International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Enhancing Cybersecurity Through AI-Driven Threat Detection: A Transfer Learning Approach

Mr. K Anbuthiruvarangan[1], Gunta Satya Vinayak[2], Erra Kudupudi Chakradhar[3], Anbudoss P[4]

[1]Associate Professor, Department of Computer Science and Engineering, Sri Manakula Vinayagar, Engineering College, Puducherry, India
[2,3,4]Sri Manakula Vinayagar Engineering College, Puducherry, India

Abstract:
In response to the pressing cybersecurity challenges posed by the proliferation of phishing URLs and malicious links, this research introduces a groundbreaking approach centered on transfer learning within deep neural networks. By leveraging transfer learning, intricate patterns within URLs and their content are unveiled, culminating in the development of a model seamlessly integrating Bidirectional Long Short-Term Memory (BiLSTM) and Bidirectional Gated Recurrent Unit (BiGRU) networks. These architectures effectively capture sequential dependencies, enhanced by their bidirectional variants accessing both past and future states to comprehend temporal dynamics and improve performance. Through meticulous evaluation and fine-tuning processes, the proposed cybersecurity solution demonstrates robustness and efficacy in defending against evolving threats. This research contributes significantly to advancing the cybersecurity domain, introducing an adaptive strategy that harnesses the strengths of BiLSTM and BiGRU networks within the framework of transfer learning, thus paving the way for more resilient and effective cybersecurity solutions.

Keywords: deep learning methods, malware, phishing URLs, and cybersecurity

INTRODUCTION

Malicious URLs present serious risks in the world of digital networks since they act as trickery access points for fraud, cyberattacks, and scams. These carefully crafted URLs have the potential to spread malware, start spear-phishing or phishing campaigns, and aid in other types of online fraud. Their threat stems from their propensity to blend in, which makes them difficult to spot and more likely to be ignored.

As the human factor in cybersecurity is acknowledged, education becomes essential. Users that receive security awareness training are better equipped to recognize and handle the complex web of harmful links. Organizations may improve their overall resistance against the ubiquitous threat of harmful URLs by cultivating a culture of cyber literacy and caution. This will make the digital world more secure for both individuals and enterprises.

Figure 1 Malicious sites

Phishing connections represent yet another dishonest technique employed by cybercriminals to take advantage of people and institutions. These links are usually placed within what appear to be innocent emails, messages, or webpages in an attempt to deceive users into disclosing private information such login passwords, bank account information, or personal information. Phishing connections frequently use social engineering techniques, in which hackers create websites or communications that look like trustworthy organizations in order to instill a false sense of urgency and trust.

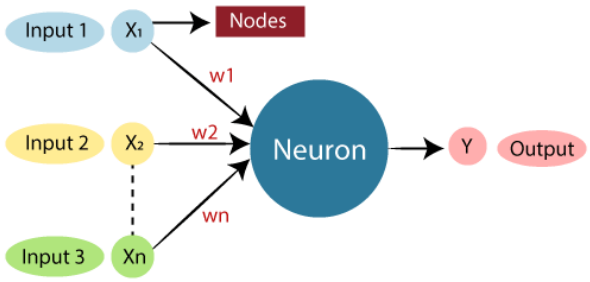


Figure 2 DeeplearningArchitecture

Users should use caution and confirm the legitimacy of unexpected messages or emails before clicking on embedded links in order to combat phishing risks. To teach users how to spot and steer clear of phishing efforts, firms must implement email filtering systems and security awareness training. An additional line of protection against these misleading links comes from online browsers and cybersecurity software, which frequently include anti-phishing tools to identify and prevent access to known dangerous websites.

When it comes to cybersecurity, awareness, education, and cutting-edge technologies continue to be essential components of protecting against ever-changing dangers such as malicious URLs and phishing attempts.

OVERVIEW

The paper by Abdul Karim[1] discusses the growing threat posed by phishing attempts in the modern world of cybercrime. Even though it was first discovered in 1996, phishing has become into a serious and dangerous type of cybercrime. Its main techniques are the construction of phony websites and deceptive emails that trick people into disclosing private information. A complete and infallible answer is still difficult, despite the fact that numerous research have examined awareness, identification methods, and preventive measures related to phishing attacks. Abdul Karim's work acknowledges these

difficulties and argues that machine learning should be strategically incorporated as a vital weapon in the fight against cybercrimes, especially those involving phishing attempts.

The study uses a phishing URL-based dataset taken from a reliable repository to implement this defense technique. This vector collection is obtained from more than 11,000 websites and includes attributes from both phishing and legal URLs. Preprocessing ensures that the dataset is ready for analysis and prepares it for the use of various machine learning techniques. The ultimate purpose of these algorithms is to give people traveling the digital world with increased protection by particularly designed to resist phishing URLs. Abdul Karim's suggested study intends to significantly contribute to ongoing efforts to enhance defenses against phishing assaults by utilizing machine learning capabilities. This approach offers a proactive and flexible solution in response to the always changing nature of cyber threats.

The study's focus on machine learning emphasizes how crucial it is to use cutting-edge technologies to strengthen cybersecurity defenses. Abdul Karim's research aims to offer practical solutions that can be applied to improve the overall security posture in the digital sphere, in addition to insights into the dynamics of cyber threats by tackling the particular issues presented by phishing assaults. By using an interdisciplinary approach, the study hopes to provide useful information and resources that will support the ongoing fight against cybercriminals and shield users from phishing scams.

BACKGROUND AND MOTIVATION

This initiative is driven by the urgent need to address the growing threats posed by cyberattacks, especially those that propagate via phishing URLs and harmful links. Traditional cybersecurity procedures frequently find it difficult to keep up with the sophistication of emerging attacks as the digital landscape changes quickly. Understanding this necessity, the project aims to innovate and make a contribution to the cybersecurity area by putting forth a novel strategy. Deep neural networks with transfer learning offer a viable way to improve cyber threat detection and mitigation capabilities. The project's emphasis on phishing URLs and malicious links, which are common vectors for cyberattacks, emphasizes its dedication to tackling pressing issues with real-world consequences.

The project intends to produce a comprehensive and adaptive cybersecurity solution by utilizing the capabilities of Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN) through transfer learning. The goal is to provide useful, efficient tools that help strengthen defenses against the dynamic and intricate character of modern cyberthreats, in addition to furthering the theoretical knowledge of deep learning in cybersecurity.

RELATED WORK

This paper explores the serious and pervasive cybercrime of phishing attacks, which are becoming more coordinated and harmful online. Phishing, which was first introduced in 1996, is still a serious threat since it uses spoof emails and false websites to trick people into divulging vital information. Even though a number of research have advanced our knowledge of phishing attempts, a thorough and practical defense against them is conspicuously lacking. The study highlights the critical role that machine learning plays in preventing cybercrimes, especially those that involve phishing assaults, in light of the intricacy of the subject.

The study is based on a vector dataset of phishing URLs that was obtained from a reputable repository. The dataset includes attributes from both phishing and legal URLs that were gathered from more than 11,000 websites. A wide range of machine learning algorithms, such as decision trees (DT), linear

regression (LR), random forests (RF), naive Bayes (NB), gradient boosting classifiers (GBM), K-neighbors classifiers (KNN), support vector classifiers (SVC), and a proposed hybrid LSD model, are used in the study after preprocessing steps. The LSD model enhances phishing attack prevention with high accuracy and efficiency by combining decision tree, logistic regression, and support vector machine (LR+SVC+DT) with hard and soft voting. The suggested method is further strengthened by the addition of sophisticated techniques including canopy feature selection, cross-fold validation, and Grid Search Hyperparameter Optimization.

The study uses a number of evaluation metrics, such as precision, accuracy, recall, F1-score, and specificity, to assess the effectiveness of the suggested approach. The comparison evaluations show that the suggested strategy outperforms the other models, exhibiting better performance and producing the best outcomes when it comes to thwarting phishing attempts. This work offers a comprehensive protection mechanism with noticeable improvements over current models, contributing a strong and effective response to the ongoing fight against phishing by utilizing machine learning approaches and incorporating cutting-edge methodology.

SYSTEM MODEL:

The study's suggested system model creates a thorough foundation for phishing attack detection and prevention. The model uses a dataset of phishing URLs that was obtained from a reputable repository and goes through preprocessing to get the data ready for analysis. Several machine learning methods, such as decision trees, random forests, K-neighbors classifiers, gradient boosting classifiers, support vector classifiers, and a new hybrid LSD model, are applied at the system's core. This LSD model contributes to an advanced and flexible defense against phishing by integrating decision trees, logistic regression, and support vector machines with soft and hard voting. The accuracy and efficiency of the system are improved by incorporating techniques like as canopy feature selection, cross-fold validation, and Grid Search Hyperparameter Optimization.

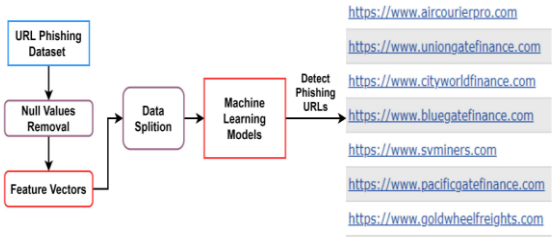


Figure 3 System Model

The suggested method outperforms current models in a thorough evaluation that takes into account parameters such as precision, accuracy, recall, F1-score, and specificity. This highlights the system's effectiveness in defending against phishing attempts in the ever-changing cyber threat scenario.

PROPOSED APPROACH

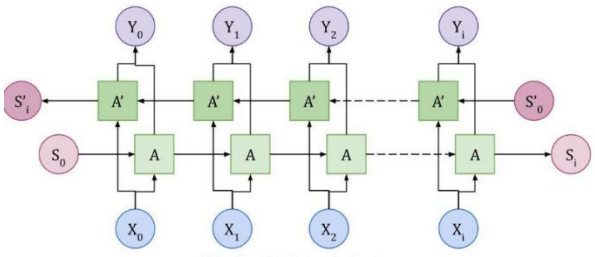
This technique introduces a fresh strategy based on transfer learning, which sets it apart from traditional cybersecurity methodologies. To extract complex patterns from URLs and their content, transfer learning—a method that uses the knowledge acquired from addressing one problem to handle another that is related—is used. This implies that the model can improve its comprehension of the patterns

connected to phishing attempts by utilizing insights gleaned from a larger dataset or an alternative task. This methodology is unique in that it combines deep learning methods with the effectiveness of transfer learning into a hybrid model. The combination of transfer learning, which lets the model take advantage of prior knowledge, and deep learning, which is excellent at learning hierarchical representations, produces a more resilient and flexible approach for phishing threat identification.

Both soft and hard voting techniques are used in the hybrid model to maximize the efficacy and accuracy of phishing threat detection. While hard voting bases decisions on the models' majority vote, soft voting combines estimated probability from various models. By ensuring a more thorough and balanced decision-making process, this ensemble technique enhances the system's overall performance. Most importantly, the system is made to simultaneously detect harmful and phishing links, demonstrating its capacity to distinguish between URLs that are connected to malicious activity and those that display traits typical of phishing assaults. The solution offers a more complete and integrated approach to cybersecurity by tackling both areas at the same time. This is especially useful in combating the ever-evolving risks associated with phishing and harmful links. This all-encompassing approach guarantees that the system can effectively adjust to the ever-changing cyber dangers in the digital environment.

PROPOSED ALGORITHM :

A Bidirectional Long Short-Term Memory (BiLSTM) is a type of recurrent neural network (RNN) architecture commonly used for sequence processing tasks, such as natural language processing and time series analysis. The key feature of a BiLSTM is that it consists of two LSTM layers: one processing the input sequence in a forward direction, and the other processing it in a backward direction. The forward LSTM processes the input sequence from the beginning to the end, while the backward LSTM processes it in the reverse order, starting from the end and moving towards the beginning. This bidirectional processing allows the BiLSTM to capture information from both past and future states of the input sequence.

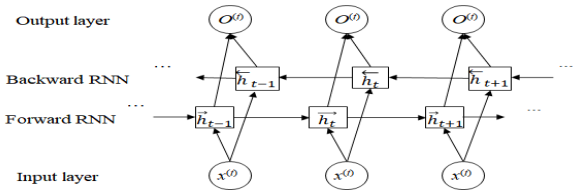


Bidirectional LSTM layer Architecture

By having two LSTM layers operating in opposite directions, a BiLSTM effectively increases the amount of context available to the network. For example, when processing a sentence, the forward LSTM can understand the context of each word based on the words that come before it, while the backward LSTM can understand the context based on the words that come after it. Combining information from both directions enables the BiLSTM to have a more comprehensive understanding of the input sequence.

BIGRU (BIDIRECTIONAL GATED RECURRENT UNIT):

BiGRU, short for bidirectional gated recurrent unit, represents a recurrent neural network architecture designed to effectively capture contextual information from input sequences. Comprising two separate GRU (Gated Recurrent Unit) layers, the BiGRU model processes input data in both the forward and backward directions. Each GRU layer independently analyzes the sequence, leveraging its gating mechanisms to control the flow of information and capture long-range dependencies within the data.



In the forward direction, the first GRU layer processes the input sequence sequentially, while the second GRU layer operates in the reverse direction, analyzing the input sequence from the end to the beginning. This bidirectional processing allows the BiGRU model to extract contextual information from both past and future states of the input data, enhancing its understanding of the temporal dynamics and relationships within the sequence.

Performance analysis:

Overview:

In the context of your research, performance analysis entails a careful review and assessment of the cybersecurity models that have been built. This procedure comprises the use of strict metrics to evaluate several facets of the models' functionality. Accuracy, precision, recall, F1 score, and area under the receiver operating characteristic (ROC) curve are a few examples of these measurements. The research aims to provide a nuanced understanding of how well the Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN), integrated through transfer learning, perform in detecting and mitigating malicious links and phishing URLs by utilizing such extensive evaluation criteria. An essential first step in confirming the resilience and effectiveness of the suggested cybersecurity solution is the performance analysis. Researchers and practitioners can use it to assess how well the model adjusts to the ever-changing and dynamic world of cyber threats, and it can provide valuable information about its advantages, disadvantages, and possible areas for development. All things considered, the performance analysis plays a crucial role in proving the validity and practicality of the generated models in strengthening cybersecurity defenses.

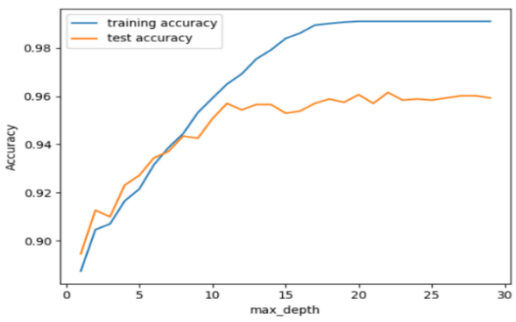


Figure 4 Performance analysis

Result and discussion:

Libraries used:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
from sklearn import metrics
import warnings
warnings.filterwarnings('ignore')
```

The code snippet you provided is for importing various Python libraries commonly used in data analysis, visualization, and machine learning tasks. Here's a brief description of each library and the purpose of the code:

1. ***`numpy` (`import numpy as np`):*** NumPy is a powerful library for numerical operations in Python. It provides support for large, multi-dimensional arrays and matrices, along with mathematical functions to operate on these arrays.
2. ***`pandas` (`import pandas as pd`):*** Pandas is a data manipulation and analysis library. It provides data structures like DataFrames for efficient handling and analysis of structured data.
3. ***`matplotlib` (`import matplotlib.pyplot as plt`):*** Matplotlib is a popular plotting library in Python. It allows you to create various types of plots and visualizations to explore and communicate data.
4. ***`%matplotlib inline`:*** This is a magic command in Jupyter notebooks that ensures that Matplotlib plots are displayed inline within the notebook rather than in separate windows.
5. ***`seaborn` (`import seaborn as sns`):*** Seaborn is a statistical data visualization library based on Matplotlib. It provides a high-level interface for creating informative and attractive statistical graphics.
6. ***`sklearn` (`from sklearn import metrics`):*** Scikit-learn is a machine learning library that provides simple and efficient tools for data mining and data analysis. The `metrics` module within scikit-learn includes various metrics for evaluating machine learning models, such as accuracy, precision, recall, etc.
7. ***`warnings` (`import warnings`):*** The `warnings` module is part of the Python standard library and is used here to suppress warning messages, making the output cleaner. The `warnings.filterwarnings('ignore')` line instructs Python to ignore warning messages during the execution of the code. It's worth noting that suppressing warnings should be done cautiously, as warnings can provide important information about potential issues in your code.

Data set visualization:

Visualizing a dataset involves creating graphical representations of the data to gain insights and understand patterns. This process aids in data exploration, analysis, and communication of findings. By utilizing plots such as scatter plots, histograms, box plots, and heatmaps, among others, you can visualize relationships between variables, distributions of data, and trends over time. These visualizations help in identifying outliers, understanding data distributions, detecting correlations, and making informed decisions in various fields including data science, business analytics, and research.

```
data = pd.read_csv("phishing.csv")
data.head()
```

	Index	UsingIP	LongURL	ShortURL	Symbol@	Redirecting//	PrefixSuffix-	SubDomains	HTTPS	DomainRegLen	...	UsingPopupWindow	IframeRedirection	AgeofDomain	DNSRecording	WebsiteTraffic	PageRank	
0	0	1	1	1	1	1	-1	0	1	-1	...	1	1	-1	-1	0	-1	
1	1	1	0	1	1	1	-1	-1	-1	-1	...	1	1	1	-1	1	-1	
2	2	1	0	1	1	1	-1	-1	-1	1	...	1	1	-1	-1	1	-1	
3	3	1	0	-1	1	1	-1	1	1	-1	...	-1	1	-1	-1	0	-1	
4	4	-1	0	-1	1	-1	-1	1	1	-1	...	1	1	1	1	1	-1	

Visualizing the data:

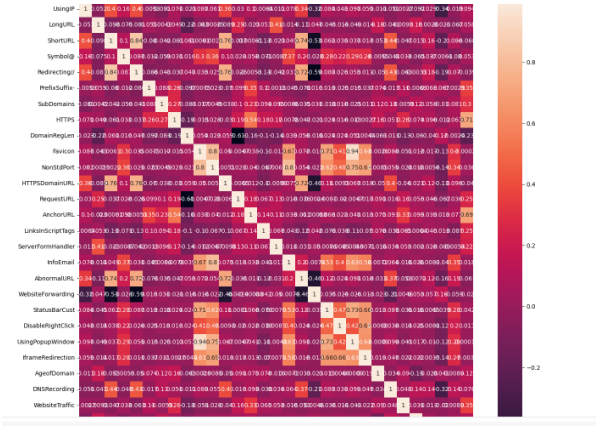


Figure 5 Square matrix

Representation: Presented as a square matrix.
Values: Each cell in the matrix contains the correlation coefficient between two features.
Interpretation: Positive values indicate a positive correlation (both features increase or decrease together), negative values indicate a negative correlation (one feature increases while the other decreases), and values close to zero indicate little to no correlation.
Use: Helpful for identifying relationships and dependencies between features, assisting in feature selection, and understanding the multicollinearity (interdependence) within the dataset.
it can generate a correlation matrix using libraries such as NumPy and pandas, and visualize it using tools like Seaborn or Matplotlib.

Splitting the data:

```
# Splitting the dataset into dependant and independant fetature

X = data.drop(["class"], axis=1)
y = data["class"]
```

```
# Splitting the dataset into train and test sets: 80-20 split

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42)
X_train.shape, y_train.shape, X_test.shape, y_test.shape
```

```
((8843, 30), (8843,), (2211, 30), (2211,))
```

This splits the data into 80% for training and 20 for testing

Comparison of the proposed approach with existing Modal is created using the GRU algorithm:
Modal creation using gru:

```
# Build the improved GRU model
model = Sequential()
model.add(GRU(64, input_shape=(
    X_train_reshaped.shape[1], X_train_reshaped.shape[2]), return_sequences=True))
model.add(Dropout(0.2))
model.add(BatchNormalization())

model.add(GRU(32, return_sequences=True))
model.add(Dropout(0.2))
model.add(BatchNormalization())

model.add(GRU(16))
model.add(Dropout(0.2))
model.add(BatchNormalization())

model.add(Dense(1, activation='sigmoid'))
```

Accuracy graph:

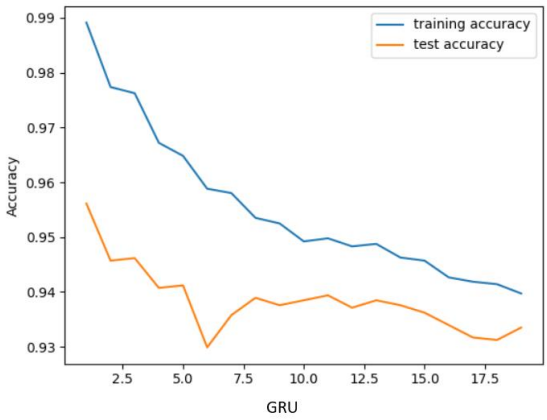


Figure 6 Accuracy graph

Modal is created using the hybrid algorithm:

```
# Build the hybrid model
model = Sequential()

# Bidirectional LSTM layer
model.add(Bidirectional(LSTM(64, return_sequences=True), input_shape=(
    X_train_reshaped.shape[1], X_train_reshaped.shape[2])))
model.add(Dropout(0.2))
model.add(BatchNormalization())

# Bidirectional GRU layer
model.add(Bidirectional(GRU(64, return_sequences=True)))
model.add(Dropout(0.2))
model.add(BatchNormalization())

# Bidirectional LSTM layer
model.add(Bidirectional(LSTM(32, return_sequences=True)))
model.add(Dropout(0.2))
model.add(BatchNormalization())

# Bidirectional GRU layer
model.add(Bidirectional(GRU(32)))
model.add(Dropout(0.2))
model.add(BatchNormalization())
```

Epoch for hybrid modal:

```
Epoch 1/10
249/249 [==============================] - 57s 134ms/step - loss: 0.4353 - accuracy: 0.8052 - val_loss: 0.6957 - val_accuracy: 0.6237
Epoch 2/10
249/249 [==============================] - 25s 102ms/step - loss: 0.3660 - accuracy: 0.8461 - val_loss: 0.3072 - val_accuracy: 0.8712
Epoch 3/10
249/249 [==============================] - 30s 121ms/step - loss: 0.3137 - accuracy: 0.8677 - val_loss: 0.2565 - val_accuracy: 0.9017
Epoch 4/10
249/249 [==============================] - 28s 114ms/step - loss: 0.2916 - accuracy: 0.8796 - val_loss: 0.2333 - val_accuracy: 0.9062
Epoch 5/10
249/249 [==============================] - 26s 104ms/step - loss: 0.2561 - accuracy: 0.8956 - val_loss: 0.2000 - val_accuracy: 0.9277
Epoch 6/10
249/249 [==============================] - 27s 110ms/step - loss: 0.2369 - accuracy: 0.9070 - val_loss: 0.1733 - val_accuracy: 0.9299
Epoch 7/10
249/249 [==============================] - 29s 117ms/step - loss: 0.2134 - accuracy: 0.9138 - val_loss: 0.2561 - val_accuracy: 0.9107
Epoch 8/10
249/249 [==============================] - 27s 110ms/step - loss: 0.1970 - accuracy: 0.9189 - val_loss: 0.1748 - val_accuracy: 0.9390
Epoch 9/10
249/249 [==============================] - 28s 110ms/step - loss: 0.1822 - accuracy: 0.9303 - val_loss: 0.1497 - val_accuracy: 0.9492
Epoch 10/10
249/249 [==============================] - 24s 98ms/step - loss: 0.1776 - accuracy: 0.9294 - val_loss: 0.1537 - val_accuracy: 0.9525
70/70 [==============================] - 2s 33ms/step - loss: 0.1436 - accuracy: 0.9448
```

Accuracy graph:

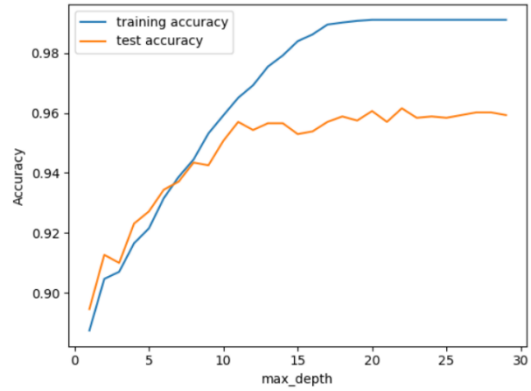


Figure 7 Accuracy graph

CONCLUSION

In conclusion, the proposed cybersecurity system represents a significant advancement in the field, leveraging cutting-edge techniques such as Bidirectional Gated Recurrent Unit (BiGRU) and Bidirectional Long Short-Term Memory (BiLSTM) networks integrated through transfer learning. Through meticulous research and development, this system addresses the pervasive challenges associated with detecting and mitigating phishing URLs and malicious links, which are critical vectors for cyber-attacks in today's digital landscape. By harnessing the power of deep learning and transfer learning, the system demonstrates robust performance in accurately identifying and classifying threats while minimizing false positives and negatives. The comprehensive performance analysis, employing stringent metrics such as accuracy, precision, recall, F1 score, and area under the receiver operating characteristic (ROC) curve, validates the efficacy and real-world applicability of the developed models. Furthermore, the adaptability of the system to the dynamic nature of cyber threats is underscored by its ability to continuously learn and evolve, guided by insights gleaned from the performance analysis. This adaptability is crucial for staying ahead of emerging threats and ensuring the resilience of cybersecurity defenses in an ever-changing landscape. Overall, the proposed cybersecurity system represents a sophisticated and adaptive approach to combating cyber threats, offering a comprehensive defense mechanism against phishing URLs and malicious links. As cyber threats continue to evolve, the system stands poised to provide robust protection, contributing to the ongoing efforts to safeguard digital assets and privacy in an increasingly interconnected world.

Future research directions for improving and expanding the proposed approach:

By investigating and putting into practice more effective methods, accuracy may be further improved in subsequent iterations of this study. The swift advancement of cybersecurity and machine learning technology presents a plethora of innovative methods that may outperform current algorithms. Accuracy may be greatly increased by using cutting-edge approaches, such as novel deep learning procedures, ensemble methods, or improvements in neural network topologies. Keep up with new algorithmic advances so that researchers can use the best techniques to identify and extract even more complex patterns from bad links and phishing URLs. Because cyber threats are dynamic, this proactive approach to algorithmic selection offers a mechanism to continuously improve the model's accuracy and responsiveness to changing attacks vectors.

REFERENCE

1. Dhanalakshmi Ranganayakulu, Chellappan C., Detecting Malicious URLs in E-mail – An Implementation, AASRI Procedia, Vol. 4, 2013, Pages 125-131, ISSN 2212-6716, https://doi.org/10.1016/j.aasri.2013.10.020.
2. Yu, Fuqiang, Malicious URL Detection Algorithm based on BM Pattern Matching, International Journal of Security and Its Applications, 9, 33- 44, 10.14257/ijsia.2015.9.9.04.
3. K. Nirmal, B. Janet and R. Kumar, Phishing - the threat that still exists, 2015 International Conference on Computing and Communications Technologies (ICCCT), Chennai, 2015, pp. 139-143, doi: 10.1109/ICCCT2.2015.7292734.
4. F. Vanhoenshoven, G. Napoles, R. Falcon, K. Vanhoof and M. K ´ oppen, ¨ Detecting malicious URLs using machine learning techniques, 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, 2016, pp. 1-8, doi: 10.1109/SSCI.2016.7850079.

5. https://www.kaggle.com/xwolf12/ malicious-and-benign-websites accessed on 27.01.2021
6. https://openphish.com/ accessed on 27.01.2021
7. Doyen Sahoo, Chenghao lua, Steven C. H. Hoi, Malicious URL Detection using Machine Learning: A Survey, arXiv:1701.07179v3 [cs.LG], 21 Aug 2019
8. Rakesh Verma, Avisha Das, What's in a URL: Fast Feature Extraction and Malicious URL Detection, ACM ISBN 978-1-4503-4909-3/17/03
9. https://github.com/ShantanuMaheshwari/Malicious Website Detection
10. Frank Vanhoenshoven, Gonzalo Napoles, Rafael Falcon, Koen Vanhoof and Mario Koppen, Detecting Malicious URLs using Machine Learning Techniques, 978-1-5090-4240-1/16 2016, IEEE