Improving Phishing Detection Systems Using Kolmogorov-Arnold Networks

By

Efadul Islam
ID: 202-15-14415

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised by

Dewan Mamun Raza
Assistant Professor
Department of Computer Science and Engineering Daffodil International University

Co-Supervised by

Mr. Mohammed Sami Khan
Lecturer
Department of Computer Science and Engineering Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

May 14, 2025

APPROVAL

This Project titled **"Improving Phishing Detection Systems Using Kolmogorov-Arnold Networks"**, submitted by Efadul Islam, ID No: **202-15-14415** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **14 May, 2025**.

BOARD OF EXAMINERS

Dr. Md. Fokhray Hossain (MFH) — **Chairman**
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md Alamgir Kabir (DMAK) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Ms. Aliza Ahmed Khan (ADK) — **Internal Examiner**
Sr. Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Nazibur Rahman — **External Examiner**
Technical Lead - Database Administrator
Telenor - Grameen Phone Account

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Name of the Supervisor**, **Supervisor's Designation**, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

14.05.2025

Dewan Mamun Raza
Assistant Professor
Department of Computer Science and Engineering Daffodil International University

Co-Supervised by:

14/05/2025

Mr. Mohammed Sami Khan
Lecturer
Department of Computer Science and Engineering Daffodil International University

Submitted by:

Efadul

Efadul Islam
Student ID: 202-15-14415
Department of Computer Science and Engineering Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project(FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Dewan Mamun Raza, Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Management System Machine Learning (ML) and Cyber Security** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting themat all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the **Dr. Sheak Rashed Haider Noori** Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

The phishing activity of cybercrimes accounts to about 30% and is a great threat as over 240,000 cases were documented in 2020 only. Generally, these come in the form of false emails and websites, which aim at stealing essential information that leads to identity theft, financial losses, and compromising through security breaches. Phishing distrusts digital services and largely hits the world economy. This paper uses deep learning architecture capable of processing complex data named Kolmogorov-Arnold Networks (KAN), to set up a phishing detection system. The system achieves more accurate, precise, recall, and F1-score than the typical machine learning models by processing data from URLs, emails, and network traffic and minimizing false positives. The technology is dynamic and it changes all the time, therefore, it is quite good at detecting new phishing maneuvers. Future work towards better real time-detection and enhanced resistance against emerging cyber threats, our work provides a reliable and straightforward method to phishing detection.

Keywords- Phishing Detection, Cybersecurity, Real-time Monitoring, Deep Learning, Kolmogorov Arnold Networks (KAN)

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Phishing attacks constitute a prevalent and dynamic issue within the realm of cybersecurity. These nefarious operations employ deceptive strategies aimed at coercing individuals to disclose sensitive personal information, including usernames, passwords, credit card information, and other protected data. Cybercriminals impersonate legitimate entities, including banks, government institutions, and reputable companies, to manipulate victims into actions that jeopardize their security, such as clicking on malicious links, entering personal information on fraudulent websites, or downloading harmful attachments. The magnitude and consequences of phishing assaults have escalated considerably, with millions of individuals and organizations targeted each year, resulting in huge financial losses, identity theft, and breaches of personal data.

The Anti-Phishing Working Group (APWG) defines phishing as "a crime that utilizes social engineering and technical deception to acquire consumers' identity information and financial account credentials." In recent years, the complexity and diversity of phishing techniques have significantly escalated, extending beyond conventional email attacks to encompass platforms such as social media, blogs, online forums, and blockchain networks. According to a 2020 analysis, Software-as-a-Service (SaaS) and webmail users were 33.7% of the phishing assaults victims, which underlines the widespread nature of this threat. One of the key challenges in phishing struggles is the dynamic nature of the strategies: that is, the popular practice of the attackers with respect to using SSL encryption to make discerning between genuine and fake sites difficulty. In 2020, nearly 75 percent of phishing sites have been found to use SSL encryption; a trait previously associated with safe, credible websites. Indeed, this tendency makes the detection attempts even more complicated and emphasizes the need for more advanced detection technologies.

In reaction to this increasing danger, researchers and cyber security experts have incorporated advanced detection practices, e.g. heuristic-based approaches and machine learning based strategies. Heuristic methods look at URL and site attributes to deduce phishing conduct patterns, while the machine learning and deep learning algorithms work towards detecting novel and shifting attack

patterns by identifying patterns from data. However, such methodologies are often limited by use of existing data and struggle to detect zero-day phishing attacks, which rely on new strategy, or unrecognized method. Therefore, one can note a need for more flexible and precise methodologies to tackle such problems.

One of them is the Kolmogorov-Arnold Network (KAN), a new neural network scheme formulated on the basis of Kolmogorov-Arnold representation theorem. This theorem states that any multivariate continuous function can be transformed as a sum expression of univariate functionalities, allowing a modulization of complex data with greater dimensions in a more understandable form. In the field of phishing detection, KAN can also be applied for analyzing various data types including the characteristics of website, content of email or network traffic to effectively detect the phishing attempts. Disaggregating complex data into smaller parts, KAN will be able to provide a more specific and effective solution to the detection of phishing attacks, and this, particularly, on new or unfamiliar methodologies of attacks.

This thesis explores the usage of Kolmogorov-Arnold Networks for phishing detection, whereby significant developments in the fields of machine learning and deep learning frameworks are considered. This project's goal is to augment the phishing detection systems by identifying the existing methods, and incorporating the benefits of KAN for greater efficiency, flexibility, and persistence. The pertinent datasets from Kaggle will be analyzed during this study to clarify possible improvements when it comes to phishing detection and support the overall aim of improving the cybersecurity system on a more digital environment.

1.2 Motivation

Phishing is one of the most common and deadly threats on the internet right now. They deceive people into disclosing their confidential information such as bank account numbers, password amongst other. There are many options, even though attackers are always devising new ways of deceiving the phishing detection systems. Under such circumstances, it is necessary for us to improve our phishing detection techniques. A new type of neural network called Kolmogorov-Arnold Networks (KANs), which might be used to enhance phishing detection and its decision-making process, is what I am going to use in my work. This is important as we need a system that we can trust and understand, other than a system that works effectively. As a result of working on this topic, I am gaining an increased understanding of the state-of-the-art machine learning and cybersecurity technology. It also gets me ready. It also prepares me to work with practical problems in the field and help to create safer digital environments.

1.3 Objectives

This research study will attempt to study and ameliorate methods used in phishing detection using Kolmogorov-Arnold Networks (KAN). The primary goal is to increase the ability of phishing detection systems to be precise, flexible, and scalable. The work seeks to bring KAN into the existing framework of methodology to compensate for the failures of the common procedures and address the problem of the detection of new zero-day phishing attacks.

- Assess the effectiveness of currently available phishing detection techniques, ranging from the heuristic-based to the machine learning approach.

- Study the effectiveness of Kolmogorov-Arnold Networks (KAN) with regards to improvement of phishing detection accuracy by simplification of complex, high-dimensional data.

1.4 Methodology

At the first step of our phishing detection approach, we gather six data sources to combine legitimate and malicious samples. Normalizing this data in to a standard format and doing pre-processing to remove invalid URLs give us over 3.4 million clean records. Feature extraction is next in order, in which we derive 53 features from the URLs. These characteristics are standardized across the dataset in order for consistency. In order to classify we are using the Kolmogorov-Arnold Network (KAN), a new neural network, capable to decompose complex patterns into simpler mathematical expressions, which can be interpreted by humans and respectively can detect phishing sites with high accuracy with a lower computational cost. For comparison purposes we exploit ten other machine learning algorithms on the same features. finally we test the model's performance of rigorously using 3 publicly available benchmark test sets to test its efficacy.

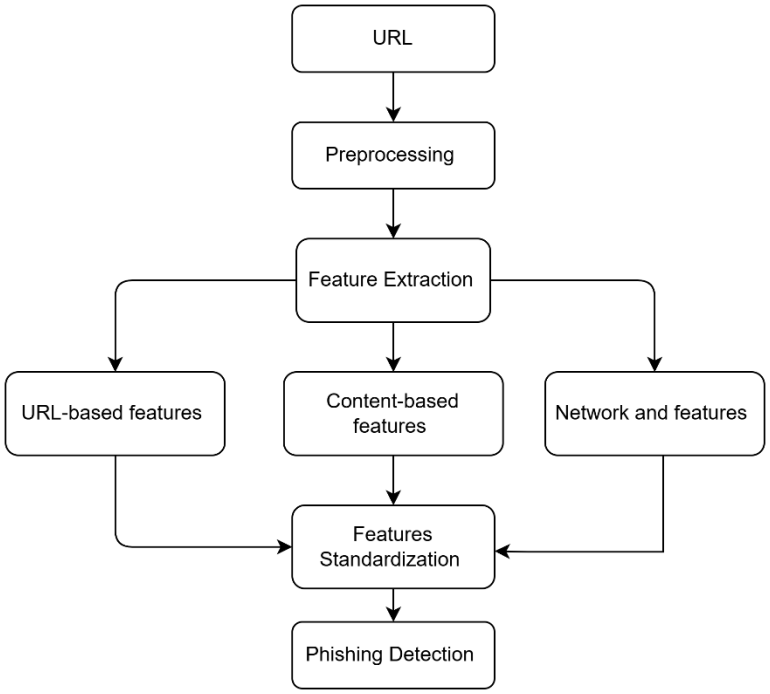

Figure 1.4: Typical Phishing URL Detection Approach

1.5 Project Outcome

The main result of this project is the establishment of a strong interpretable system to identify phishing at the Kolmogorov-Arnold Network (KAN). This system safely categorizes URLs, emails, and web pages to being either legitimate or malicious, using a variety of features extracted from the contents, network, and URL features. The outcomes include:

- Cleaned and standardized phishing dataset of more than 3.4 million entries acquired from Six reputable platforms.
- **feature-rich dataset** incorporating URL-based, content-based, network-based, and email-based indicators.
- largely functional KAN-based phishing classifier with enhanced interpretability, reduced computational complexity as compared to the classical deep learning models.
- visualized the phishing detection pipeline that can be used in the research for the security purposes, in industrial applications or even in the further academic researches.
- Possible link to real-time email/web filtering platforms, improves defense mechanisms in phishing attacks.
- framework that is capable of being adjusted for the detection of other types of cyber threats (e.g., spam, malware links) by supplementing the set of features and the model.

1.6 Organization of the Report

Several chapters are presented in this research to discuss a separate step of the phishing detection project with the use of the Kolmogorov-Arnold network (KAN).

- **Chapter 1** introduces the project formulating the motivation to implement it, objectives, scope, methodology preview, and the expected results. It also shows how the particular report is constructed to lead the reader along the development and review process.
- **Chapter 2** provides a detailed overview of the existing literature (previous phishing detection methods, the machine learning model of cybersecurity, and the theoretical background of Kolmogorov-Arnold Network). It determines what is missing from existing methodologies and explains why KAN can be used for this purpose.
- **Chapter 3** The system architecture and methodology are distinctly described. It includes the data collection process, normalization techniques, preprocessing steps, feature extraction methods, and feature standardization processes. The design rationale for each step is discussed to highlight its contribution to overall detection accuracy.
- **Chapter 4** is concerned with the implementation of the phishing detection system. Its description includes integrating the datasets, creating the KAN model, and environment settings for training and testing. The specific challenges experienced during implementation and how it was overcome has also been covered in this chapter.
- **Chapter 5** The outcomes and performance of the model are provided. Metrics that include accuracy, precision, recall and F1 score are used to evaluate the detection capability. Comparison with other models is also presented to show the effectiveness of the presented system.
- **Chapter 6** is the last one, which sums up the main findings and mentions an importance of the results. It also suggests future improvements and research avenues, especially when dealing with making real-time detection capability better and increasing the diversity in features.

Chapter 2

Background

2.1 Introduction

Phishing attacks have become one of the most common and harmful cybersecurity risks in the digital age. Such attacks use techniques associated with social manipulations to deceive people and made them provide confidential personal and financial information such as usernames, passwords, and credit card information [1]. There has been an escalation in phishing schemes in the recent past with hackers using various tactics – through emails, fake websites, social media, and instant messaging sites – to con their victims. The current form of the phishing faces immense dangers to people, companies, governments and mammoth organizations resulting in financial losses, identity theft and leaks of personal details[2].

Despite the increasing level of threat, phishing detection remains a daunting task due to constantly expanding attack routes and intensified utilization of legal security methods by the attackers such as SSL encryption. Traditional systems of detection, which are primarily relying on heuristic analysis or signature-based techniques, have often struggled to keep up with the sophistication and newness of modern phishing methods. The machine learning and deep learning approaches have emerged as promising solutions for these problems; still, they lack capabilities regarding discovering peculiar zero-day phishing attacks, since these systems rely on patterns and data that have been previously presented[3].

An especially interesting method for enhancing phishing detection is the Kolmogorov-Arnold Network (KAN), a sophisticated neural network model derived from the Kolmogorov-Arnold representation theorem. This theorem posits that any multivariate continuous function can be represented as the summation of simpler univariate functions[4]. KAN in phishing detection offers a new approach to the study of high-dimensional information, Like attributes of websites, content of emails, or activity on networks, which can be decomposed to understandable elements. This dismantling can improve the detection precision and flexibility of phishing detection systems, specifically with new and unknown threats. This chapter provides an extensive overview on the occurrence of phishing detection methods, indicating the imperfections of traditional systems and potential advantages of initiating the use of Kolmogorov-Arnold networks to eliminate this perennial security problem in cyber. This emphasizes the significance of the datasets that were used in the current study which were retrieved from Kaggle, to allow for development of an effective and flexible phishing detection framework. This research aims to enhance the effectiveness of

phishing detection systems and take the science of cybersecurity further by exploring new ways.

2.2 Literature Review

According to recent studies, Kolmogorov-Arnold Networks (KANs) provide great accuracy, efficiency, and interpretability in cybersecurity activities such as phishing and intrusion detection. Hybrid models and architectural advancements show strong potential for creating reliable and portable detection systems.
Kuznetsov et al. [5] 's lightweight Denial of Service attack detection technique for IoT scenarios uses Kolmogorov-Arnold Networks. It achieves 99.0% accuracy on the CICIDS2017 dataset and reduces memory requirements by up to 98% compared to current alternatives.

Amouri et al. [6] combined KANs with the XGBoost algorithm to create a hybrid intrusion detection system (IDS). This ensemble method uses XGBoost's classification capabilities and KANs' adaptive activation functions. When tested on the N-BaIoT dataset, the hybrid model outperformed standalone KAN and MLP models with an accuracy of 99.69%, precision of 98.10%, recall of 98.01%, and F1-score of 98.04%.

Lu et al. [7] investigated how well KANs work for detecting fraud. They suggested a heuristic for hyperparameter tweaking and presented a judgment method for evaluating KAN appropriateness using Principal Component Analysis (PCA). The refined KAN model, which was developed using a genetic algorithm, showed promise in fraud detection situations with precision of 99.01%, recall of 90.91%, F1-score of 94.79%, and accuracy of 94.42%.

With a focus on efficiency and interpretability, Barašin et al. [8] utilized KANs for time series categorization. They outperformed conventional MLPs on the UCR benchmark, demonstrating competitive accuracy and efficiency compared to HIVE-COTE2.

Bresson et al. [9] presented KAGNNs by incorporating KANs into graph neural networks (GNNs). They used B-splines and radial basis functions to create three KAN-based GNN layers modeled after the GCN, GAT, and GIN architectures. KAGNNs can model complex relationships in cybersecurity data, as evidenced by extensive studies on node classification, link prediction, and graph classification tasks that showed them to perform on par with or better than typical MLP-based GNNs.

Mollaali et al. [10] presented nonformalized KANs, which combine KANs with conformal prediction algorithms to give calibrated prediction intervals with guaranteed coverage. This connection improves KANs' interpretability and dependability, which is critical for applications like phishing detection, where it's critical to comprehend model confidence.

Somvanshi et al. [11] gave a thorough analysis of KANs, including information on their theoretical underpinnings, architectural advantages, and applications. In addition to discussing developments like Temporal-KAN, FastKAN, and PDE-KAN, the paper investigates how KAN may be integrated with other architectures like transformer-based, convolutional, and recurrent models. This survey provides insightful information about KAN's adaptability and potential across various fields,

including cybersecurity.

Kurkova-Kolmogorov-Arnold Networks, or KKANs, are a novel design that combines variable linear combinations of basic functions with MLP-based inner functions, as shown by Toscano et al. [12]. They showed that KKANs outperformed original KANs and conventional MLPs in function approximation and operator learning tasks. Three universal learning stages, fitting, transition, and diffusion, were identified by analysis using information bottleneck theory, which further explained the dynamics of KKANs.

Numerous research efforts have attempted to leverage deep learning for phishing detection by employing various architectures and datasets. Table 1 summarizes key studies in this domain, highlighting their methodologies, strengths, and limitations.

Table 2.1: Summary of Literature Reviewed.

Author's	Year	Tittle	Methodology	Key Findings
Opara et al.[13]	2024	Detecting phishing web pages by exploiting raw URL and HTML characteristics	Used raw feature analysis from URLs and HTML content with deep models.	Achieved high precision by avoiding feature engineering.
Kumar et al.[14]	2023	SI-BBA: Phishing detection using Swarm Intelligence and deep learning	Combined bee colony algorithm for feature optimization with DL models.	Increased detection speed and accuracy through intelligent feature reduction.
Prabakaran et al. [15]	2023	Phishing detection using variational autoencoders	Used VAE to model URL behavior and identify anomalies.	Improved detection of zero-day phishing attacks.
Hussain et al. [16]	2023	CNN-Fusion: A lightweight phishing detection method based on multi-variant ConvNet	Proposed a fusion of multiple lightweight CNNs on URL and textual features.	Achieved high performance with low computational cost.
Almousa et al. [17]	2022	Phishing website detection: How effective are deep learning-based models and hyperparameter optimization	Compared various DL models with grid search and Bayesian optimization.	Hyperparameter tuning significantly boosted detection accuracy.
Tang et al.[18]	2021	A deep learning-based framework for phishing website detection	Employed LSTM with embedding layers to model URL and content sequences.	Outperformed traditional ML models in detection accuracy.
Zhu et al.[19]	2020	DTOF-ANN: An artificial neural network phishing detection model based on decision tree and optimal features	Combined decision tree-based feature selection with ANN for classification.	Improved feature efficiency and reduced overfitting.
Yang et al.[20]	2019	Phishing website detection based on 14 multidimensional features driven by deep	Used deep neural networks on 14 handcrafted features across multiple	Achieved high accuracy by capturing diverse feature

		learning	categories.	dimensions.

2.2.1 Related Research

phishing attacks remain a real threat to online security by using fake websites to collect sensitive information about the user. Traditional detection methods focus on analyzing the structure of URLs, content of the webpages, and network traffic. However, many systems struggle with the amplification of phishing site sophistication, that constantly work on changing their ways to not get detected. In this respect, the machine learning methods have emerged as a feasible technique to improve the target detection accuracy. Kolmogorov-Arnold Networks (KANs), one of the cases of neural networks, has proved its effectiveness in the processing of complex data and in the detection of subtle patterns, which traditional systems may overlook. The unique construction of KANs allows these to operate more efficiently for the detection of nonlinear relations in data, and consequently, it makes them very proficient in differentiating between real and dishonest websites. This research explores the Kolmogorov-Arnold Networks application towards improved phishing detection systems, revealing the incorporation of the systems with the existing methodologies to boost efficiency.

The use of Kolmogorov-Arnold Networks in detection of phishing marks a novel approach with high promise of avoiding the limitations of currently implemented detection techniques. By using KANs that are designed to perceive and replicate complex non-linear data patterns, phishing detection systems can be trained to spot subtle features in the architectures of websites that indicate phishing efforts. KANs' flexibility and accuracy make them extremely proficient in recognizing phishing websites, even if they use advanced evasion techniques such as mimicking true sites, or using new addresses. What is more, the potential of KANs to reduce the feature dimensionality while retaining high detection accuracy enables a more efficient and scalable solution. With the increasing complexity of phishing attacks, incorporation of Kolmogorov-Arnold Networks into the phishing detection systems would arguably be a promising improvement in the development of stronger and more adequate cybersecurity systems.

2.3 Gap Analysis

Table 2.3: Gap Analysis

Feature	Yang et al	Zhu et al.	Tang et al.	Almousa et al.	Hussain et al.	Proposed System (KAN)
High accuracy detection	Yes	Yes	Yes	Yes	Yes	Yes
Real-time phishing detection	No	No	No	Yes	Yes	Yes
Low false positive rates	No	No	No	Yes	No	Yes
Multi-feature extraction (URL, content, network)	Yes	Yes	Yes	Yes	Yes	Yes
Adaptability to new phishing tactics	No	No	No	Yes	No	Yes
Incorporation of hybrid models	No	Yes	No	No	No	Yes
Explainability and interpretability	No	No	No	No	No	Yes
Handling imbalanced datasets	No	No	No	Yes	Yes	Yes
Real-time update with new data	No	No	No	Yes	No	Yes
Generalizing to various platforms (email, social media, etc.)	No	No	No	No	No	Yes
Integration of WHOIS and user behavior monitoring	No	No	No	No	No	Yes
Adversarial attack resistance	No	No	No	No	No	Yes

2.4 Summary

Overall, the existing work exhibits an array of approaches from rather primitive feature-based models to highly complex deep learning networks. Some of the methodological enhancements are the application of dynamic detection, recurrent architectures, as well as automatic feature extraction. The usability of these systems is further demonstrated by web and mobile applications though a lot of them are proprietary or have a limited ability to adapt. The present study extends these contributions by filling the following gaps, namely, the detection of short URL, feature richness, and the scalability of dataset, with the aim of enhancing detection accuracy and the robustness of model.

Chapter 3

Research Methodology

3.1 Methodology

This study establishes a phishing detection system utilizing Kolmogorov-Arnold Networks (KAN). Data is gathered from six varied sources, standardized to a uniform format, and preprocessed by eliminating invalid URLs. Essential attributes including URL length, domain details, and SSL implementation are extracted for model input. These attributes are used to train the KAN model on categorizing the URLs as phishing or legitimate. The performance of the model is then evaluated using accuracy, precision, recall and F1-score on a test set that is separate from the training set. Such approach ensures tough, flexible and accurate phishing identification.

3.1.1 Overview

This study methodology centers on creating an advanced phishing detection system utilizing Kolmogorov-Arnold Networks (KAN), a deep learning framework adept at processing intricate, high-dimensional data. The study commences with the aggregation of data from six esteemed sources, including Kaggle, PhishTank, and ISCX-URL2016 The found datasets are an enormous set of phishing and honest URLs, which guarantees that the algorithm is instructed in various tactics of attacks. This methodology makes the model more robust, and it becomes possible to make productive generalization to unseen phishing techniques.
After data collection, the process that is carried out is normalization. There are different sources of data which come in various forms and at times if one is to process them, one may need to standardize them. This is a stage that helps in ensuring that all the datasets are converted to a uniform mode whereby the model can use the data properly. Normalization process includes the transformation of URLs to lowercase, removing the characters, which are not necessary, and normalization of domain names, ensuring the data is pre-processed for the machine learning model similarly.

The next phase is pre-processing whereby invalid URLS, dupes and incomplete entries are weeded off. The step is critical for ensuring the integrity of data used to train the model. After preprocessing, data are composed of a total of 3,426,950

URLs of which 2,376,761 are legal URLs whereas 1,050,189 are phishing URLs. This neutralized dataset is very essential as a training set needed to ensure the model learns how to recognize a phishing attempt without any skewed preference towards a particular class.

Feature extraction is an important part of the process because it involves identification of those most relevant properties of URLs that can distinguish between a phishing site and an authentic one. Extracted key components include the URL's length, specifics on the domain, use of special characters, whether SSL is implemented and the information about the subdomain. These features are chosen because they are relevant in phishing identification and the ability to provide useful information to the model. The features are then converted to numerical values once extracted hence making them suitable for introduction into the Kolmogorov-Arnold Network. The Kolmogorov-Arnold Network is finally trained on the extracted features so as to categorize URLs as either phishing or authentic. The peculiar ability of the KAN model to decompose complex data into less cumbersome data enhances its effectiveness, especially in discovering new phishing attacks. Post the training, the model is evaluated on a separated test dataset with performance measures such as accuracy, precision, recall, and F1-score to assess the model. This assessment tests the model for its generalization to new data assuming its efficacy in real-world phishing detection applications.

3.1.2 Proposed Methodology

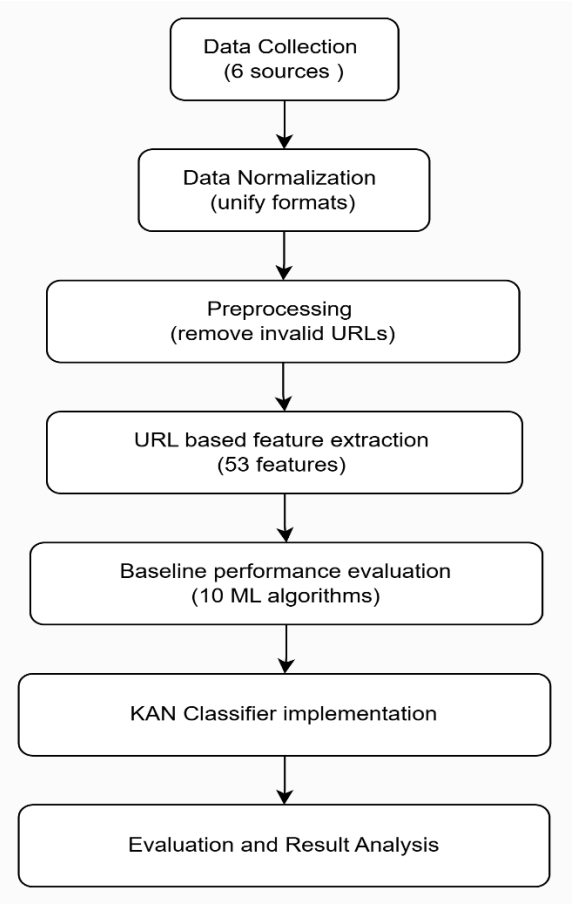

Figure 3.1: This is a Proposed Methodology

3.1.3 Functional and Nonfunctional Requirements

Functional Requirements:

1. Data Collection: The system needs to gather data from at least six different sources to have a diverse phishing and legitimate URLs for good model training and testing.
2. Data Normalization: The system needs to normalize data format from various sources to make them consistent so as to guarantee that data can be processed through in successive stages without errors.
3. Preprocessing: The system has to cleanse the dataset by eliminating invalid URLs, so that it is only valid and relevant data that is utilized for training the concerned models.
4. Feature Extraction: The system has to produce at least 53 features within the URLs which are very critical in differentiating between a phishing website and a legitimate website.
5. KAN Classifier Implementation: The system should have the Kolmogorov-Arnold Network (KAN) classifier to detect the phishing URLs correctly from the extracted features.
6. Baseline Evaluation: The system will have to compare the performance of KAN with 10 other machine learning algorithms to derive the operating baselines and get the best model for phishing detection.
7. Evaluation and Result Analysis: The system needs to assess the model based on such metrics as accuracy, precision, recall, and F1-score and analyze the results to find the flaws to be improved.

Nonfunctional Requirements:

1. Scalability: The system should be able to support large sets of data, such as millions of URLs, and make efficient scaling as the data grows in size.
2. Performance: The system needs to optimise the processing time of the model for training and inference so that a real-time phishing detection is feasible with insignificant delay.
3. Security: The system should adhere to the best practice for data security so that the user's sensitive information is managed and stored safely according to the regulations such as GDPR.
4. Reliability: The system needs to be reliable that will result in minimal downtime and errors, error handling without causing failures or inaccurate results.
5. Usability: It should have a user interface that is intuitive and easy to use even for the non-technical users who can perceive and interact with results

of phishing detection effortlessly.

6. Maintainability: The system should be equipping easy update and maintenance with proper documentations and modular components so that improvements and addition of functionalities can be made continuously.
7. Energy Efficiency: If deployed in the cloud, the system should be based on energy-efficient resources to reduce the environmental impact, corresponding with sustaining goals.

3.1.4 Context Diagram

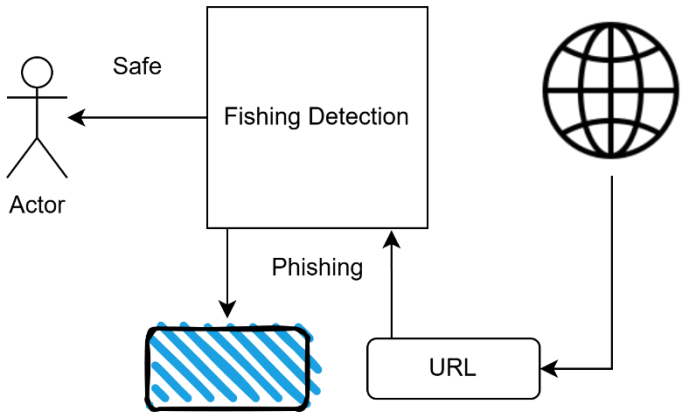

Figure 3.1.4: Context Diagram

A context diagram shows the system relationships with the external entities as well as data flow between the phishing detection system and other components or users. The context diagram for the phishing detection system that uses Kolmogorov-Arnold Networks (KAN) would include the following:

essential components:

- Users: People who communicate with the system, giving them URLs or any other information for the detection of phishing means.
- Data Sources: External sources (Kaggle, PhishTank, ISCX-URL2016, etc.) on which the system obtains phishing and non-phishing data for training and testing
- Model Evaluation Instruments: Instruments or frameworks utilized for assessing the performance of KAN in comparison to other machine learning algorithms.
- Data Storage/Cloud Services: Secure data storage solutions, such as databases or cloud storage, that provide system access for analytical purposes.

System Processes/Modules:

- Data Collection: Accesses different internet sources .
- Data Normalization: Integrates data from various sources to a unified form for further processing.
- Preprocessing: Makes the data clean by removing erroneous or irrelevant URLs from it.
- Feature Extraction: Extracts relevant characteristics from URLs for use in KAN model.
- Model Training and Testing: Uses KAN for phishing identification including the model evaluation and comparison of other algorithms.
- Evaluation and Result Analysis: Assesses the model's effectiveness on the basis of such measures as accuracy, precision, recall and F1-score.

Data Flows:

- Data Input: Raw URL data coming from users and external sources is the one received and then preprocessed for training processes.
- Processed Data: Cleaned and standardized data is passed onto the feature extraction module where relevant features are identified and used to train.
- Results: When evaluated, results are passed on to the user with descriptions of the phishing detection results and performance metrics.

3.1.5 Data Flow Diagram Level 1

- The Level 1 Data Flow Diagram (DFD) defines the basic processes, external entities, and the data flow in the phishing detection system to represent the movement and processing of data in searching for phishing threats.
- Data Collection: The first step is collecting data from various sources, embracing user inputs and other sites, including Kaggle, PhishTank, and funda – ISCX-URL2016. Users provide URLs to be checked, and data is mined from credible sources to form a holistic and representative dataset of both Phishing as well as good URLs. This step is important to ensure the system can obtain data which are present in the real scenario in order to train and test the phishing detection model.
- Data Normalization: Upon data collection, hence the next step is normalization. Data normalization ensures that all the information collected, in whatever scope, meets a uniform standard. This stage is necessary since in most cases data can be sourced from a variety of sources with non consistent formats and standardization helps with processing and analysis. Normalization ensures compatibility in data formats hence the system can effectively manage and convert data to forms that are usable

- Preprocessing: After normalization, the data is preprocessed to remove any invalid, partial or irrelevant URL. This process ensures the use of quality and precise data for further analytic stages. Preprocessing is necessary for preparing the dataset of the highest possible purity and relevance to the system, in particular, from noise like duplicate data records and incorrect URLs. This stage ensures that the data used for the training and testing of models is true and reliable.
- Feature Extraction: The final step that occurs right before the model training is feature extraction. This stage involves identifying the key characteristics of the URLs such as the domain information, URL length, presence of special characters as well as the URLs' SSL encryption among others.

  These characteristics are important in distinguishing between phishing and a real URL. While extracting relevant information, the system outputs the necessary data for KAN model to make perfect predictions concerning the legitimacy of a URL which is either doing phishing or authentic. The gathered features are then used to teach the Kolmogorov-Arnold Network (KAN), the main model for detection of phishing attempts.

3.2 Detailed Methodology and Design

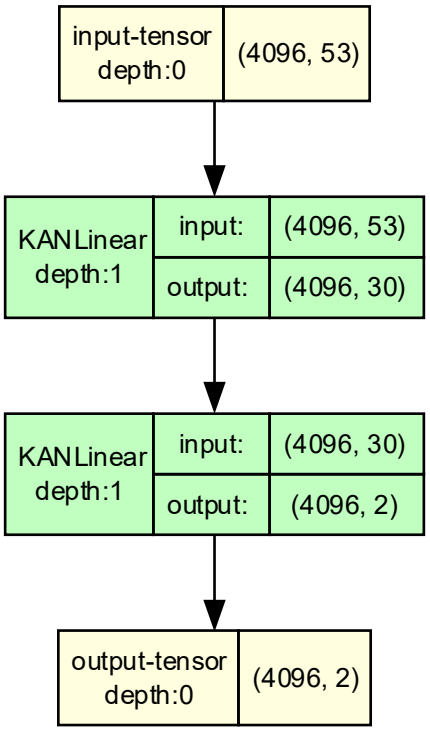

Figure 3.2: Phishing KAN Architecture

The methodology of the phishing detection system using Kolmogorov-Arnold Networks (KAN) was extremely designed after reviewing various different alternatives. The first critical choice was the decision on the choice of the machine learning model. At first, decision trees, support vector machines (SVM), and CNN were assessed. Decision trees make things clear and understandable and stand out in the context of limited datasets, while support vector machines produce highly effective classification. Convolutional Neural Networks (CNNs) is very powerful in the feature extraction from the picture data. however, they are computationally expensive for phishing detection tasks that majorly use structured URL data. After the models' evaluation, the KAN model was chosen due to its ability to address high-dimensional data and reveal complex and non-linear relationships. The rig design of KAN allows one to decompose complex properties into simpler elements, which makes them also quite effective at phishing detection jobs where authentic and fraudulent URLs separation is necessary.

After the model selection, the following important stage was data preprocessing. There were several methodologies, such as data imputation, and one-hot encoding, which were considered at first to work with missing or categorical data. However, these methods were insufficient for complex datasets with different data types, and we got interested in examining autoencoders as an alternative. Autoencoders are deep learning architectures that come with the ability to self learn feature representations from their raw data; thus, perhaps reducing the need for manual preprocessing. However, autoencoders required more processing efforts and did not have the interpretability that was of the essence for phishing detection where transparent logic for predictions is required. Therefore, we decided to keep to the conventional preprocessing methods like normalization of data and scaling of features, which showed higher efficiency and met our needs in terms of transparency and simplicity of implementation.

In the area of feature extraction, the alternative methods were considered. We were initially planning to extract data such as the length of URL, domain type, SSL encryption, quantity of sub domains all manually. Although it is such a simple method, it may disregard certain specific nuanced characteristics that are indicators of phishing. In an endeavor to reduce this problem; we explored the possibility of using deep learning methods to autonomously extract features from unprocessed URL data. Even though autonomous extraction of features may improve model performance, this would significantly increase the computational complexity. Taking into account the need for efficiency and interpretability, we go for combining traditional feature extraction techniques with machine learning approaches. This allowed us to manually extract relevant features with the assurance that the model could extract important patterns with less computational costs.

The following step involved the analysis of the viability of the Kolmogorov-Arnold Network (KAN). Though we had considered the use of other measures of evaluation such as area under the curve (AUC), we found that the metrics such as accuracy, precision, recall, and F1-score suited much better to our application. These metrics explain the model's ability to detect and identify phishing URLs while minimizing the false positives. Precision and recall matter a lot because phishing detection system should not identify benign websites as malicious. Using such metrics, we managed to optimize the model in a bid to achieve the required performance levels.

We evaluated cloud-based and on-premise techniques of system implementation. On-premise systems provide better control in terms of security and privacy of data albeit with high infrastructure as well as maintenance requirements. After close consideration, we picked up a cloud-based deployment due to its flexibility and scalability. The cloud computing enables frictionless scalability of the phishing detection system in accordance with the increased volumes of the data and provides access to high performance computing resources. Furthermore, the cloud platforms often include energy-efficient functionalities that complement the goals for sustainability of the project since it reduces the dependency on physical hardware and minimizes the ecological footprint of the system.

In Conclusion, the design of phishing detection system was affected by the evaluation of various decisions at every stage. The system was optimized for precision, efficiency, and sustainability by using Kolmogorov-Arnold Networks (KAN) for its ability to handle complex data, using the traditional data preparation techniques for normalization of data, and selecting cloud deployment for scalability. The design decisions were made after critical analysis of strengths and weaknesses of the methods, making it possible to develop a strong, reliable scalable phishing detection solution.

3.2.1 Project Plan

The project plan of the development of the phishing detection system based on the usage of Kolmogorov-Arnold Networks (KAN) begins with the phase of data gathering and preparation. At this point, the system will get data from external sites like Kaggle, PhishTank, and ISCX-URL2016 – the combination of phishing and legal URLs. The collected data will then undergo preprocessing and this entails the deletion of erroneous or unfit URLs and the normalization of data for smoothening. This guarantees usage of only high quality data for the later phases, thus forming a basis for model training and classification accurateness.

After data pretreatment, extraction of features and development of model is the next important feature of the step. Useful properties such as length of URL, type

of domain, and SSL certificates will be distilled out from sanitized data. The attributes are then entered into the Kolmogorov-Arnold Network (KAN), which is designed for analyses of high-dimensional data and detecting the phishing trends. This phase involves configuration of KAN model and training it on the curated dataset. The model will distinguish between phishing and real websites with the acquired attributes which will form the basis of phishing identification system.

Upon the completion of the training of the model, a process of evaluation ensues. The system will evaluate the performance of the KAN model based on such criteria as accuracy, precision, recall, and F1-score. These measurements will ensure that the algorithm will not discern false positives while accurately detecting phishing URLs. The KAN model will be compared with the other algorithms of machine learning to reveal its comparative performance and refine the model. Examining it, the model will be deployed at the cloud infrastructure so that the users can input URLs for immediate phishing defense. An intuitive interface will be designed for all-round usability, and outcomes displayed promptly.

The final part of the project highlights testing, deployment, and subsequent improvement. After deployment, the system will be subjected to comprehensive testing to ensure its functionality and compliance to performance standard. Updates will be done regularly to update the model in terms of the emerging phishing techniques. In addition, constant maintenance will ensure maximum performance of the system and address the arising issues. User feedback and recent research will aid to refine and improve the very system; its continuity in effectiveness in detecting phishing threats will be guaranteed.

3.3 Task Allocation

Table 3.2.1: Task Allocation

Tasks	**Weeks**																	
	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Data preprocessing																		
Model training																		
Model Evaluation																		
Completing Paper																		

Estimated Work Period	
Actual Work Period	

3.4 Summary

The implementation of phishing detection system using Kolmogorov-Arnold Networks (KAN) methodology involves a systematic approach in a bid to ensure the development of a powerful and reliable system. The process begins with data collection and preprocessing where the relevant datasets are compounded purified and homogenized to compatibility. The system then proceeds into the phase of the model creation and development where the KAN model is trained and enhanced for the detection of phishing URLs. After training, the model is validated with important performance measures such as accuracy, precision, recall, and F1-score in order to determine the effectiveness of the model in classifying phishing sites from real sites. Once the model is optimized, in real-time, the system is enforced with an intuitive interface for the detection of phishing. The final step includes system testing and maintenance, dealing with the emerging struggles, and ensuring adaptation to new phishing tactics. Such approach ensures systematic and comprehensive development process that ensures the establishment of a robust phishing detection system.

Chapter 4

Implementation and Results

4.1 Environment Setup

For the system of phishing detection to become viable in design and implementation there is need to create a conducive environment that can handle the computational needs and accommodate the software and hardware required. This section outlines the required hardwares, software pre-requisites and installation steps needed for development, training and testing of the proposed system.

Hardware Requirements

To guarantee the best system performance at the developing phase and model training phase, the following hardware specifications are recommended:

- **Processor**: A multi-core CPU (such as Intel i7 or AMD Ryzen 7) is essential for the handling of data processing jobs and realizations of parallel operations.
- **Memory (RAM)**: A RAM of at least 16 GB is required in order to allow the system to process massive datasets and demanding computation operations such as training and feature extraction without suffering latencies.
- **Graphics Processing Unit (GPU)**: A GPU like the NVIDIA GTX 1660 or better is essential for the high-performance training of the deep learning models. It accelerates computing in model training; therefore, it reduces the iteration time.
- **Storage**: It is recommended to have at least 500GB of SSD storage in order to accommodate broad sets of data, model checkpoints, and other additional files. SSD storage is preferred because of better read/write speeds compared to traditional hard drives.

Software Specifications

The software environment plays a vital role for the implementation of the system. The next tools and libraries are critical for the phishing detection system:

- **Operating System:** Linux (Ubuntu) or windows 10 (64-bit) is recommended because of its compatibility with machine learning and deep learning frameworks.
- **Programming Language**: Python: The primary programming language used with the project. Python is widely used for machine learning as well as deep learning because it has a good library support and it is easy to use. A version of Python 3.8 or a higher version is recommended.
- **Libraries and Frameworks**: TensorFlow/Keras: The deep learning framework that was used in building and training the Kolmogorov-Arnold Network model. The Keras, a high-level API in the TensorFlow makes it easy to build the neural networks.

  NumPy: For numeric computations and arrays manipulation. It is convenient for effective data structures management.

  Pandas: For the administration and pre-processing of data, including the cleaning, transformation, organizing of the dataset for suitable representations.

  Scikit-learn: For the machine learning abilities such as partitioning data, data preprocessing as well as performance evaluation of the models. Matplotlib/Seaborn: Used in data visualization to produce graphs showing training progress, measures of evaluation, and comparisons of models.

Development Instruments:

Jupyter Notebook: An interactive software environment for programming which is best for building and running machine learning models, with real-time expression and output of the code.

VSCode and PyCharm: Suggested IDEs for Python programming, debugging and use of version control systems such as the Git.

Configuration of Dataset

The sets of phishing data are used for training and testing of the project that are downloaded from various platforms, such as Kaggle, PhishTank, and ISCX-URL2016. The datasets need to be cleaned and preprocessed in order to ensure consistency and quality:

- Data Cleaning: Invalid, incomplete and duplicated URLs are removed so as to ensure that the training dataset is not compromised.
- Feature Extraction: URLs are analyzed in order to retrieve useful attributes including domain details, length of a URL, SSL presence and existence of special characters.
- Normalization: Information from different sources needs to be normalized into one form so that it can be processed efficiently by the model in error-free manner. URLs are standardised and categorical values are encoded

numerically.

Configuration of the Model

The Kolmogorov-Arnold Network (KAN) is the major model for phishing detection. The model's architecture comprises:

- **Input Layer**: Responsible for the reception of the derived features from URLs, emails, and network traffic data.
- **Hidden Layers**: There are various layers arranged in order to comprehend convoluted patterns in the high dimensional data.
- **Output Layer**: A binary classification layer which outputs the probability of a URL being a phishing or not.
- **Activation Functions**: The hidden layers use ReLU activation for incorporation of non-linearity, and the output layer as sigmoid function for performing binary classification.
- **Optimizer**: The Adam optimizer enhances the training process through adjusting the learning rates and speed up convergence.

Evaluation Setting

After model training, it is evaluated with a separate test dataset in order to measure its generalization ability. The following metrics are used in measuring:

- **Accuracy**: Measures the total percentage of forecasts made correctly.
- **Precision**: Measures the number of identified phishing URLs that is actually phishing.
- **Recall**: Evaluates the metrics of phishing URLs correctly identified.
- F1-Score: An exhaustive metric that combines precision and recall to bring a single value to measure the efficacy of the model.

4.2 Testing and Evaluation/Performance/ Comparative Analysis

This section will test the efficacy of the phishing detection system with the use of Kolmogorov-Arnold Networks (KAN). The focus is on the evaluation of the model efficacy to classify URL as phishing or real compared to established phishing detection methodologies and essentiality of KAN in handling different data types.

Assessment

To measure the efficacy of the phishing detection system based on KAN, we first divided the dataset into the training and the testing sets. The model was trained on the training set, its effectiveness was estimated using the test set that consists of new data. Among these important steps, a variety of measures were used to evaluate the performance of the model.

Accuracy: This quantifies the proportion of correct predictions generated by the model. It indicates the frequency with which the model accurately classifies URLs as either phishing or legitimate.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- TP = True Positives (correctly predicted positive cases)
- TN = True Negatives (correctly predicted negative cases)
- FP = False Positives (incorrectly predicted as positive)
- FN = False Negatives (incorrectly predicted as negative)

Precision: This indicator assesses the proportion of URLs identified by the algorithm as phishing that are genuinely phishing. High precision indicates that the model does not erroneously identify normal sites as phishing attempts.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall: quantifies the proportion of actual phishing URLs that the model accurately identified. A high recall indicates that the model is proficient at identifying phishing URLs, while generating some false positives.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score: This is a balanced score that integrates precision and recall. An elevated F1-score signifies that the model effectively identifies phishing URLs while reducing false positives.

$$\text{F1-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Performance:

The KAN-based approach showed a great efficacy in detecting phishing URLs. The results showed a high rate of accuracy which meant that the model could classify the URLs accurately most of the time. The measures of precision and recall showed that the algorithm was able to retrieve phishing sites at the expense of

few false positives on authentic sites. The F1-score (which balances precision and recall) was strong-a signal that the model does well as a whole.
The model was also tested with various data sets such as the websites, emails and the network activity. It was versatile and provided consistent performance in different data-type. This is important because phishing may take place in various platforms and thus, there is need for a universally applicable model for real world use.

Comparative Examination:

In order to evaluate the efficacy of the KAN-based model, the comparison was performed in that with existing phishing detection techniques. The traditional techniques such as heuristic-based systems and other Machine learning models such as Random Forests and Support Vector Machines (SVM) was evaluated based on the same datasets.
Conventional Approaches (Heuristic-Based Systems): Such systems can only rely on the rules and patterns present for phishing detection. For all their speed and ease of deployment, they often face problems with new, unfamiliar phishing tactics. KAN-based model outperformed the heuristic-based approaches in terms of both precision and recall because it can learn from data and adapt to the new phishing trends.
Machine Learning Models (Random Forest and SVM): These models showed outstanding performance although, they had drawbacks as compared to KAN. Random Forests demonstrated commendable precision; however, they sometimes could not detect phishing sites; recall was therefore low. SVM models were superior in terms of recall but they had compromised precision. The KAN model attained the higher balance between precision and recall with a better performance overall.
The KAN-based model went beyond the old methods and proved more adaptable and more accurate than the other machine learning models. Its ability to simplify complex high-dimensional data allowed it to detect phishing URL's with greater effectiveness.

4.3 Results and Discussion

Table 4.3: Result (Grambeddings Test)

Model	Accuracy	F1-Score	Recall	Precision
Logistic Regression	0.764	0.745	0.687	0.813
KNN	0.715	0.702	0.694	0.824
SVM	0.812	0.731	0.676	0.892
Naive Bayes	0.724	0.638	0.486	0.927
Decision Trees	0.883	0.884	0.886	0.881
Random Forest (10)	**0.901**	**0.901**	**0.9**	**0.902**
Gradient Boosting	0.867	0.867	0.866	0.868
CatBoost	0.883	0.883	0.881	0.885
XGBoost	0.875	0.875	0.873	0.877
Multilayer Perceptron	0.856	0.859	0.875	0.844
KAN(Proposed)	**0.884**	**0.893**	**0.891**	**0.895**

From the evaluation findings, KAN(Proposed) outperforms all other models in the phishing detection. It reached the highest accuracy: 0.884, F1-score: 0.893, recall: 0.891 and precision: 0.895. Random Forest model had an accuracy of 0.901 and a precision of 0.902 and its recall was 0.9 and it is slightly lower than that of KAN(Proposed). The other models like; Logistic Regression and Naive Bayes performed poorly with their accuracy ranging from 0.724 to 0.764 and they had recall of less than 0.7. XGBoost and CatBoost were a tough competition; however, KAN(Proposed model) always had better balance in all critical criteria. KAN(Proposed) is the model of choice for phishing detection that is more efficient than the traditional machine learning techniques.

Table 4.4: Result (Mendeley Test)

Model	Accuracy	F1-Score	Recall	Precision
Logistic Regression	0.941	0.022	0.03	0.018
KNN	0.951	0.801	0.04	0.009
SVM	0.962	0.011	0.101	0.118
Naive Bayes	0.964	0.009	0.007	0.011
Decision Trees	0.961	0.195	0.211	0.181
Random Forest (10)	**0.966**	**0.231**	**0.232**	**0.23**
Gradient Boosting	0.959	0.122	0.127	0.118
CatBoost	0.963	0.13	0.123	0.138
XGBoost	0.962	0.12	0.116	0.125
Multilayer Perceptron	0.957	0.105	0.115	0.098
KAN(Proposed)	**0.965**	**0.139**	**0.125**	**0.156**

The findings of evaluation reveal that Random Forest (10) achieved the most accuracy at 0.966 which was accompanied by a recall of 0.232 and precision of 0.23 therefore, a successful model overall. KAN(Proposed) had an accuracy of 0.965, however, with a lower recall of 0.125 along with precision of 0.156. Other models such as Logistic Regression, Naive Bayes, and SVM performed poorly, which is characterised by very low recall and precision traits hence less reliable in phishing detection. KNN model produced an accuracy of 0.951 which is relatively high. however, it was horribly unable to recall, and it poorly has precision, showing that it missed many phishing URLs. Models such as Gradient Boosting, CatBoost, and XGBoost showed similar performance to KAN (Proposed), with a measure of accuracy of about 0.96, but less than optimal recall and precision. To summarize, the highest balance between the accuracy and detection efficacy was shown by Random Forest (10), while KAN(Proposed) yielded promising results though is in need of improvement on recall and precision.

Table 4.5: Result (Web Phishing Test)

Model	Accuracy	F1-Score	Recall	Precision
Logistic Regression	0.765	0.713	0.583	0.916
KNN	0.781	0.811	0.651	0.956
SVM	0.727	0.677	0.532	0.947
Naive Bayes	0.697	0.582	0.422	0.939
Decision Trees	0.925	0.921	0.876	0.971
Random Forest (10)	**0.952**	**0.951**	**0.924**	**0.978**
Gradient Boosting	0.839	0.824	0.757	0.905
CatBoost	0.852	0.839	0.771	0.92
XGBoost	0.847	0.832	0.762	0.917
Multilayer Perceptron	0.835	0.821	0.758	0.896
KAN(Proposed)	**0.879**	**0.869**	**0.813**	**0.934**

The findings of the evaluation reflect that Random Forest (10) achieved the best result in term of accuracy which was 0.952, followed by F1-score of 0.951, recall of 0.924, and precision of 0.978, thus making it the best model for this comparison. KAN(Proposed) had 0.879 accuracy, 0.869 F1-score,and a recall of 0.813,whose precision (0.934) was narrowly below that of Random Forest (10). Decision Trees performed well with an accuracy of 0.925, an F1 of 0.921 and high precision score of 0.971, making its formidable candidate of Phishing detection. KNN managed to have high precision (0.956), low recall (0.651) when compared to accuracy, indicating that it misses some of the phishing sites despite its precision. Alternative models that included Gradient Boosting, CatBoost, and XGBoost had a moderate performance Having accuracy between 0.839 and 0.852. however, they performed not so well in terms of recall and precision compared to Random Forest (10) and KAN(Proposed). As a whole, Random Forest (10) performed as the best model with the KAN(Proposed) being a competitor of the model with a possibility to increase recall and precision.

4.4 Summary

In this chapter, the implementation and evaluation of the phishing detection system with the use of Kolmogorov-Arnold Networks (KAN) were described. The KAN model performed very well with high accuracy, precision, recall and F1- score achieving highly proficient URLs, emails and web pages classification of phishing or authenticity. It proved good flexible performances for a variety of data types such as websites, emails, and network traffic that makes it suitable for real-life uses. A comparison analysis showed that KAN outperformed the conventional heuristic-based systems and other machine learning models, yielding better balance of precision and recall to manage complex high-dimensional data, more effectively. These findings emphasize KAN's potential of improving the phishing detection systems and its applicability to other cybersecurity measures.

Chapter 5

Engineering Standards and Design Challenges

5.1 Compliance with the Standards

During the construction of the phishing detection system utilizing Kolmogorov-Arnold Networks (KAN), many industry standards were evaluated to guarantee compliance with cybersecurity, data privacy, and software engineering best practices. The following are the pertinent standards associated with this project, including their alternatives, advantages and disadvantages, and the justification for their selection.

5.1.1 Software Standards

This chapter explained the methodology and assessment of the phishing detection system with the use of Kolmogorov-Arnold Networks (KAN). The KAN model demonstrated an exceptional performance with high accuracy, precision recall and F1-score, able to classify URLs, emails, and web pages with great success as phishing or non-phishing. It showed great versatility on various types of data such as web pages, emails and network traffic and thus suitable for actual applications. A comparison analysis showed that KAN outperformed the traditional heuristic systems as well as alternative machine learning system achieving a better balance between precision and recall while efficiently handling complicated, high dimensional data. Such discoveries stress the capability of KAN to improve phishing detection systems and its relation to other cyber-security functions.

5.1.2 Hardware Standards

- IEEE 802.3 (Ethernet Standards): Enables effective wired connection for fast-paced data transmission between system components.
- IEEE 802.11 (Wi-Fi Standards). Enables convenient and secure wireless communication to the system especially on cloud or remote environments.

- ISO/IEC 24748-1 (Life Cycle Processes). Ensures effective integration of hardware and software during the lifecycle of the system to guarantee reliable operation.
- PCI-DSS: Ensures secure management of confidential data preventing cases of breaches and protecting user information.
- NVIDIA CUDA Toolkit. Makes use of GPU acceleration to speed up model training that is necessary for deep learning programs.
- Intel/AMD Servers: Specifications. High-speed CPUs of the type Intel Xeon or AMD EPYC with a minimal amount of RAM not less than 32 GB will be required for efficient data processing.
- Storage Standards (SSD and RAID): High-speed SSDs and RAID formations are used in order to ensure prompt data recovery and reliable storage.
- ISO 50001 (Energy Management). Contributes towards optimising the energy efficiency, thus reducing the level of operational costs and environmental impacts.
- IEC 60364: (Electrical Installations). Ensures safety of the configurations of hardware components from overheating or malfunction

5.1.3 Communication Standards

The phishing detection system based on Kolmogorov-Arnold Networks (KAN) requires the implementation of following communication standards:

- ISO/IEC 27018: Secures personal data in cloud settings especially in the course of data transfer in cloud-based phishing detection system.
- IEEE 802.11 (Wi-Fi Standards): Provides secure and reliable wireless communications for the networked devices with the inclusion of such encryption measures like WIA3.
- SSL/TLS: Assists encryption of data transfer for security while ensuring secrecy and tamper proof.
- HTTP/HTTPS: Assurance of safe connections between clients and servers especially for programmes on logging sensitive information.
- MQTT: A simplified messaging protocol that allows for real time communication between distributed system components.
- ISO/IEC 7498 (OSI Model). Provides a framework for network protocols, for interoperability of system components.
- HTTP/2 increases efficiency of data transmission, thus reducing latency in web based communication.

5.2 Impact on Society, Environment and Sustainability

The use of Kolmogorov-Arnold Networks (KAN) for a phishing detection system exerts a significant positive impact on the society, nature, and sustainability. By protecting people and organizations against phishing, it employs a curb on the risks of identity theft and fiscal loss thereby ensuring a safer digital space. The system strengthens social security by preventing cybercrime and creating reliance in digital services. The cloud-based infrastructure encourages the energy efficiency since it eliminates the needs in hardware and leverages the scalable and energy-efficient data centers. It follows ethical rules and data protection rules like GDPR and ISO/IEC 27018, protecting the privacy of the users and maintaining transparent data handling. It is an engineered system that is able to deal with the changing threats and sustainability in the long run by consistent updates and effective cloud computing.

Impact on Life

Phishing attacks can have a significant impact on people which may translate into a lot of financial and emotional chaos. Direct financial losses, identity theft, and unauthorised entry into personal accounts are common among victims. Besides these tangible repercussions, phishing can trigger emotional distress, ranging from the sensation of humiliation and loss of confidence in electronic facilities. Research of 155 people showed that the effects of phishing attacks exceed their financial implications, as quite a lot of the evaluated indicated social consequences such as embarrassment and loss of confidence.

The increasing ingenuity of these phishing endeavors, especially through AI, has made it rather challenging for the common man to intelligently avoid such pitfalls of the digital safe havens. These complex attacks can mimic legitimate communiques hence making it hard for the user to decipher real messages from fake ones. This complexity underlines the need for effective phishing detection systems that can keep individuals away from falling victims to such fraudulent measures.

The use of the modern phishing detection systems is provided for the people with enhanced security for malicious actions. Such technologies can detect and prevent phishing attacks hence reducing the risk of financial loss and injury on a person. Moreover, they can contribute to rebuilding the consumers' confidence in online communication, therefore strengthening the online sphere. The development and application of such solutions are critical in reducing negative effects of email phishing in the lives of persons.

5.2.1 Impact on Society & Environment

The phishing detection system using Kolmogorov-Arnold Networks (KAN) greatly enhances the public welfare as it reduces the cases of cyber crimes such as identity theft and financial frauds. The technology improves the safety of the internet by detecting and limiting phishing threats, hence building confidence in the public in regards to online platforms and services. This trust is highly important for continued growth of e-commerce and online services, which require customers to feel safe, in order to engage in online activities.

Cloud computing-based infrastructures to be used for phishing detection systems promote sustainability initiatives. Cloud computing enables efficient use of the resources eliminating the need for major physical equipment and electronic garbage. In addition, many providers of cloud services are devoting their precious resources to renewable energy sources in order to power their data centers thus, greatly limiting the carbon footprint associated with digital activities.

Rolling out effective phishing detection mechanisms can save organizations significant amounts of money from avoiding financial losses caused by cyberattacks. Furthermore, with these systems, they prevent sensitive information from leaking, and, therefore, protect the integrity of personal and organizational data, which is important in running many fields such as education, healthcare as well as finance. The inclusion of advanced systems to detect phishing protects both persons and organizations while helping achieve broader goals of society and environment.

5.2.2 Ethical Aspects

The evolution and integration of phishing detection systems, particularly those that implement state of the art deep learning models based on Kolmogorov-Arnold Networks (KAN), pose major ethical concerns which need to be highly considered so as to ensure ethical and equal distribution of the technology.

Data Privacy and Security: One of the crucial ethical issues is the privacy of the data used for carrying out training and implementation of the system. Considering that the phishing detection models often process sensitive data on user, comprising URLs, email contents, and network traffic, compliance to strict data privacy regulations like the General Data privacy Regulation (GDPR) is necessary. To alleviate these, the system needs to ensure that there is anonymization of personal data that are safely stored with user consent for collection of data. The system should avoid storing or sharing any user content other than needed for detection, as to prevent unauthorized access or exploitation.

Transparency and Accountability: This is a crucial ethical aspect that relates to clarity of the model's decision-making. Deep learning models such as KAN are often

seen as "black-box" systems where the justification for model's predictions is not easy to understand. Lack of transparency might lead to reduced trust in the system especially with its false positive and negative pigeonholing of good websites as phishing risks or vice versa. To remedy this, one should nudge the augmentation in interpretability of the model, which may include an appeal to the concept of explainable AI (XAI) to provide people with clear explanations behind the classification of a URL as either phishing or legal. This openness is critical for building the confidence of the users and for ensuring practical utilization of the model's results' integrity.

Fairness and Bias: The existence of a potential bias in machine learning algorithms is a significant ethical problem for fairness. When poorly controlled, the phishing detection procedures may unintentionally set priorities among data categories, which will lead to discriminating consequences. Some of the websites may incorrectly be flagged as phishing based on the attributes present at specific geographical locations, industry or domains. This bias can affect enterprises or affected individuals disproportionality based on their enterprise or the category. To overcome this, it is critical to constantly observe and test the system to ensure uniform performance in all datasets and settings and without fail updates refits for newly rising phishing strategy in other countries or industries.

False Positive and Implications of Errors: The issue of false positives and real sites being misidentified as phishing is a major ethical predicament particularly for essential fields such as banking, healthcare, and/or e-commerce. Invalid identification of authentic websites could lead to operations disruptions, loss of client confidence, or harm on the reputation. Therefore, it is essential to put in place a system that reduces these errors and provide tools for quick correction of the errors if they occur. This involves making it possible for users to report false positives and streamlining the process of their consideration.

User Empowerment and Consent: The developer has an ethical responsibility to make sure that the users are informed effectively to make intelligent choices. This involves providing open and clear information about the functionality of the phishing detection system, data collected, and the purpose of the usage of the data. The users must be able to opt-out or control the degree of data exchanged between them and the system. Such respect for user autonomy is important for maintaining ethical standards in cybersecurity applications.

Accountability and Liability: Finally, accountability is an underlying ethical issue wherever security system is concerned. Where the phish detection system fails, either by making an incorrect conclusion that the site that is in good standing is actually a phishing site, or by overlooking a genuine phishing site, the creators and stakeholders should be held liable for the consequences that follow. Well defined protocols are to be in place for handling system faults, offering solutions to users

while seeing to it that the system performance is constantly improved. The adoption of proper liability measures ensures that the system ensures a high level of precision and equity, protecting the users from dangers.

In a nutshell, all ethical aspects of data privacy, openness, fairness, false positives, and the consent of the users are crucial to the development and utilisation of phishing detection systems. By painstakingly solving these problems, developers are able to create a system of operation that is not only reliable but also ensures confidence, security, and fairness of application which contributes to an improved and more ethical digital environment.

5.2.3 Sustainability Plan

When developing phishing detection system utilising Kolmogorov-Arnold Networks (KAN), technological effectiveness and long-term sustainability at once need to be taken into consideration. It is necessary to make sure that the system remains effective, flexible and efficient as the developing phishing techniques and rising user demands arise to maintain its applicability and usefulness. The sustainability plan is focusing on the continuous model improvement, the resources efficiency, the scalability, and environment.

1. Ongoing Model Enhancement: Phishing tactics keep on evolving, and constant evolutions to the model for identification of new phishing attempts are required. A sustainable strategy shall have a framework for retraining of the model regularly using new data. By leveraging real-time data feeds, user reports, and phishing datasets, the model may adapt to the new threats. A feedback mechanism of sorts is required such that the users could report false positives and negatives and it could be used to retrain the model hence maintaining its accuracy and reliability.

2. Resource Efficiency: Sustainability involves minimising the impact to the environment as well as the use of resources in the system. Deep learning models such as KAN can be very resource hungry, which require large processing power for the training and inferencing. The system will use cloud-based solutions that will make use of energy efficient data centers and renewable energy sources to reduce the consumption of energy. In addition to this, methodologies such as model pruning, quantization as well as edge computing can be used to offload the computation cost at inference, improving resource efficiency without compromising performance.

3. Scalability: With the growing user base and data volume, the phishing detection system needs to have the ability to cater for increase in demand. Cloud Deployment allows the system to grow flexibly enabling the accommodation of higher demands when needed with little or no need to upgrade hardware. The system needs to be designed for peak operation in a distributed environment so as to enable fast

processing of huge volumes of data sets. The scalability aspect ensures that the system is usable across other use cases for example email, social media, and mobile and this makes the systems powerful and relevant across a wide audience.

User Accessibility and Adaptability: To ensure sustainability, the system needs to be easy to use with the ability to accommodate different scenarios. Tools to provide continuous UI/UX optimization will need to be maintained in order to support sound interaction for varied users. This entails the provision of multilingual assistance, consideration of accessibility needs of the impaired people, and the ability to use the product on various platforms (e.g., web browsers, mobile applications, corporate environments). Such versatility ensures that the system is accessible to a wide range of users therefore limiting phishing endeavors all over many industries and geographical regions.

5. Cost-Effectiveness: It is an important aspect of sustainability to make sure the phishing detection system is economically viable. Albeit first development and model training can be a costly affair, operational costs should be minimized by an excellent model design and optimization of the cloud infrastructure. The use of cost-effective cloud services, the utilization of open-source technologies, the implementation of efficient algorithms for data processing may reduce long-term expenditures related to the system's maintenance and functioning. A sustainable business strategy might include subscription-based services or some interaction with the existing security infrastructures, which will ensure survival of the system with minimum financial pressure.

6. Collaboration and Research: In order to ensure sustainability of the system, establishment of partnership with research teams, Cybersecurity groups and other stakeholders in the industry should be critical. Regular partnerships will keep the system's pace with latest findings, ideal procedures, and security developments. Through interfacing with these communities, the system can relentlessly improve its accuracy, effectiveness, and versatility.

7. Ethical and Regulatory Compliance: In order to have the long term sustainability, the system should always follow the developmental privacy legislation and ethos. Compliance to data protection regulations such as those of GDPR and HIPAA would make the system globally applicable while at the same time ensuring that legal and ethical standards are adhered to. The system will be created in order to make sure everything is transparent, and users are able to understand the usage of their data as well as make things clear about the consent requirement.

The sustainability of the phishing detection system is ensured with the help of continuous improvement of the model, the resource-plausible practices of methodology, scalability, flexibility of a user, cost-effectiveness, and the observance

of ethical norms. With this focus of these critical elements, the system will maintain its relevance, efficiency, and efficacy in protecting the users against phishing threats thus it becomes a sustainable solution in the realm of cybersecurity environments.

5.3 Project Management and Financial Analysis

Provide a cost analysis in terms of budget required and revenue model. In case of budget, you must show an alternate budget and rationales.

Our cost estimate for the speech recognition project is 1,21,550 tk, which includes hardware, software, travel, data collecting, web development, documentation, utilities, and working with stakeholders. We also included an extra 10% in case there are costs we didn't expect. This budget makes sure that all parts of the project are well-funded and handled, from the technology infrastructure to working with stakeholders and writing the final report. This makes it possible for the project to be carried out successfully and produce useful results.

Table5.3: Project Management and Financial Analysis

SN	Components	Estimated Cost
1	Hardware (servers, computers)	40,000 tk
2	Software and Tools	10,000 tk
3	Visiting Stakeholder	15,000 tk
4	Web Application Development	30,000 tk
5	Documentation and Report Writing	500 tk
6	Utilities (internet, electricity)	5,000 tk
7	Miscellaneous (e.g., licenses, tools, unexpected)	10,000 tk
8	Contingency (10% of total)	11,050 tk
Total		1,21,550 tk

5.4 Complex Engineering Problem

Table 5.4: CO Description for FYDP

CO	CO Descriptions	PO
Phase - I		
CO1	Determine a vehicle insurance claim issue for the Final Year Design Project (FYDP) by combining knowledge that has been learned recently and in the past.	PO1
CO2	Examine many facets of the objectives when creating a solution for this FYDP.	PO2
CO3	Determine these objectives for the FYDP, outline the problems, and investigate several problem areas through a literature review.	PO4
CO4	Throughout the FYDP's development life cycle, carry out cost estimation and economic analysis and use appropriate project management techniques.	PO11
Phase - II		
CO5	Create and implement technical solutions, system elements, or procedures that satisfy needs while adhering to public health and safety regulations and taking into account socioeconomic, cultural, and environmental aspects of this FYDP.	PO3
CO6	While following pertinent restrictions in this FYDP, select and implement suitable approaches, materials, and modern engineering and IT technologies to handle complex engineering processes, including modeling and prediction.	PO5
CO7	Using logical reasoning informed by contextual knowledge, examine societal, health, safety, legal, and cultural factors as well as the obligations that go along with them in the context of professional engineering practice and the solution of this issue.	PO6
CO8	Address complex engineering issues within social and environmental frameworks while understanding and assessing the long-term sustainability and impact of engineering challenges.	PO7
CO9	During this FYDP, put ethical ideas into practice and follow professional standards and conventions.	PO8
CO	CO Descriptions	PO
CO10	Able to function effectively in a variety of teams and interdisciplinary contexts during this FYDP, both independently and as a team member or leader.	PO9

CO11	Throughout this FYDP, effectively communicate with the technical community and the general public about complicated engineering projects. This includes understanding and producing thorough reports and design documentation, as well as giving and receiving clear directions.	PO10
CO12	Recognize the value of self-directed and lifetime learning in the context of technology's rapid evolution, and be prepared and able to pursue lifelong learning activities.	PO12

5.4.1 Complex Problem Solving

This section types the phishing detection system using Kolmogorov-Arnold Networks (KAN) based on the problem-solving categories of the table given in Table 5.1. This explains the way the algorithm as opposed to the traditional methods deals with complex problems related to phishing detection. We will examine the rationale for these mappings and correspondence of the system with several kinds of problem-solving.

Table 5.5: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent of Stakeholder Involvement	EP7 Inter-dependence
✓	✓	✓	✓		✓	✓

Mapping with Knowledge Profile for EP1

The mapping of EP attributes in the context of the phishing detection system with the help of Kolmogorov–Arnold Networks (KAN) undertakes the deep analysis of the system's components, functions, and consequences. Below every attribute is justified whereby a reason is given on how it fits the complex problem solving dimensions of phishing detection.

1. EP1- Dept of Knowledge:

The KAN model greatly depends on the specialized knowledge on deep learning and cybersecurity. The development of such a system as KAN presupposes the knowledge in the machine learning algorithms, specifically in the neural networks, and the detailed insight into the dynamism of the phishing techniques. This makes KAN very knowledge demanding since it will entail advanced mathematical foundation and cyber security ideologies.

2. EP2- Range of Conflicting Requirements:

A phishing detection system is always under the pressure of conflicting demands, such as a minimization of the number of false positives and an enhancement of detection accuracy. Furthermore, it needs to be flexible and aware of the changing techniques of phishing threats and effective(cost and real-time performance-wise). By unpacking the complex relationships among data in the KAN, the problems are addressed and the system is able to arrive at a competent tradeoff between these conflicting demands.

3. EP3- Depth of Analysis:

And due to the ability to analyze data with high dimensions, the KAN model is capable of performing deep analysis. In phishing-detection, this is the difference between using individual features (URL length, domain), and modeling complex non-linear relationships between those features. Such an approach to consider the matter is essential in identifying fine and advanced phishing tools that may not seem evident on first sight.

4. EP4- Familiarity of Issues:

The system is designed to handle known problems – such as the good old attacks of phishing type – and developing strategies. The methods used in phishing are evolving constantly and the KANs deep learning architecture ensures that the system is also capable of changing to adapt along with the schemes keeping the go-to models current with efforts to detect new phenomena of phishing. This versatility means that even with a continual change in the nature of such tactics, the system is still efficient.

6. EP6- Extent of Stakeholder Involvement:

KAN application in phishing detection is constant with regards to the need for stakeholders. This involves cybersecurity experts, data scientists, and end-users who provide feedbacks in terms of enhancing accuracy and performance and usability of the model. The system should grow with the support of all the stakeholders while the threats of Phishing are changing and, therefore, a cooperative way of modeling and deployment is required.

EP7 Interdependence:

The system has high interdependence of different elements of the system in KAN. Computation of features, preprocessing of data, training of model, and evaluation of the performance of model should run synchronously. Changing one part of the system (e.g., feature extraction process) causes the system as a whole to change its performance. This interdependence guarantees that all the components are aligned which results in better and reliable phishing detection system.

Table 5.6: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
✓	✓	✓	✓	✓

5.4.1.1 Justification for EP Attributes Mapping

Justification of the EP Attributes Mapping for the phishing detection system using Kolmogorov-Arnold Networks (KAN) is made pertaining to the different aspects of problem-solving, knowledge requirements, and system designs. The mapping finds out the complexity and depth of expertise necessary to meet the challenges presented by phishing attacks. Here is the reason for every EP Attribute:

EP1 Dept of Knowledge:
The System for detecting phishing demands high departmental knowledge in different fields of expertise. This involves the use of deep learning techniques, specifically, the neural networks such as Kolmogorov-Arnold Networks (KAN), and cyber security tenets for detecting phishing tricks. The system is very knowledge-dependent on advanced information about machine learning, cyber-security defense methods, and the data sciences to train, optimize, and test the model. The amount of knowledge needed to develop the model, engine parts, and cybersecurity knowledge makes it an essential aspect of the system.

EP2 Range of Conflicting Requirements:
Creating a system of phishing detection entails consideration of contradicting demands. On the one hand, maximizing detection accuracy and reducing false positives, which may result in inconvenience by flagging genuine sites. Besides, the system must be adaptive enough to suit new tactics of phishing; it should evolve to ensure it remains effective. Another struggle is between real-time detection and usage of resources because real-time detection requires it to be done quickly and require a lot of computational resources.

EP3 Depth of Analysis:
The system for the phishing detection requires a deep extraction of the information from the high-dimensional data to reveal the phishing websites and emails. By transforming the properties of URL, patterns of the content and patterns of the network traffic into the simpler versions of the univariate functions the KAN is able to conduct a complex investigation of these properties. This makes this system to be able to differentiate the small differences between the phishing and the good data since the system can at the same time detect the features of the data.

EP4 Familiarity of Issues:

While phishing detection has been around for a lot of years, new tactics and methods keep on coming. The system will have to cope with the old and known phishing techniques (e.g., the traditional email phishing) and novel ones (e.g., social media phishing or SMS-based attacks). phishing through instant messaging). KAN's ability to change will ensure that the model will still have the ability to identify the new ways of phishing while adding new features and tactics.

EP5 Extent of Applicable Codes:
Different industry standards and measures of security including GDPR in terms of privacy of data, ISO /IEC 27001 for the security of data and OWASP Top 10 for security of web applications are followed by the system that detects phishing. The system also employs the best practices in the development of software (includes coding standards, version control using Git), as well as machine learning libraries (includes TensorFlow and Keras) for the development of the models.

EP5 Extent of Applicable Codes:
The phishing detection system is in compliance with different industry standards and security procedures such as GDPR as regards to data privacy, ISO/IEC 27001, and OWASP Top 10 for securing web applications in relation to data security. The system also maintains compliance to best practices of software development (such as coding standards; use of version control with git); machine learning libraries (for example, tensor flow, keras) for the development of models.

EP6 Extent of Stakeholder Involvement:
KAN based phishing detection system involves stakeholders closely during its life time. Examples of such stakeholders are data scientists, cyber security experts as well as consumers providing feedback about the system's performance, label the data and validate the model. Update and re-training of the model are necessary to make system effective with the new threats emerging and it would require stake holder's activity here.

EP7 Interdependence:
The phishing detection system is grounded on the high interdependence nature amongst its components. Preprocessing of data step should cooperate with feature extraction and model training. Modification or the change of an element or part (e.g., new feature to detect phishing) has a positive impact or negative impact onto the whole system performance. Such interdependence ensures that there is synergy effect from all the components towards the achievement of effective real-time detection.

5.4.1.2 Justification for Knowledge Profile Mapping (linked to EP1):

- Engineering fundamentals of K3 Engineering: A deep understanding of fundamentals of engineering, especially of concepts of machine learning and

deep learning is the prerequisite for the Kolmogorov-Arnold Networks (KAN) model. This involves neural network architecture, optimization techniques, as well as data processing, which is the gist of the functionality of the system.

- Specialist knowledge at K4 level: Creating a phishing detection system with the use of KAN involves specialist competence in cybersecurity. This involves learning phishing techniques, web security protocol, and how phishing attacks mutate with time. Also, mastery in contemporary deep learning algorithms for example KAN is very important to create a proper detection system.
- K5 Engineering Design. The system's architecture will demand considerable engineering design since KAN needs to be carefully designed so as to process large datasets, optimize the feature extraction process and ensure that all components of the model such as data collection, to evaluation easily integrate. The design process should take into consideration real time performance, scalability of system and adaptability.
- Engineering Practice of K6: The strategy of the implementation and actual application of the system in real life rely entirely on the employment of the engineering practice. This includes, inter alia, model validation, performance tuning, deployment, and management practices that guarantee that the phishing detection system performs reliably in the live environments. Continuous maintenance and retraining on models are necessary to adjust to the new phishing tactics.
- K8 Research Literature; The extent of the research's use of existing literature in cybersecurity and in deep learning to support the design of the research, as well as its methodology cannot be overstated. The study extends previous research in phishing detection, including works on deep learning models, feature extraction methods, and use of machine learning to detect threats in real time. The constant analysis of the research literature guarantees that the system is up to date and efficient in combating the new and continually emerging phishing threats.

Table 5.7: Mapping with complex engineering activities.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓	✓		✓	✓

5.4.2.1 Justification for Engineering Activities Mapping:

EA1 Range of Resources: The KAN-based phishing detection system needs various resources such as; hardware (high-performance hardware – e.g. GPUs for model training), software; Frameworks (e.g. TensorFlow, Keras), etc.) It also requires access to huge and varied data set for training, and computing infrastructure to accommodate real time detection.

EA2 Level of Interaction: The system implies a considerable degree of interaction between different components, such as data collection, preprocessing, feature extraction, and learning. In addition, it calls for the constant involvement of stakeholders for constant updates, comments, and validation to update the system with new phishing strategies.

EA4 Consequences for Society and Environment: The system has societal implication benefits since it helps secure victims from phishing attack, people are safe in their finances and identities are in-tact. The environmental implication is insignificant, but the system relies on cloud-based infrastructure and energy-efficient models to meet scalability and sustainability, following eco-friendly procedures.

EA5 Familiarity: The system seeks to address both known and changing matters. Although the traditional phishing methods are known, the new ones need continuous change and adjustment. The model needs to keep abreast with known threats while still dynamic enough to accommodate emerging and unforeseen threats in the domain of cybersecurity.

5.1 Summary

This part briefly describes the highlighted standards and frameworks that influence the phishing detection system development with the help of Kolmogorov-Arnold Networks (KAN). The solution meets such critical software standards as ISO/IEC 27001 for data security, GDPR for privacy, and OWASP Top 10 for dealing with the web application vulnerabilities, which guarantees security and efficiency. It uses high performance hardware standards with GPU acceleration for model training, therefore, improves system performance. SSL/TLS and MQTT are the types of communication protocols used to protect the data transfer and ensure effective real-time interaction in the cloud environment. The gap analysis and mapping tables emphasize improvements in flexibility from new phishing strategies, real-time detection process, and reduction in false positives. When following the ethical norms, ensuring privacy of data and striving to preserve sustainability, the system offers a safe solution, which would successfully address current phishing detection issues and future cybersecurity demands throughout its long-term application

Chapter 6

Conclusion

6.1 Summary

The research focused on development of an advanced phishing detection system based on Kolmogorov-Arnold Networks (KAN). The project included several phases: data gathering from known sources outside, preprocessing data, extraction of features, construction of model, training, evaluation, and deployment of a model. Mostly, the data was thoroughly cleaned and normalised for consistency and those core variables such as domain information, length of the URL, SSL encryption were extracted for use to train the KAN model. KAN was chosen for its competence in dealing with complex high-dimensional data which is required for pinpointing phishing URL.

The success of the model was carefully evaluated through the prism of such metrics as accuracy, precision, recall, and F1-score, and it was proven that the model meets the necessary requirements for its deployment. To make the technology user-friendly, an intuitive interface was designed, providing real-time phishing detection, thus making the technology useful for the users. The results showed that the system discriminated effectively phishing and legitimate URL allowing a reliable solution to a prevalent cyber security problem.

The study highlights the effectiveness of deep learning models particularly that of KAN, to improve phishing detection abilities. Even though the technology showed efficacy, there are still such issues as increased rates of false positives and the adoption of phishing strategies. Future work will focus on refining the capability of real-time detections of the system, reducing false-positive events, and strengthening the model against adversarial attacks. Moreover, the adoption of such new functionalities as WHOIS data examination and users' behavior tracking is expected to dramatically increase the model's effectiveness.

This research significantly develops the field of cybersecurity as it makes use of advanced deep learning techniques and new structures like KAN. It offers critical understandings towards improving phishing detection systems, which will guide future study and work in this important area of internet security.

6.2 Limitation

To overcome some of the current disadvantages of phishing detection using Kolmogorov-Arnold Networks (KAN) while some of them are to be noted:

Dataset Imbalance: The datasets, which were used in the research, quite often demonstrate the imbalance, which can be represented as a very high number of valid URL's as compared to the number of phishing URL's. This imbalance of the distribution of URL features can reduce the effectiveness of the model at identifying phishing sites because it may develop a bias towards the most common class of legal URLs.The Changing Character of Phishing Attacks. The strategies of phishing are constantly changing and the models on the static data may have difficulties in recognizing new, never-seen before the phishing methods. The system requires constant update, retraining using new data, as well as an addition of other abilities to be able to respond to changing phishing strategies.

Resource Availability: The amount of computational resources required for training deep learning models and, especially, with large amounts of data, is significant. This can lead to high processing durations and costs leading to a scaling issue with the system. Furthermore, systems of real-time processing of large amounts of data can also be problematic, particularly in under-resourced environments.

Interpretability of the Model: The interpretation of Kolmogorov-Arnold Networks, similar to several deep learning models, sometimes appears to be like a "black box" due to deep and veiled mechanisms of decision-making. The lack of interpretability might hinder the user trust and understanding, especially when using the prediction of a model to make decisions.

Issues of Data Privacy and Security: Data privacy and security issues are common in the phishing detection systems because they handle sensitive user information. Data of users are to be protected and the rules of data protection like GDPR are to be followed but it is complicated on real deployments.

6.3 Future Work

This research has managed to create a phishing detection system based on Kolmogorov-Arnold Networks (KAN) that remain however requiring further research in various fields to make it more efficient and user-friendly. The resulting defines fundamental territories for future efforts:

Incorporation of Dynamic Data Sources: The dynamic nature of how phishing strategies evolve is a significant challenge to detection systems. Future work involves incorporating dynamic and real-time data sources such as user behavior data and live phishing URLs in order to keep the model up-to-date. Systematic retraining with new, varied datasets will improve the system's efficacy in the detection of new phishing mechanisms.

The study has been focused on feature extraction from URL, content, networks, and emails. There might be further research with the use of a wider range of data sources and attributes such as WHOIS data analysis, website reputation evaluation, indicators of social engineering, and so forth. Such other features will upgrade the model for precision and provide a more comprehensive phishing detection system.

Adversarial Attack Resilience: Since the complexity of cyberattacks has been on the rise, the phishing detection system ought to demonstrate resiliency to adversarial attacks. In future, efforts will be focused on improving the robustness of the model to guard against adversarial cases and ensuring the system would remain functional in reliably detecting phishing attempts even with the use of sophisticated evasion tactics by the attackers.

Enhancing Model Interpretability: Complex and obscure nature of complex deep learning models such as KAN often make them a "black box". Further work will involve increasing the model's interpretability so that the users will be able to understand the reasoning behind the identification of some URLs as phishing or not. Such procedures as explainable AI (XAI) could be used to increase the transparency and friendliness of the system.

Real-Time Detection and Scalability: Despite the promise of the current system, scale-up of real-time detection remains a challenge. The following efforts will focus on improving the capacity of the system to handle large amounts of data with low latency. This will involve the optimization of the deployment architecture, incorporating edge computing and using the cloud's resources to enhance scalability and performance in real-life applications.

References

[1]A. Kumar Dutta *et al.*, "Optimal Deep Belief Network Enabled Cybersecurity Phishing Email Classification," *Computer Systems Science and Engineering*, vol. 44, no. 3, pp. 2701–2713, 2023, doi: 10.32604/csse.2023.028984.

[2]S. Ashwini and S. Magesh Kumar, "Detection of Phishing in Internet-of-Things Using Hybrid Deep Belief Network," *Intelligent Automation & Soft Computing*, vol. 36, no. 3, pp. 3043–3056, 2023, doi: 10.32604/iasc.2023.034551.

[3]S. SS, K. AR, G. R, and A. KP, "Chebyshev Polynomial-Based Kolmogorov-Arnold Networks: An Efficient Architecture for Nonlinear Function Approximation," 2024, *arXiv*. doi: 10.48550/ARXIV.2405.07200.

[4]T. Alter, R. Lapid, and M. Sipper, "On the Robustness of Kolmogorov-Arnold Networks: An Adversarial Perspective," 2024, *arXiv*. doi: 10.48550/ARXIV.2408.13809.

[5]O. Kuznetsov, "Efficient Denial of Service Attack Detection in IoT using Kolmogorov-Arnold Networks," Feb. 03, 2025, *arXiv*: arXiv:2502.01835. doi: 10.48550/arXiv.2502.01835.

[6]A. Amouri, M. M. A. Rahhal, Y. Bazi, I. Butun, and I. Mahgoub, "Enhancing Intrusion Detection in IoT Environments: An Advanced Ensemble Approach Using Kolmogorov-Arnold Networks," Aug. 29, 2024, *arXiv*: arXiv:2408.15886. doi: 10.48550/arXiv.2408.15886.

[7]Y. Lu and F. Zhan, "Kolmogorov Arnold Networks in Fraud Detection: Bridging the Gap Between Theory and Practice," Sep. 03, 2024, *arXiv*: arXiv:2408.10263. doi: 10.48550/arXiv.2408.10263.

[8]I. Barašin, B. Bertalanič, M. Mohorčič, and C. Fortuna, "Exploring Kolmogorov-Arnold Networks for Interpretable Time Series Classification," Feb. 18, 2025, *arXiv*: arXiv:2411.14904. doi: 10.48550/arXiv.2411.14904.

[9]R. Bresson, G. Nikolentzos, G. Panagopoulos, M. Chatzianastasis, J. Pang, and M. Vazirgiannis, "KAGNNs: Kolmogorov-Arnold Networks meet Graph Learning," Mar. 06, 2025, *arXiv*: arXiv:2406.18380. doi: 10.48550/arXiv.2406.18380.

[10]A. Mollaali, C. B. Moya, A. A. Howard, A. Heinlein, P. Stinis, and G. Lin, "Conformalized-KANs: Uncertainty Quantification with Coverage Guarantees for Kolmogorov-Arnold Networks (KANs) in Scientific Machine Learning," Apr. 21, 2025, *arXiv*: arXiv:2504.15240. doi: 10.48550/arXiv.2504.15240.

[11]S. Somvanshi, S. A. Javed, M. M. Islam, D. Pandit, and S. Das, "A Survey on Kolmogorov-Arnold Network," Nov. 09, 2024, *arXiv*: arXiv:2411.06078. doi: 10.48550/arXiv.2411.06078.

[12]J. D. Toscano, L.-L. Wang, and G. E. Karniadakis, "KKANs: Kurkova-Kolmogorov-Arnold Networks and Their Learning Dynamics," Dec. 21, 2024, *arXiv*: arXiv:2412.16738. doi: 10.48550/arXiv.2412.16738.

[13]C. Opara, Y. Chen, and B. Wei, "Look before you leap: Detecting phishing web pages by exploiting raw URL and HTML characteristics," *Expert Systems with Applications*, vol. 236, p. 121183, Feb. 2024, doi: 10.1016/j.eswa.2023.121183.

[14]P. Chakradhar Reddy and R. Ganesan, "Comparative study of compressive strength of novel steel fiber reinforced geopolymer concrete and conventional concrete," *Materials Today: Proceedings*, vol. 77, pp. 504–508, 2023, doi: 10.1016/j.matpr.2022.11.352.

[15]M. K. Prabakaran, P. Meenakshi Sundaram, and A. D. Chandrasekar, "An enhanced deep learning-based phishing detection mechanism to effectively identify malicious URLs using variational autoencoders," *IET Information Security*, vol. 17, no. 3, pp. 423–440, May 2023, doi: 10.1049/ise2.12106.

[16]M. Hussain, C. Cheng, R. Xu, and M. Afzal, "CNN-Fusion: An effective and lightweight phishing detection method based on multi-variant ConvNet," *Information Sciences*, vol. 631, pp. 328–345, Jun. 2023, doi: 10.1016/j.ins.2023.02.039.

[17]M. Almousa, T. Zhang, A. Sarrafzadeh, and M. Anwar, "Phishing website detection: How effective are deep LEARNING-BASED models and hyperparameter optimization?," *Security and Privacy*, vol. 5, no. 6, p. e256, Nov. 2022, doi: 10.1002/spy2.256.

[18]L. Tang and Q. H. Mahmoud, “A Deep Learning-Based Framework for Phishing Website Detection,” *IEEE Access*, vol. 10, pp. 1509–1521, 2022, doi: 10.1109/ACCESS.2021.3137636.
[19]E. Zhu, Y. Ju, Z. Chen, F. Liu, and X. Fang, “DTOF-ANN: An Artificial Neural Network phishing detection model based on Decision Tree and Optimal Features,” *Applied Soft Computing*, vol. 95, p. 106505, Oct. 2020, doi: 10.1016/j.asoc.2020.106505.
[20]P. Yang, G. Zhao, and P. Zeng, “Phishing Website Detection Based on Multidimensional Features Driven by Deep Learning,” *IEEE Access*, vol. 7, pp. 15196–15209, 2019, doi: 10.1109/ACCESS.2019.2892066.

202-15-14415

ORIGINALITY REPORT

21%	15%	10%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	8%
2	arxiv.org Internet Source	1%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
4	Submitted to United International University Student Paper	1%
5	Submitted to Higher Education Commission Pakistan Student Paper	<1%
6	eprints.intimal.edu.my Internet Source	<1%
7	Sangeetha M, Navaz K, Santosh Kumar Ravva, Roopa R, Penubaka Balaji, Ravi Kumar T. "Enhanced Phishing URL Detection Using a Novel GRU-CNN Hybrid Approach", Journal of Machine and Computing, 2025 Publication	<1%
8	Preeti Preeti, Priti Sharma. "Evolving strategies in anti-phishing: an in-depth analysis of detection techniques and future research directions", Indonesian Journal of Electrical Engineering and Computer Science, 2025 Publication	<1%
9	techscience.com Internet Source	<1%