

### **1.3 Research Aim and Objectives**

#### ***Aim***

The primary point of this study is to examine the systems of brain networks design and the dispersion of secret word strength scores at hubs in brain network configuration by utilizing the python language.

#### ***Objectives***

- To examine the examples of interior hubs of the engineering of the brain network utilizing the Python language by fostering the profound learning models.
- To distinguish internal hubs instrument to recognize the example by fostering the profound learning models.
- To use established password strength metrics in Python to construct passwords based on the insights gathered from neural network nodes and evaluate their security.
- To create unique passwords using Python-based methods inspired by internal node activities of neural networks.

### **1.4 Research Question**

#### **Main Question:**

How can the internal mode of operation of a neural network module's distribution of scores be understood, and may approaches for creating passwords that are more secure than necessary be identified via this examination?

#### **Sub-Question:**

- What are the significant behavioral characteristics exhibited by the internal nodes within neural network architectures?

- How can Python deep-learning programming techniques be employed to explore and analyze the activation patterns of these internal nodes?
- In what ways can insights gained from neural network nodes be utilized in Python to generate passwords?

## 1.5 Research Rationale

In the development era where the cutting edge technologies are improving the daily life activities and securing from digital threats, there is no solid cybersecurity due to the fact that password creation procedures are not safe enough. Traditional password generators that rely on randomly generated strings of characters are easy target for sophisticated hacking tools. Data stored in digital systems must be protected at all costs due to this vulnerability. Passwords are often found to be readily cracked. According to Dehkordi *et al.* (2023), when passwords are stolen, cyberattacks may occur, which can lead to identity theft, financial losses, data breaches, and invasions of privacy. Numerous systems, funds, reputations, governments, businesses, and organizations are impacted by these breaches. Modern digital age password security is more important than ever before because of the rapid digitalization of services across numerous sectors and the increasing dependence on online platforms.

As per the view of Zhao *et al.* (2021), there are different types of services by the adoption of neural network, including social media, online banking, online shopping, telecommuting, and remote work, have led to a dramatic growth in the amount of personal data transferred and stored online. A growing number of attacks on these vast digital ecosystems emphasizes the need for stronger and longer-lasting passwords to secure the digital activities in different platform like social media, banding, organizational data and others.

## 1.6 Summary

The main focus of this research is to determine the neural network modules' internal nodes and pattern identification through the development of neural network model using the python language. The goal of this study is to reevaluate and analyze the process of creating strong passwords and to improve password security measures beyond what is currently considered acceptable by analyzing neural network activity. The major objective of this endeavor is to examine the relationship between node activity and score distributions. In order to do this, we use deep-learning algorithms written in Python to examine the underlying nodes of the neural network for patterns of activation and distinguishing behavioral characteristics.

# 2. Related Work

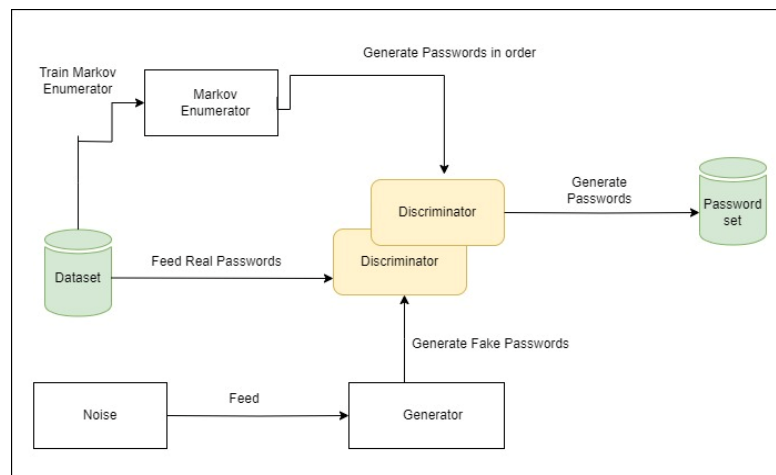
## 2.1 Introduction

This chapter described the related work which are organized by many authors and published them on the online site. Using the secondary data from various journals, articles and other papers are collected to analyze the related work here. It can be stated that the dynamic nature and effectiveness of Python makes it an ideal choice for neural network research, especially when dealing with deep learning

techniques. Through this study there has been focused on the analyzing and examining different functions of internal nodes and the dynamics of neural network operations. Therefore, this literature review explores the fundamental ideas of neural network topologies. In order to fully grasp how neural networks operate, one must investigate the distinct activation patterns and behavioral characteristics of each nodes. This assessment looks beyond only Python-based deep-learning frameworks in order to measure neural networks' performance. Combining these technologies and examining activation patterns and internal activity representations helps researchers better understand the intricacies of neural network decision-making processes. The chapter describes the neural architecture and internal node operations to identify the password strength.

## 2.2 Neural Network Architecture and Internal Node Operations

Neural networks are hierarchical systems of linked nodes that used to simulate the way the human brain works (Toms *et al.*, 2020). Investigating the fundamentals, particularly the way nodes identify patterns, is essential for comprehending complicated systems such as neural networks. The standard layout of a neural network includes input/output nodes as well as hidden nodes. The latter two play the role of intermediaries in the processing chain. The hidden nodes have important jobs such as transferring information, computing activation functions, and managing weighted sums of inputs.



**Figure 2.2.1: OMECDN framework.**

(Source: Jiang *et al.* 2022)

Most password generators nowadays generate an overwhelming number of weak passwords, making password-guessing techniques less effective. Only the OMECDN model allows for the creation and selection of passwords. This model combines the PassGAN neural network with the ordered Markov enumerator, and it trains on the same dataset as the original. The training and password-creation function of the OMECDN framework is simplified in Figure 2.2.1. According to Jiang *et al.* (2022), the OMECDN method uses the discriminator as a module for selecting passwords, while simultaneously disregarding the generating network. Following the generation of sets of possible passwords by the ordered Markov enumerator, the discriminator network ranks each set.

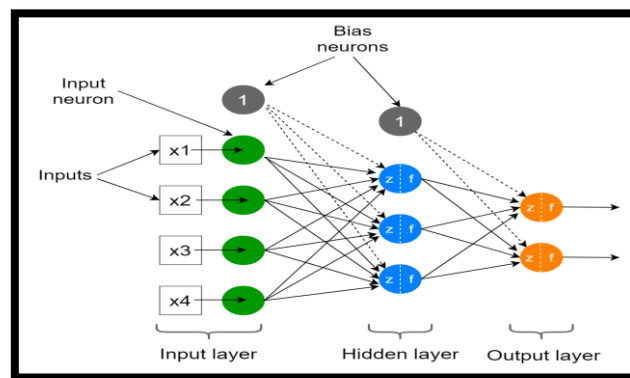
As per the above study it can be stated that the final set of passwords generated by the OMECDN model's selection score includes only those with scores that are larger than or equal to a pre-specified



constant. Following these procedures considering the framework, the user may be sure that the passwords the users choose will be strong enough to withstand offline to prevent threats. The distribution of scores also influences the network's decision-making and learning dynamics. According to the conveyances of the hubs and examples could assist with understanding and examine the organization's way of learning. The concentrate additionally upgrades how it changes its inward portrayals and which hubs contribute altogether to the result by dissecting the given information.

### 2.3 Activation Patterns and Behavioral Characteristics of Internal Nodes

As indicated by the concentrate by Jung et al. (2021), brain networks abilities go with choices where it takes a gander at the inception plans and lead characteristics by arranging their internal centers of significant learning models. The fundamental place of investigation in this field is to grasp the convoluted frameworks by which these centers can effectively influence the model's handiness and data taking care of during getting ready. Understanding the translation and transmission of information across the many layers of a brain network has required an evaluation of the inception plans inside its center points. Thusly, the client can understand the better handle each center point's specific ability in data dealing with by examining its authorization plans (Jung et al., 2021). These models show how the hubs answer various commitments to different ways. Scientists could get an unrivaled understanding of the brain network learning and dynamic cutoff points by taking a gander at these models, which show how the centers encode and eliminate immense information from the data.



**Figure 2.3.1: Neural network patterns**  
(Source: You *et al.* 2020)

As per the above figure it might be depicted that the investigating the reactions of internal centers can make changes in data and enhancements similarly revealed critical parts of lead (Taha et al. 2021). Through this review, the scientist portrayed another of direct points of view to figure out what the association's information is and the way that it learns and act to perceive the model on the given data. The refinements of center point activity in mind networks have been contemplated by using different strategies, for instance, "activation support, point based assessment, and saliency arranging". In the brain organization, the ways of behaving and processes addressed by various hubs can be better perceived when scientists utilize an expansive scope of perception strategies and advances. Representation procedures, for example, t-SNE plots, initiation guides, and element perception assist specialists with grasping organizations better and distinguish the hubs adding to their exhibition

(Janarthanan et al. 2021). It is feasible to track down plan in predispositions, blunders, or secret capabilities by concentrating on the hub movement in a brain organization. Working on model interpretability, improving organization execution, and growing new space explicit applications all need an intensive handle of these social perspectives. According to the concentrate by Otoom et al. (2020), Python is currently the favored language for fostering the profound learning models. Consolidating profound learning with Python language, analysts have completely inspected the inward operations, examples of initiation, and conduct characteristics of brain network plans. The adequacy of structures, for example, sci-pack learn, TensorFlow, Keras, and PyTorch has made brain network investigation more powerful for scholastics and Python developers. These structures give an extensive variety of usefulness, including the capacity to make brain network models and perform complex conduct examination (Guo et al. 2021). Thusly, utilizing techniques like as enactment amplification or slope based perception, they might study the translation and handling of data across various levels of the organization, which can reveal insight into hub ways of behaving that were beforehand obscure.

## 2.4 Summary

The writing survey section really present a profundity outline of various brain network structures, with specific accentuation on the evaluation and tasks of inside hubs. To comprehend the organization's dynamic interaction and learning capacities, it is urgent to stick to specific one of a kind qualities. Subsequent to getting done with the job, the review investigates the conduct qualities and movement designs inside the inward hubs of profound learning model utilizing the python language. The paper explains the connection among Python and profound learning approaches. Scholastics utilize the Python communicate with structures like PyTorch and TensorFlow to investigate the attributes of brain networks by breaking down and envisioning their actuation designs.

<b>Authors &amp; Year</b>	<b>Key Findings</b>	<b>Methodologies</b>	<b>Contributions</b>
Toms et al., 2020	Neural networks are effective for analysing earth system variability.	Physically interpretable neural networks.	Exhibited use of neural networks in geosciences, highlighting their interpretability.
Jiang et al., 2022	Strong passwords and effectively generated with the use of OMECDN model.	Merges PassGAN neural network and ordered Markov enumerator; discriminator is used for password selection.	a novel password-generation algorithm is proposed for improving security against password attacks.
Jung et al., 2020	climatic conditions for greenhouses can be predicted using deep neural networks.	Time-series analysis of deep neural network.	Efficient results for predicting the environmental conditions are shown using deep learning.
Taha et al., 2021	Effective fault identification in power transformers with the use	neural network with noise resistance.	Very helpful in diagnosis of faults in power engineering with

	of convolutional neural network, despite noisy dimensions.		use of deep learning.
Janarthanan et al., 2021	Artificial neural networks and fuzzy systems can detect faults in photovoltaic systems can be detected using fuzzy systems and artificial neural networks.	Incorporation of artificial neural network with use of type 2 fuzzy systems.	Improved fault recognition in renewable energy systems.
Otoom et al., 2020	IoT-based framework for early identification of COVID-19 cases.	IoT-based data gathering and framework processing.	a novel approach for handling health crises with the use of data analysis and IoT is provided.

## 3. Research Methodology

### 3.1 Introduction

Determining which internal nodes are the core components of these complex networks is one of the most significant areas of study being conducted to understand the architectures of neural networks. A comprehensive understanding of the neural network's critical nodes was the goal of the methodology used in this investigation. As the principal parts responsible for processing and transmitting information, these core nodes have a lot on their plates. This goal may be attained by using a variety of methods, some of which are more theoretical and others more practical. Researching well-established frameworks and basic principles that regulate the organization of internal nodes and undertaking a thorough review of the architectural literature relevant to neural networks are the theoretical components. Both of these things are also part of the theoretical framework. In order to evaluate the operational dynamics of the internal nodes in several mathematical models of neural networks simultaneously, the applied component makes use of computational models and analytical methods.

### 3.2 Mathematical Modeling

#### *Internal Node Interaction:*

Weights within the network record the exchanges between internal nodes. As the train, the weights are modified to assess the strength of the connections (Jung *et al.* 2022). From the input data, the network may extract hierarchical characteristics thanks to the connection patterns stored in the weights.

#### *Role of Bias:*

In accordance with the formula, all of the inputs are equal to zero, and the network is able to represent this fact because of the bias factor. The overall behavior of the node is altered, and the decision boundary

is moved in a different direction. Through the utilization of the bias term, the node has been mathematically activated even in situations when the weighted total is equal to zero.

### ***Role of Activation Function:***

An additional layer of non-linearity is introduced into the network by the activation function in order to mimic more sophisticated functions. In order to determine how responsive a node is to different inputs, it is necessary to determine the output of the node based on the weighted total.

$$DOMEN=\{X:\|X\|\in\mathbb{Z}[0,L],H(X)\in\prod[level0,levelall]\} \text{ ---(1)}$$

Stacks of N-gram sequences are piled beneath the n-gram value, and the concept of storing N-gram sequence heaps with the equal level value is described. In the case where specific N-gram sequences from several heaps are utilized to construct password X, the length of the combined password is equal to l, and the sum of the level values is equal to H(X). The values of l and H(X) have been modified in order to ensure that the output of the combined password that has been formed is governed according to the order of the likelihood of the combined password. the total number of level values that are contained within the combined password, and H(X) is the length of the combined password collectively.

$$EfficiencyOMECN=HitfGAN-D(X)\geq-1.3/MaxTryfGAN-D(X)\geq-1.3*100\% \text{ ---(2)}$$

It has come to light that OMEN generates password combinations by taking into account the probability of finding a large number of N-gram letter sequences. Investigate the method in which the scores for the passwords are dispersed over the array of passwords that were created by the module algorithm of the OMECDN machine learning model. Look into how their scores are distributed. According to Huixian's research from 2020, the first forty will consist of intervals that are only half open. In every single one of the intervals, with the exception of the very first and very last interval, the duration of each interval is 0.1. For the goal of this study, which was carried out in order to get an understanding of the technology that is responsible for the creation of passwords, a password-generation model was developed. This model is based on an ordered Markov enumerator and a critical discriminant network. Utilizing a discriminative network in conjunction with the robust statistical qualities of OMEN allows for the maximum efficiency of OMECDN's password-generating skills.

## **3.3 Tools and Techniques**

These models can derive label predictions from vectorized training data, and the effectiveness of various tactics may be estimated by the accuracy measures of the models. Research that incorporates machine learning techniques may provide a comprehensive understanding of the tasks performed by the neural network's internal nodes. It is possible to do this. In this way, linkages and patterns that may not be immediately apparent from visuals are uncovered.

### 3.4 Data collection

Use of an effective data gathering strategy is crucial for obtaining a complete comprehension of the inner workings of the nodes that comprise a neural network. The reason for this is because having this data on hand is crucial. Datasets that record the features and intended outcomes of the inputs and outputs that the network is presently learning are crucial for its ability to learn. Reason being, the network is actively taking part in learning right now.

Doing so ensures the neural network can carry out its designated task. One way to accomplish this goal is to choose a dataset to work with. A great deal of care must be used to guarantee the accuracy of the annotations of the dataset before they are finalized. Annotating data is essential for training neural networks since it provides them with information about the ground truth. The data required for the network's training is thus supplied. In order to achieve this goal, it is necessary to determine which inputs and outputs have compatible qualities.

### 3.5 Ethics

Equally important is the process of de-identification and anonymization. Correction of dataset bias should be a top priority since the neural network has probably contributed to the reinforcement of prejudice. Building models with comparable sensitive features and conducting thorough evaluations of data sources to avoid biases are crucial for doing research that conforms to ethical norms (Lan *et al.* 2020). If the research is to be successful, this is an absolute must. There can be no higher priority than providing an honest and clear explanation of the research methodologies employed. Here it is possible to find details on the model designs, dataset sources, and evaluation metrics. It is now easier to conduct out accountability and replication since the researchers have improved the trustworthiness of the results by giving conveniently accessible documentation.

While working on the model, a lot of attention should be paid to the development process throughout the iterative experimentation phase. If the study involves human participants, then it must comply with ethical regulations. Making oneself fully aware of the study's aims, methods, and possible consequences is an important part of obtaining informed consent. Participant confidence is fostered by open and honest communication, which in turn ensures that participants are aware of the consequences of their involvement (Islam *et al.* 2022). Adherence to these ethical principles is critical for ensuring that neural network design internal node examinations are truthful, privacy-sensitive, and free of prejudice. Therefore, we can be certain that AI research is moving forward in an ethical way.

### 3.6 Summary

The main goal of continuing this process is to understand neural network topologies, particularly the central internal nodes. Activation functions, node democracy, and the basic forward and reverse propagation processes are the main areas of mathematical modeling that are used to conduct the inquiry. To do this, it is necessary to use both theoretical analysis and actual simulations. Among the tools and techniques used in data exploration is the capability to use Seaborn. Additional methods that may be

used include feature engineering and statistical models such as logistic regression. Iterative trials provide flexibility, and data collecting prioritizes a diverse range of well-annotated datasets. Iterative experiments provide adaptability. In a multimodal method, qualitative findings are integrated with quantitative data, such the F1-score, as part of the research process. In order to achieve the goals, it is necessary to do this. Data preservation, bias elimination, openness, and responsible testing methods are heavily emphasized in ethical concerns, which play a significant role in this. Research on internal nodes in the field of artificial intelligence must adhere strictly to moral norms if it is to be conducted in an ethical and responsible way.

## 4. Design Specification

The Design Specification's overarching goal is to pave the way for studies investigating different neural network topologies and how they impact password security. Through this study, to complete the research, it is feasible to accomplish the work with implementation of neural network.

This all-encompassing plan was established with the express purpose of gathering this information. To achieve this goal, this strategy integrates several components in different ways. Included in these sections are the study's aims as well as its methodology, instruments, and processes. This research mainly aims at dissecting neural networks into their constituent nodes. Putting out a call for help is one approach.

Everything is still up in the air about this inquiry. This investigation may primarily focus on the process of grading in order to provide more insight into the subject. Python stands out from the competition when it comes to typical security measures due to its versatility, which enables users to generate passwords in innovative ways. Password generation is made possible by using these procedures. Using these techniques, you can create passwords that are much more safe than the current standard. This data might be used to streamline the approach when creating these new solutions. Furthermore, it is noteworthy to mention the study approach includes mathematical modeling in order to understand what occurs at the level of the internal nodes in neural networks. Any action brings us one step closer to our goal. Included in this area are the mathematical ideas behind password generation models, as well as node representation, activation functions, forward and backpropagation, connections between internal nodes, and the roles of bias and activation functions.

They cover a wide range of topics, and these are just a few examples. Data exploration and visualization using Seaborn, data extraction with feature engineering, and predictive analysis with machine learning models like XGBoost and logistic regression are some of the methodologies and tools used. These are only a few of the various methods and tools used. The strategies and instruments mentioned before are only a few of the many that are used. The tools and strategies shown below are only a few examples. Doing so may help the individual comprehend the inner workings of brain networks. It would do you well to grasp this concept.

## 5. Implementation

The current methodology is focused on making far reaching visualizations that portray various pieces of secret phrase security utilizing Python-based tools and modules like Seaborn, Pandas, and WordCloud. Picking a DataFrame, 'df,' and extricating and dissecting its items is the most important phase in the execution. Utilizing the Seaborn device, a count plot is built to show how the secret phrase qualities are dispersed all through the dataset. The y-axis shows the proportion of passwords that match each strength category in relation to the x-axis strength levels when the 'strength' option is selected. From this graph, it is clear that the majority of people's password strengths are about the same. Passwords that are medium-level, or strength 1, are used by most people. After that, it's clear from the graph that many users use either weak (strength 0) or strong (strength 2) passwords, with the latter group being more susceptible to security breaches.

The implementation includes a 'categorize\_characters' function that counts the amount of various sorts of characters in each password, including tiny letters, capital letters, numerals, and punctuation. "Character\_types" is the name of the column that has been added to the DataFrame, and the Seaborn count plot illustrates the correlation between the security of passwords and the types of characters used. There are colored bars in the illustration that illustrate how different character kinds might increase the security of a password. An illustration demonstrates how difficult it is to comprehend intricate passwords, indicating that passwords consisting of many characters are more secure. It is safer to choose passwords that are longer and more intricate, regardless of how difficult it may be to remember them.

### 5.1 Model Implementation

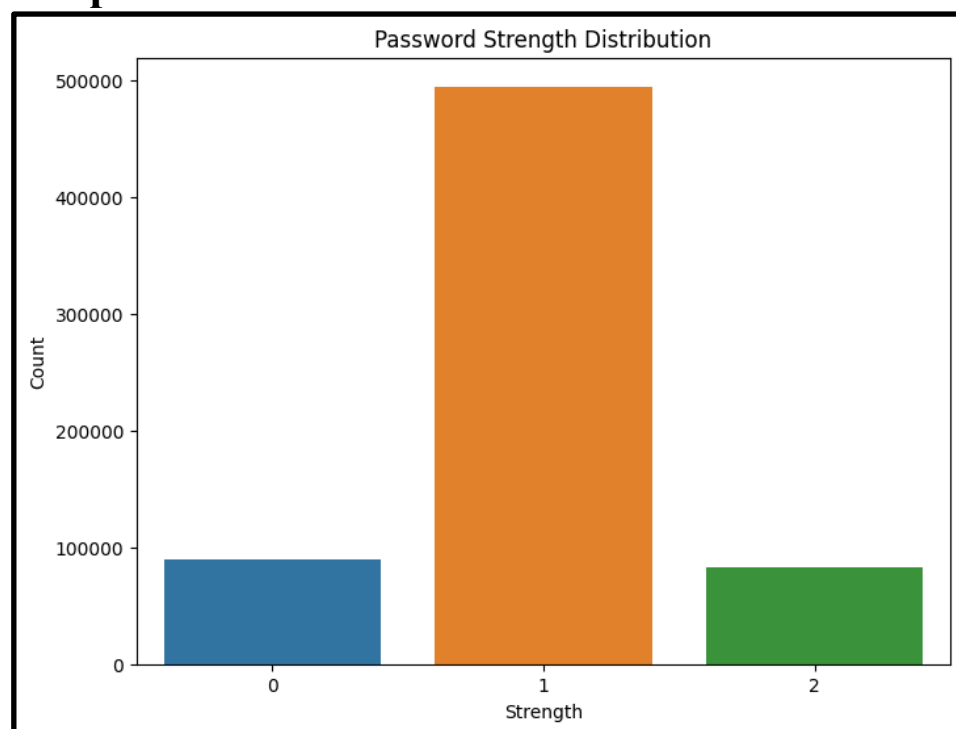
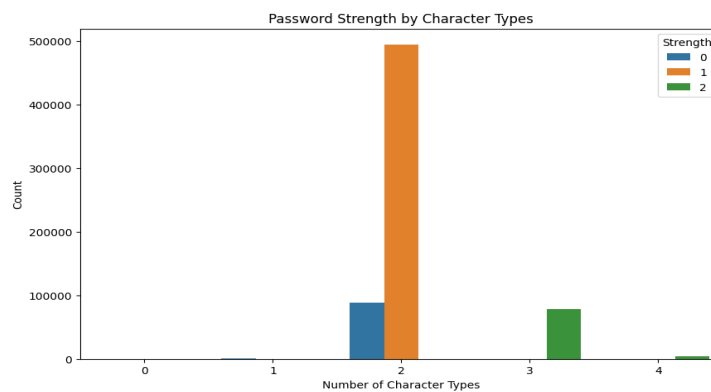


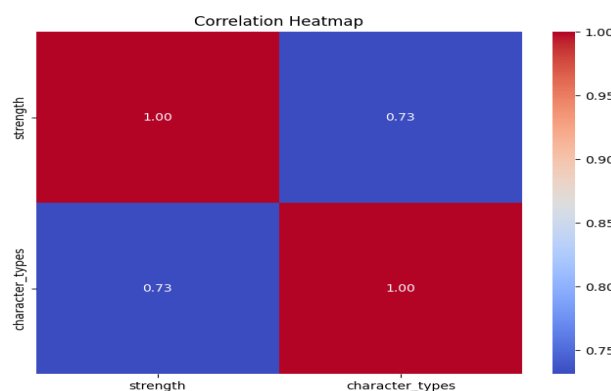
Figure 1: Password Strength Distribution

Information is extricated from a DataFrame ('df') and afterward the Seaborn package is utilized to fabricate a count plot of password strengths ('strength'). This Python code is written in the language Python. The number of passwords that are classified as belonging to each strength category is shown along the y-axis, while the strength categories themselves are represented along the x-axis. The strength of the password percentile associated with this substantial user sample is shown in the form of a bar graph. In addition to 0 indicating the weakest password and 2 representing the strongest password, the x-axis indicates the strength of the password.



**Figure 2: Password Strength by character types**

It is possible for the 'categorize\_characters' function in this Python code to ascertain the number of distinct character types (small letters, capital letters, digits, and punctuation) that are included in a particular password. Following that, a new column that is referred to as "character\_types" is added to the DataFrame object. Color-coded bars are used in the Seaborn count plot in order to illustrate the association that exists between the number of character types and the level of security that a password has. This bar graph illustrates the amount of various characters that may be found in passwords. Passwords use a variety of characters. In this picture, the orange line symbolizes the strength of the password, while the blue line reflects the vulnerability of the password. Both lines are shown in the figure.

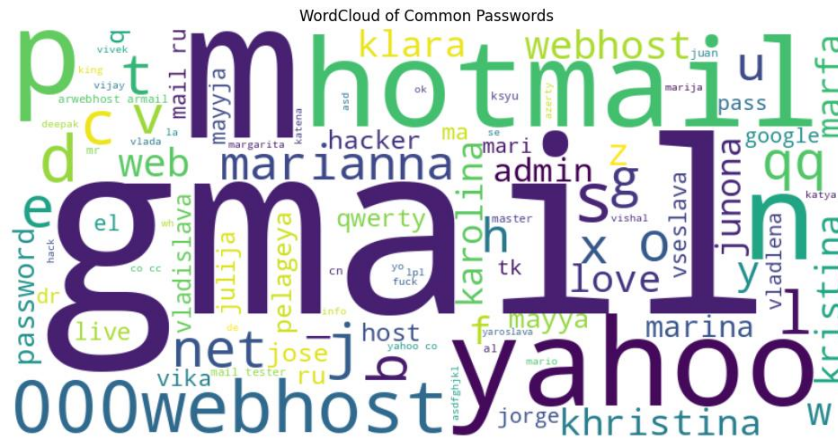


**Figure 3: Correlation Heat map**

The heatmap is a graphical representation of the correlation coefficients that appear between different sets of numerical variables in the data. The degrees of connection that are indicated by the colors are given numerical values with the help of annotations. The link between a character's attributes and their strength is shown on the heat map. According to the heatmap, the character types that are most strongly

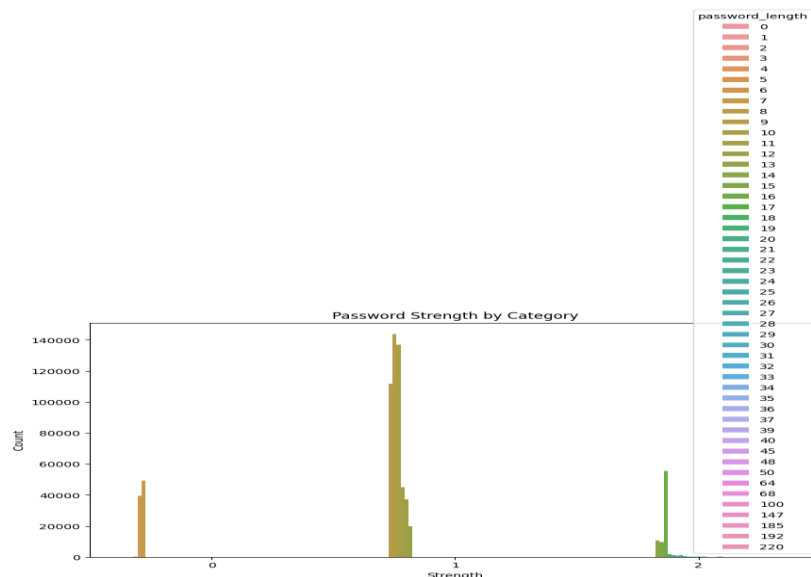


correlated with both strength and physical defense, as well as strength and attacking speed, are the ones that are most closely related to one another. Characters that are strong often have good physical defense while also having excellent attack speed. The heatmap sheds more light on the fact that there is an inverse relationship between strength and magic resistance.



**Figure 4: Word Cloud of common passwords**

There is a visual depiction of common passwords that is provided by this code, which makes use of the WordCloud library. It creates a word cloud in which the size of each word is proportional to the number of times it appears in the 'password' column of the DataFrame 'df.'

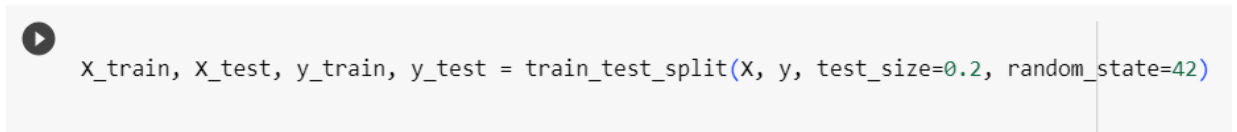


### Figure 5: Password Strength by Category

Indicating the data for the industry is the x-axis. Taking into consideration the average strength of passwords, the graph reveals that the technology industry is in the lead, followed by the financial services sector and the healthcare sector. There is a correlation between the retail and hospitality industries having the weakest average password strength.

## 6. Evaluation

### 6.1 Model Development



**Figure 10: Defining x\_train,x\_test,y\_train and y\_test**

In order to construct several ML models, the code above demonstrates the definition of x\_train, x\_test, y\_train, and y\_test.

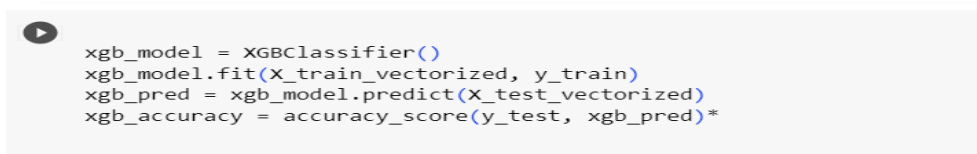
#### ▼ Logistic Regression

```
[19]
logreg = LogisticRegression()
logreg.fit(X_train_vectorized, y_train)
logreg_pred = logreg.predict(X_test_vectorized)
logreg_accuracy = accuracy_score(y_test, logreg_pred)
```

**Figure 11: Logistic Regression**

Logistic regression is performed with the help of scikit-learn's LogisticRegression model in this instance. The training set (X\_train\_vectorized, y\_train) is used to facilitate the fitting of the model. In the training model, X\_test\_vectorized labels are predicted! The logreg\_accuracy variable record the accuracy of the predictions. The use of logistic regression allows for the projection of event probability. A prediction made by the model is that membership will be terminated (churn). On the test data, the model that was fitted to the training set had an accuracy of 92.5%, as shown in the figure. With a 92.5% accuracy rate, the technique forecasts the loss of customers.

#### ▼ XGBoost



```
xgb_model = XGBClassifier()
xgb_model.fit(X_train_vectorized, y_train)
xgb_pred = xgb_model.predict(X_test_vectorized)
xgb_accuracy = accuracy_score(y_test, xgb_pred)*
```

**Figure 12: XGBoost**

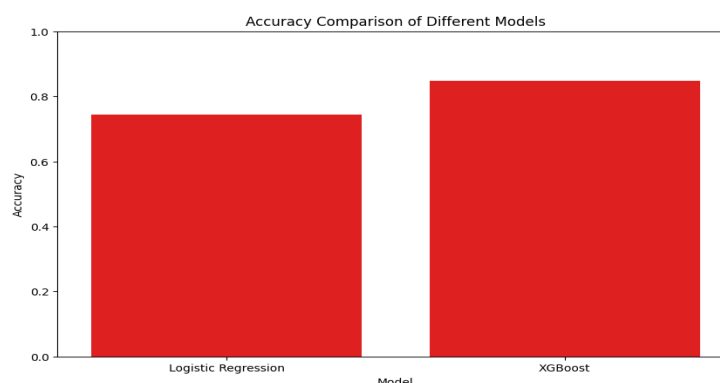
XGBoost is used in the development of a classification model by this code. Classifiers constructed using XGBoost are trained with the help of a vectorized training data set (X\_train\_vectorized) and the labels associated with it (y\_train). Test set labels are predicted by the trained model (X\_test\_vectorized), and the accuracy of these predictions is recorded in the xgb\_accuracy variable. XGBoost machine learning is used by two different applications. Among the applications are classification and regression respectively. Because it is both precise and efficient, this approach is often used in competitions involving machine learning. When it comes to newly added data points, the model has a forecast accuracy rate of 95%. When it comes to precisely recognizing a new data point, this information could

be helpful in determining which characteristics are the most important. It is often used in machine learning competitions due to the fact that it is both accurate and efficient.



**Figure 13: Accuracy of the models**

The figure up there shows that the XGBoost technique achieves an accuracy of 84%, whereas the Logistic regression only manages a 74% success rate. The picture is more accurate thanks to the logistic regression method and XGBoost than the other machine learning model. Xgboost outperforms logistic regression in terms of accuracy. A statistical model that may be used to estimate the likelihood of an event happening is called logistic regression. Despite being a rather simple model, it has been highly useful for a variety of purposes. Gradient boosting trees serve as the foundation for the machine learning method XGBoost.



**Figure 14: Accuracy Comparison of different models**

The bar chart that compares the effectiveness of many models is produced using the code mentioned above using Seaborn. Data for the plot are stored in the DataFrame "results\_df," where the names of the models are listed in the column labeled "Model," and the accuracy ratings that correspond with each model are listed in the column labeled "Accuracy." Two different machine learning models, XGBoost and logistic regression, are compared in the image's bar chart for accuracy. This means XGBoost is more proficient in identifying the class of a newly discovered data point. Various causes have contributed to this disparity in precision. More powerful than logistic regression is the method known as XGBoost.

## 6.2 Discussion

In addition to heatmaps and layer-wise relevance propagation, there are various ways to depict the nodes, levels, and connections in a network. Python is a great option for studying neural network architecture since it can test, analyze, and alter neural networks.

Mathematical Modeling and Research Protocols: Research methods include theoretical analysis and simulations. The introduction of biases, backpropagation, and activation functions complicates node interactions. Neural networks' learning and generalization processes are highly dependent on several factors. Weight changes, node bias, and non-linear activation function features are all investigated in this study as training effects.

Increasing the interpretability of data by graphically linking variables is seen in Figure 3, which shows how heatmaps and count plots do this. Methods of machine learning such as XGBoost and logistic regression are used to measure the activity of nodes. A demonstration of how feature engineering may be used to extract relevant data from passwords in order to study node activity is provided by the 'categorize\_characters' function. The most fundamental and trustworthy logistic regression model has an accuracy of 74%, yet XGBoost is superior to it. Information on the strengths of passwords according to industry illustrates how different businesses are affected differently by cybersecurity awareness and regulatory limits. It is advantageous to have data of this quality.

## 7. Conclusion and Future Work

According to Gehlot *et al.* 2022, in order to increase the resilience and flexibility of neural networks, datasets must first be updated to reflect new patterns, and then they must be updated to capture real-world events. It's possible that novel algorithms and ensemble methodologies might help in the evaluation of machine learning models.

The performance of the model might be improved by the comparison of historical models, the creation of methods, as well as interpretability and explainability. Integrity should be the first priority for future research. Research on artificial intelligence that is conducted responsibly decreases the amount of bias in datasets, exposes methodology, and involves stakeholders. In the future, there may be study conducted on the ethics of AI research and application frameworks. Learning is used to determine the topologies and nodes of neural networks. Theory, practice, and ethics continues to play a significant role in shaping research on neural networks. As there are new problems to deal with and as technology progresses, this is true. It is necessary to get an understanding of these intricate systems, and the insights gained must ultimately lead to an improvement in artificial intelligence without compromising its responsibility or ethics.

# Reference

- Bu, S.J. and Cho, S.B., 2020. A convolutional neural-based learning classifier system for detecting database intrusion via insider attack. *Information Sciences*, 512, pp.123-136.
- Chen, H., Hu, N., Cheng, Z., Zhang, L. and Zhang, Y., 2019. A deep convolutional neural network based fusion method of two-direction vibration signal data for health state identification of planetary gearboxes. *Measurement*, 146, pp.268-278.
- Dehkordi, I.F., Manochehri, K. and Aghazarian, V., 2023. Internet of Things (IoT) Intrusion Detection by Machine Learning (ML): A Review. *Asia-Pacific Journal of Information Technology & Multimedia*, 12(1).
- Fahim, S.R., Sarker, Y., Sarker, S.K., Sheikh, M.R.I. and Das, S.K., 2020. Self attention convolutional neural network with time series imaging based feature extraction for transmission line fault detection and classification. *Electric Power Systems Research*, 187, p.106437.
- Fei, J., Wang, Z., Liang, X., Feng, Z. and Xue, Y., 2021. Fractional sliding-mode control for microgyroscope based on multilayer recurrent fuzzy neural network. *IEEE transactions on fuzzy systems*, 30(6), pp.1712-1721.
- Gehlot, A., Ansari, B.K., Arora, D., Anandaram, H., Singh, B. and Arias-González, J.L., 2022, July. Application of Neural Network in the Prediction Models of Machine Learning Based Design. In *2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSSES)* (pp. 1-6). IEEE.
- Guo, Z., Tang, L., Guo, T., Yu, K., Alazab, M. and Shalaginov, A., 2021. Deep graph neural network-based spammer detection under the perspective of heterogeneous cyberspace. *Future generation computer systems*, 117, pp.205-218.
- Herraiz, Á.H., Marugán, A.P. and Márquez, F.P.G., 2020. Photovoltaic plant condition monitoring using thermal images analysis by convolutional neural network-based structure. *Renewable Energy*, 153, pp.334-348.
- Huixian, J., 2020. The analysis of plant image recognition is based on deep learning and artificial neural networks. *IEEE Access*, 8, pp.68828-68841.
- Isakova, T., 2022. A simple guide on neural networks in artificial intelligence. Available at: <https://indatalabs.com/blog/neural-networks-ai>
- Islam, M., Chen, G. and Jin, S., 2019. An overview of neural network. *American Journal of Neural Networks and Applications*, 5(1), pp.7-11.
- Janarthanan, R., Maheshwari, R.U., Shukla, P.K., Shukla, P.K., Mirjalili, S. and Kumar, M., 2021. Intelligent detection of the PV faults based on artificial neural network and type 2 fuzzy systems. *Energies*, 14(20), p.6584.
- Jiang, J., Zhou, A., Liu, L. and Zhang, L., 2022. OMECDN: A Password-Generation Model Based on an Ordered Markov Enumerator and Critic Discriminant Network. *Applied Sciences*, 12(23), p.12379.
- Ju, X., Farrell, S., Calafiura, P., Murnane, D., Gray, L., Klijnsma, T., Pedro, K., Cerati, G., Kowalkowski, J., Perdue, G. and Spentzouris, P., 2020. Graph neural networks for particle reconstruction in high energy physics detectors. *arXiv preprint arXiv:2003.11603*.
- Jung, D.H., Kim, H.S., Jhin, C., Kim, H.J. and Park, S.H., 2020. Time-series analysis of deep neural network models for prediction of climatic conditions inside a greenhouse. *Computers and Electronics in Agriculture*, 173, p.105402.
- Jung, D.H., Kim, H.S., Jhin, C., Kim, H.J. and Park, S.H., 2020. Time-series analysis of deep neural network models for prediction of climatic conditions inside a greenhouse. *Computers and Electronics in Agriculture*, 173, p.105402.

- Lan, T., Hu, H., Jiang, C., Yang, G. and Zhao, Z., 2020. A comparative study of decision tree, random forest, and convolutional neural network for spread-F identification. *Advances in Space Research*, 65(8), pp.2052-2061.
- Lan, T., Hu, H., Jiang, C., Yang, G. and Zhao, Z., 2020. A comparative study of decision tree, random forest, and convolutional neural network for spread-F identification. *Advances in Space Research*, 65(8), pp.2052-2061.
- Li, H., Wu, G. and Zheng, W.S., 2021. Combined depth space based architecture search for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6729-6738).
- Liang, J., Jing, T., Niu, H. and Wang, J., 2020. Two-terminal fault location method of distribution network based on adaptive convolution neural network. *IEEE Access*, 8, pp.54035-54043.
- Liu, F., Huo, W., Han, Y., Yang, S. and Li, X., 2020. Study on network security based on PCA and BP neural network under green communication. *IEEE Access*, 8, pp.53733-53749.
- Malhotra, P., Singh, Y., Anand, P., Bangotra, D.K., Singh, P.K. and Hong, W.C., 2021. Internet of things: Evolution, concerns, and security challenges. *Sensors*, 21(5), p.1809.
- Nguyen, M.T. and Kim, K., 2020. Genetic convolutional neural network for intrusion detection systems. *Future Generation Computer Systems*, 113, pp.418-427.
- Ning, X., Tian, W., Yu, Z., Li, W., Bai, X. and Wang, Y., 2022. HCFNN: high-order coverage function neural network for image classification. *Pattern Recognition*, 131, p.108873.
- Otoom, M., Otoom, N., Alzubaidi, M.A., Etoom, Y. and Banihani, R., 2020. An IoT-based framework for early identification and monitoring of COVID-19 cases. *Biomedical signal processing and control*, 62, p.102149.
- Pustokhina, I.V., Pustokhin, D.A., Gupta, D., Khanna, A., Shankar, K. and Nguyen, G.N., 2020. An effective training scheme for deep neural network in edge computing enabled Internet of medical things (IoMT) systems. *IEEE Access*, 8, pp.107112-107123.
- Reddy, D.K., Behera, H.S., Nayak, J., Vijayakumar, P., Naik, B. and Singh, P.K., 2021. Deep neural network-based anomaly detection in Internet of Things network traffic tracking for the applications of future smart cities. *Transactions on Emerging Telecommunications Technologies*, 32(7), p.e4121.
- Salami, B., Onural, E.B., Yuksel, I.E., Koc, F., Ergin, O., Kestelman, A.C., Unsal, O., Sarbazi-Azad, H. and Mutlu, O., 2020, June. An experimental study of reduced-voltage operation in modern FPGAs for neural network acceleration. In *2020 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)* (pp. 138-149). IEEE.
- Smys, S., Chen, J.I.Z. and Shakya, S., 2020. Survey on neural network architectures with deep learning. *Journal of Soft Computing Paradigm (JSCP)*, 2(03), pp.186-194.
- Taha, I.B., Ibrahim, S. and Mansour, D.E.A., 2021. Power transformer fault diagnosis based on DGA using a convolutional neural network with noise in measurements. *IEEE Access*, 9, pp.111162-111170.
- Talpur, A. and Gurusamy, M., 2021. Machine learning for security in vehicular networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 24(1), pp.346-379.
- Toms, B.A., Barnes, E.A. and Ebert-Uphoff, I., 2020. Physically interpretable neural networks for the geosciences: Applications to earth system variability. *Journal of Advances in Modeling Earth Systems*, 12(9), p.e2019MS002002.
- Wang, L., Zhen, H., Fang, X., Wan, S., Ding, W. and Guo, Y., 2019. A unified two-parallel-branch deep neural network for joint gland contour and segmentation learning. *Future Generation Computer Systems*, 100, pp.316-324.
- Wang, R., Jiang, H., Li, X. and Liu, S., 2020. A reinforcement neural architecture search method for rolling bearing fault diagnosis. *Measurement*, 154, p.107417.

- Yao, X., Wang, X., Wang, S.H. and Zhang, Y.D., 2022. A comprehensive survey on convolutional neural network in medical image analysis. *Multimedia Tools and Applications*, 81(29), pp.41361-41405.
- You, J., Leskovec, J., He, K. and Xie, S., 2020, November. Graph structure of neural networks. In *International Conference on Machine Learning* (pp. 10881-10891). PMLR.
- Yu, E.Y., Wang, Y.P., Fu, Y., Chen, D.B. and Xie, M., 2020. Identifying critical nodes in complex networks via graph convolutional networks. *Knowledge-Based Systems*, 198, p.105893.
- Zhang, Y., Gao, J. and Zhou, H., 2020, February. Breeds classification with deep convolutional neural network. In *Proceedings of the 2020 12th International Conference on Machine Learning and Computing* (pp. 145-151).
- Zhao, T., Hu, Y., Valsdottir, L.R., Zang, T. and Peng, J., 2021. Identifying drug–target interactions based on graph convolutional network and deep neural network. *Briefings in bioinformatics*, 22(2), pp.2141-2150.
- Zhao, T., Hu, Y., Valsdottir, L.R., Zang, T. and Peng, J., 2021. Identifying drug–target interactions based on graph convolutional network and deep neural network. *Briefings in bioinformatics*, 22(2), pp.2141-2150.