**PREDICTION OF UNIVERSITY NETWORK TRAFFIC BEHAVIOUR USING DEEP LEARNING AND OPTIMIZATION TECHNIQUE**

CHAPTER ONE

**INTRODUCTION**

**1.1 Background to the Research**

In recent years, universities worldwide have experienced a significant digital transformation, with the integration of advanced technologies into various aspects of academic and administrative activities. This transformation has led to an exponential increase in network traffic within university campuses, driven by a multitude of factors such as online learning platforms, research collaborations, administrative systems, e-library, e-voting, computer-based tests/Exams, and the proliferation of internet-enabled devices.

The efficient management of university network infrastructure has become paramount to ensure seamless operations, optimal resource allocation, and enhanced user experience. Network administrators face the challenge of accurately predicting and managing network traffic behavior to meet the evolving needs of students, faculty, staff, and other stakeholders

Universities, as dynamic hubs of knowledge and innovation, increasingly rely on advanced network infrastructures to facilitate a wide array of activities, ranging from academic research and collaborative projects to online learning platforms. With the proliferation of digital technologies, the volume and complexity of network traffic within university environments have surged, presenting challenges for effective network management. The need to anticipate, understand, and optimize network behavior has become paramount for ensuring seamless connectivity, resource allocation, and overall operational efficiency.

This research focuses on the prediction of university network traffic behavior, leveraging the synergies between deep learning and optimization techniques. As the demand for online education, research collaboration, and data-intensive applications continues to grow, there is a critical necessity to develop advanced predictive models capable of adapting to the evolving dynamics of university network traffic. Understanding and predicting the behavior of university network traffic is crucial for network administrators to proactively manage resources, optimize bandwidth allocation, and ensure a seamless digital learning environment. Traditional methods of network traffic analysis and management may fall short in handling the complexity and dynamic nature of modern university network. This research holds significance in the realm of network management for educational institutions.

Across several academic institutions, especially in Universities, detailed and up-to-date analysis of the distribution of network services is critical to delivering efficient academic and administrative services. With the development of the Internet and advanced web applications, the computer network is changing students, staff, and university community lives, the internet makes life easy. A good internet traffic regulation is required to support these web-based applications and other applications or devices that use the internet, this traffic regulation can be achieved by predicting the university network traffic using past network traffic data. Jiboon Lee (2019). Predicting network traffic patterns becomes crucial for optimizing resource allocation, ensuring quality of service, and preventing congestion.

A university is an institution of higher education and research that grants academic degrees in various subjects/fields of study and disciplines. The term "university" is derived from the Latin word "universitas," which originally referred to a corporation or society of scholars and teachers. Today, universities serve as centers of learning, research, and innovation, offering undergraduate, graduate and postgraduate programs, as well as conducting scholarly research in diverse fields. Which allowing students to pursue degrees such as bachelor's, master's, and doctoral degrees.

Universities provide a wide range of disciplines and specializations, including but not limited to arts, sciences, engineering, medicine, law, business, and humanities. They often conduct research across various disciplines, contributing to advancements in knowledge and innovation.

In addition to academic programs, universities may offer facilities such as libraries, laboratories, sports complexes, and student residences. They also foster a diverse academic community comprising students, faculty, researchers, and staff, promoting intellectual exchange, critical thinking, and personal development. Universities play a significant role in society by educating future generations, advancing knowledge through research, and contributing to cultural, social, and economic development. They serve as hubs of learning, innovation, and cultural exchange, shaping the intellectual landscape of nations and the world. The word *university* is derived from the Latin *Universitas magistrorum et scholarium*, roughly meaning "community of teachers and scholars." Source: <https://www.newworldencyclopedia.org>

A computer network is a collection of interconnected computers and other devices that are linked together to enable communication, resource sharing, and data exchange. In a computer network, individual computers, also known as nodes or hosts, are connected using various communication channels, such as wired or wireless connections. These networks can range from small, local networks within a single building to vast global networks that span continents and connect millions of devices.

Computer networks play a crucial role in enabling communication and information exchange in various contexts, including businesses, educational institutions, government agencies, and homes. They facilitate collaboration, resource sharing, and access to information, leading to increased productivity, efficiency, and connectivity in modern society.

In the elapse of time, the network has become so common and it can be assumed that nowadays almost every computer user is somehow involved in a computer network. Whether it is a LAN (Local Area Network) or the internet, a computer network is considered as a base for communication media, as it provides a cheap and fast way to transfer and share information. While the internet can provide a wider coverage of the resources and information that can be accessed from the office, home, or anywhere, provided that an internet connection is available. Even software today is designed to be used in a network, for example, antivirus software that has an automatic update feature from the internet to the software. All in all, it's due to the effectiveness and cheapness of sharing information that makes it today's necessity in a computer.

Information is an asset, as network is. Without networks, we will lose the convenient way to obtain information. Today, the internet provides people with a lot of means in information retrieval. One of the examples is Google search engine, which can be used to search a vast amount of information. In another way round, without the internet, the server which stores the information wouldn't be effective in delivering the information to the user. That's why many of the organizations today are using the internet in conjunction with their intranet in distributing information within the organization. It is proven that computer networking can provide us with a cheaper means of exchanging, sharing, and transferring information. Gillis (2019).

Network traffic describes the number of network packets trying to get through an interconnected network at a given point in time. Network traffic, also called data traffic, is broken down into data packets and sent over a network before being reassembled by the receiving device or computer. Network traffic is the main component used to measure and control the usability of a computer network. A smooth flow of data packets enables us to surf the web or share information without any issues, and it keeps the nodes in sync for further communication. When data travels over a network or the internet, it must first be broken down into smaller batches so that larger files can be transmitted efficiently. The network breaks down, organizes, and bundles the data into data packetsso that they can be sent reliably through the network and then opened and read by another user in the network. Each packet takes the best route possible to spread network traffic evenly. Gillis (2019).

Network traffic is the main component for any network. The network will be unable to understand the user request without any requests. The way modern network is used nowadays makes the traffic become one of the crucial components both for the network owner and the user itself. The network owner might want to guarantee that their network can be used properly and efficiently. On the other hand, the user wants to make any communication to another can be done quickly and with the best result. By communication, it means accessing remote data, sending an email with attachment, or simply having a chat. An example of the latest case is instant messaging that uses some kind of streaming to make the message notification appear quickly in the recipient. In the scenario described above, the network simply acts as a medium for the transfer of data, and the transmission of the data is the primary objective of the user. Now consider a computer user that is searching for some information on the internet or browsing a webpage. This data has been read from some remote location and transferred to the user's computer. This mode of data transfer is called data activity to distinguish it from data production, and it adds another level of complexity in the data transfer scenario. Tian and Li (2017).

Legend and Taqqu (1994) studied the characteristics of network traffic and found that it has self-similarity characteristics. The self-similarity of time series can be measured by the Hurst exponent. Their studies also pointed out that the Hurst exponent of the network traffic is greater than 0.5. Network traffic has self-similarity characteristics. Thus, the network traffic is predictable. Their study laid the foundation work for network traffic prediction. At present, many scholars and researchers have done a lot of research work on network traffic prediction. Tian and Li (2017)

Effective management of network traffic is essential for maintaining optimal network performance, security, and reliability. By monitoring and analyzing network traffic patterns, organizations can identify bottlenecks, optimize resource allocation, and mitigate security risks to ensure a smooth and efficient operation of their computer networks.

Network traffic behavior refers to the patterns and characteristics exhibited by data transmissions over a computer network. Understanding network traffic behavior is crucial for effectively managing and optimizing network performance, security, and resource utilization. Here are some common network traffic behaviors: Regular Patterns, Anomalous Patterns, Traffic Composition, and Protocol Distribution. Analyzing network traffic behavior helps identify normal operational patterns, detect anomalies and security threats, troubleshoot performance issues, plan network capacity upgrades, and optimize resource allocation. Network monitoring tools and techniques, such as packet sniffers, flow analyzers, and anomaly detection systems, are commonly used to capture, analyze, and visualize network traffic behavior in real-time or retrospectively.

University network traffic behavior can vary significantly depending on factors such as the size of the university, the number and types of users, the types of academic programs offered, the services provided, the technological infrastructure in place and the availability of network resources. Understanding university network traffic behavior is essential for network administrators to effectively manage network resources, ensure adequate bandwidth allocation, optimize network performance, and address security concerns such as unauthorized access, malware, and data breaches. Network monitoring and analysis tools can help administrators monitor traffic patterns, identify anomalies, and implement appropriate network management policies to support the diverse needs of the university community while maintaining network security and reliability.

The behavioural complexity of the network traffic required a critical analysis to manage the network effectively to achieve a high-level quality of service (QoS). Several key issues arising from the network complexity involve poor optimization of the resource allocation for a network, poor QoS delivery, untimely quality packet distribution, unreliable free network traffic flow, and vulnerability in network security. Also, due to limited network resources, costly network infrastructure, and increasing user space on a network, all the identified issues become adversely prominent on the network quality, and the network becomes more challenging to manage. The proper organization of network traffic helps in ensuring the quality of service in a given network.

There is a need to efficiently analyze network traffic behaviour to obtain actionable insight into the influence of the traffic on a computer network to achieve an effective QoS. A proper analysis of network traffic guarantees efficient network management. Thus, the design of a predictive model to analyze the flow of network traffic is essential. Niam-fa, (2017).

Predictive models allow network providers to proactively allocate resources based on anticipated traffic demands. By leveraging deep learning techniques, we aim to enhance the accuracy of traffic prediction and optimize network performance. Several predictive methods have been proposed in recent years. Notably, deep learning approaches have gained attention due to their ability to handle complex temporal data. The prediction of network traffic behaviours has a variety of applications, including network monitoring, resource management, and threat detection. The benefit of getting accurate predictions for network traffic is significant to effectively distribute network services to intended devices, users, and locations with a focus on optimal QoS (Nipun and Tarun, 2018). Also, the prediction of the behavior of network traffic helps in the provision, accurate network planning and scheduling of the usability of network bandwidth, resource allocation, detection of anomalies, congestion analysis, QoS automation, network design, and network decisions. Predicting network congestion is more effective than detecting congestion by measurement. Effective and accurate prediction of network traffic behaviours has a vital influence on network performance and it reduces the complexity of the network. Predicting network traffic provides short-term and long-term benefits to the university management to make better decisions (Ahmed and Sudhakar, 2020).

When network traffic is accurately predicted, it could become a benefit that helps to manage network resources to ensure the quality of service and help the network administrator improve the availability and transmission speeds of the network. Different research, designs, and experiment have been conducted or proposed for analyzing and predicting network traffic. The predictability of network traffic is of important benefit in many areas, such as dynamic bandwidth allocation, network security, network planning, predictive congestion control, optimal resource management, and so on.

Network traffic prediction is an important application of network management and network measurement research direction, which is to predict the future tendency of sometimes things to happen now or later. The network traffic prediction model can be regarded as an approximate mathematical description of traffic behaviour characteristics. Generally, the network traffic prediction model can be divided into three categories: linear model, nonlinear model, and combination model.

Network Traffic Prediction (NTP) as a sub-technique of Network Traffic Monitoring and Analysis (NTMA) is used to determine the status of the network, identify changes, and predict the network traffic behavior in the foreseeable future. Generally, the results of NTP techniques can be used in a wide range of applications, for example, QoS provisioning, fault detection, and security attack detection. The problem of predicting future network traffic volume is traditionally formulated in the form of a Time Series Forecasting (TSF) or rarely a spatiotemporal problem aimed at constructing a regression model capable of estimating future traffic volume by extracting the existing dependencies between historical and future data [Iraj](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Lohrasbinasab%2C+Iraj) et al. (2021)

Optimization techniques are methods used to find the best solution among a set of possible alternatives. This is typically achieved by modifying the system to a form that will consume less resource, whilst delivering the same output. These techniques are widely used across various fields, including mathematics, engineering, computer science, economics, and more. The choice of optimization technique depends on the problem at hand, including its characteristics, constraints, and the available computational resources. Often, a combination of techniques or customization to fit specific problem requirements is necessary for achieving optimal results. However, solving optimization problems can be challenging in real-world scenarios. This is where optimization techniques come into play. Optimization techniques are powerful and there are now many successful applications that are beginning to shape and influence our modern economy. For example, algorithmic optimization techniques are increasingly being used to help companies manage and distribute resources such as personnel and equipment.

The concept of optimization is nowadays utilized in numerous fields with its own importance. By only utilizing this technique, someone can upgrade their output without needing to increase the input. Modification of a system or approach to a certain best state often requires the need to change parameters or structure within a set of the possible solution. This can be said as the main goal in optimization. When the system or approach is very complex, sometimes it is hard to identify the best solution, so some algorithms and methods need to be applied to find the best solution. This problem is called an optimization problem. It is also becoming more prevalent for companies to embed optimization models into the infrastructures that support their day to day operations, enabling them to make better decisions. This trend is likely to continue and accelerate, and it is important that the developers and users of these systems understand the issues involved in building and solving optimization models so that they can do so effectively. In general, optimization can be divided into two categories, which are linear and nonlinear. Linear optimization is usually simpler than nonlinear because the characteristics of the function are linear. It has a well-known method called the Simplex method, which is well-suited to solving systems of equations because it can generate feasible solutions to the best solution at the same time. On the other hand, nonlinear optimization, which also has real-world problems, usually has some methods and algorithms to solve specific problems. In summary, optimization techniques play a crucial role in enhancing efficiency, reducing costs, improving decision-making, mitigating risks, fostering innovation, and promoting sustainability across various domains.

In recent years, deep learning has become the go-to method for machine learning tasks. The dramatic success of deep learning in a wide range of applications has led to a surge of interest in the field. Deep learning is a subset of AI and machine learning that can process data and solve simple problems in the near future and is making current headlines. The recent application of deep learning has resulted in a diversity of devices having intelligent capabilities, these include voice recognition in mobile phones, game playing computers, and driverless cars

Deep learning is part of a broader family of machine learning methods based on learning data representations. An observation (e.g. an image) can be represented in many ways such as a vector of intensity values per pixel, or a more complex representation like an affine transformation that results in a high dimensional vector. Most methods start by defining how the data is represented and then go on to learn a separate function. This separate function might be one that takes an input an image and outputs a vector of class scores. The learning part is typically performed using gradient descent and aims to minimize a loss function that measures the inaccuracy of the function in its ability to map the input to the output. Deep learning represents a family of machine learning methods that can automatically discover useful features from raw data by building up a hierarchy of features. This has been fairly common knowledge in the machine learning community for a long time. However, what has made deep learning so exciting in recent years is that the learning process can be directly applied to finding a useful representation of the data, rather than just learning a mapping from input to output. This is useful in many real world scenarios, for example an engineer might not have a good idea of what features to extract from raw sensor data in order to build an effective detector for a given event. If the engineer knew what features were best to extract, then building the detector could just be a simple optimization problem with respect to a classification function. Deep learning provides a means for the optimization to be performed directly on the detector, given a simple parameterization of the features.

Deep learning is a new area of machine learning research that has been introduced with the objective of moving machine learning closer to one of its original goals: artificial intelligence. The concept of deep learning is to move away from traditional methods of supervised learning, and instead progress to models that are based on an understanding of the brain, specifically the structure of the brain. These models are known as ANNs. The structure and function of the brain is a complex system and is still not fully understood despite the best efforts of neurologists; however understanding the biological brain is not the main aim of ANNs. Biological neurons are the key element of ANNs as they are the processing units of the brain. These neurons are represented by a node in the system and a typical node will receive input from other nodes (e.g. sensory organs), process this information and if the information is relevant the node will trigger the corresponding output by activating another node. Each connection, like synapses in the brain, can transmit information from one node to another and can be tuned in a similar way to biological neurons. Neural networks can be used to recognize patterns in data, and are used in the automatic generation of rule sets from example data, and the detection and categorization of patterns, and the stimulation of brain function is what inspired the development of ANNs. Statisticians have been developing models for a long time based on the theory that the data we can measure and manipulate is related to underlying causal system that generated it. Deep learning is simply a new method of doing this using ANNs as models and is the current focus of machine learning research today. **Source: https://www.techtarget.com/searchenterpriseai/definition/deep-learning-deep-neural-network**

The reason why deep learning was chosen was due to a few of the following reasons below:

1 Deep learning models yield results more quickly than standard machine learning approaches (Wehle 2017)

1. Unlike machine learning systems, which need a person to select and hand-code the applied characteristics depending on the data type, deep learning systems attempt to learn those characteristics without extra human interaction.  (Janiesch et al., 2021)
2. Deep learning is suitable for dealing with larger data and complexity ( Neha et al., 2021)
3. One of the biggest draws of deep learning is its ability to work with unstructured data (Wehle 2017)
4. One of the most significant advantages of employing a deep learning methodology is its capacity to do feature engineering.  (Janiesch et al., 2021)

Deep learning and machine learning are closely related fields within artificial intelligence, but they differ in scope, techniques, and applications. Here's a comparison between the two:

**Scope**:

**Machine Learning (ML)**: Machine learning is a broader field that encompasses a variety of techniques and algorithms for building models that can learn from data to make predictions or decisions without being explicitly programmed. It includes both traditional statistical methods and modern approaches like deep learning.

**Deep Learning (DL)**: Deep learning is a subset of machine learning that specifically focuses on artificial neural networks with multiple layers (deep architectures). Deep learning algorithms automatically learn hierarchical representations of data features directly from raw input data.

**Techniques**:

**Machine Learning (ML)**: Machine learning techniques include a wide range of algorithms such as linear regression, logistic regression, decision trees, random forests, support vector machines (SVM), k-nearest neighbors (k-NN), and more. These algorithms often require feature engineering, where domain knowledge is used to manually extract relevant features from the data.

**Deep Learning (DL)**: Deep learning techniques are based on artificial neural networks with multiple layers (deep architectures). These architectures include convolutional neural networks (CNNs) for image data, recurrent neural networks (RNNs) for sequential data, and various other architectures like deep belief networks (DBNs), autoencoders, and generative adversarial networks (GANs). Deep learning models automatically learn hierarchical representations of data features from raw input data, eliminating the need for manual feature engineering in many cases.

**Representation**:

**Machine Learning (ML)**: In traditional machine learning, feature engineering is a crucial step where domain knowledge is used to manually select, extract, and transform relevant features from the raw data. These features are then used as input to the machine learning model.

**Deep Learning (DL)**: Deep learning models learn hierarchical representations of data features directly from raw input data. This eliminates the need for manual feature engineering in many cases, as the deep learning model automatically learns to extract meaningful features from the data during the training process.

**Performance**:

**Machine Learning (ML)**: Machine learning algorithms can perform well on a wide range of tasks, especially when sufficient domain knowledge is available for feature engineering and model selection. However, their performance may plateau or decline on tasks with very high-dimensional or unstructured data.

**Deep Learning (DL)**: Deep learning algorithms have achieved state-of-the-art performance on various tasks, particularly those involving high-dimensional or unstructured data such as images, text, and audio. Deep learning models often outperform traditional machine learning methods on these types of tasks, especially when large amounts of labeled data are available for training.

**Applications**:

**Machine Learning (ML)**: Machine learning techniques are widely used in various applications, including classification, regression, clustering, recommendation systems, anomaly detection, and more.

**Deep Learning (DL)**: Deep learning techniques have been particularly successful in applications such as computer vision (e.g., image classification, object detection), natural language processing (e.g., machine translation, sentiment analysis), speech recognition, autonomous driving, and many others.

In summary, while deep learning is a subset of machine learning, it has gained prominence due to its ability to automatically learn hierarchical representations of data features from raw input, leading to state-of-the-art performance on various tasks, especially those involving high-dimensional or unstructured data. (Janiesch et al., 2021)

**1.2 Research Motivation**

The major weakness and problem in network traffic prediction (NTP) is the inability to monitor, evaluate, and predict accurate network traffic, some researchers have made contributions or proposed some model for handling network prediction, especially in a university or some higher learning environments, firstly, David and lobby, 2002 proposed a campus-wide wireless network to have a good background for developing system and application software on a network and also, as a guide for those that develop, deploy and manage WLAN technology. The result shows that students’ residential areas consumed more bandwidth and traffic than some areas. Roaming was also noticed because students used their cards to roam a lot. The limitation of this research is that it does not differentiate between different types of users (i.e. staff and students) or show aggregate results rather than the segmented result which could have assisted the management or administrator in planning better or making decisions.

Closely related to the above is Mebawondu (2014) presented research on Performance evaluation and modeling of internet network traffic in an academic institution. The aim was to establish a model of the network traffic that can serve as a tool to enhance the quality of service. Network traffic was captured using Netflow. Questionnaires were administered to students at different areas and levels to determine the availability rate of a network. The limitation of this research is the student can be biased in answering the questionnaire and a research-only measurement evaluation was carried out In today’s world improved accuracy and efficiency could still adhere.

Dahunsi et al (2015) presented research on the Analysis and Performance Evaluation of the FUTA network, the research focuses on the quality of service provided to its users. Data were collected by administering questionnaires to determine the availability, accessibility, and connectivity rate at various areas under the coverage of the network. Data were further analyzed using Statistical Package for Social Science (SPSS). The limitation of the research is that information provided during data collection could be based on the subjective reasoning of individuals which will affect the result.

Babatunde (2018) researched Network Traffic Analysis using quelling and regression techniques, the research explores the analysis of the internet traffic network of the institution (FUTA) using M/M/1 queuing model and regression techniques. The traffic data was captured using the Wireshark capturing tool at different strategic locations within the University. The captured data were analyzed using Statistical Package for Social Science (SPSS). The limitation of this research is that the network traffic was not predicted and the optimized model was not used to improve the accuracy of the prediction. Besides, improving the accuracy of prediction is possible and should be considered.

In Jihoon (2019) Prediction of University Network Traffic Using the Deep Learning Method was presented. Using a deep neural network model to predict the university network traffic using the past network traffic data, the researcher compares the network before the implementation of deep learning and after the implementation of deep learning. The system went through the training phase, testing phase, and prediction phase to familiarize itself with the current network system. The prediction accuracy of this proposed system could be further optimized via the use of optimization techniques.

Hence, this research is therefore motivated by the need to address the aforementioned limitations using deep learning models combined with genetic algorithms to optimize the prediction accuracy

**1.3 Aim and Objectives**

This research aims to develop a predictive model to aid in predicting university network traffic behaviours. The primary objective of this research is to develop a robust prediction framework for university network traffic behavior. This involves the utilization of deep learning techniques and optimization methods to enhance the accuracy and efficiency of traffic prediction models. The specific objectives of the research are to:

i. design prediction-based models for network traffic behaviour using deep learning algorithms and optimization techniques;

ii implement the model designed in i; and

iii carry out performance evaluation of the model designed in ii

**1.4 Methodology**

This research started with a review of existing works in the prediction of network traffic. A framework that aims to solve some of the limitations of the existing systems is then proposed. The research methodology aimed to predict the network traffic using deep learning models and a Genetic algorithm to optimize the solution. The research methodology is in seven steps to achieve the research objectives, and the steps are: Data Collection, Data Preprocess Algorithm selection, Training/Testing, Prediction, Optimization, and Performance evaluation metric

**1.4.1 Data collection**

Good capturing tools will be installed to provide an effective capturing of the university network traffic at strategic locations within the University. To do this, Wireshark which is a real-time capturing tool will be used to capture the flow of packets to derive the arrival rate

**Data preprocess**

The data preprocessing will follow this process: **Data Cleaning,** Data Normalization, and Data reduction.

**1.4.2 Algorithm/model selection**

The prediction models proposed for this research:

long short-term memory with Genetic Algorithm,

Restricted Boltzmann machine with Genetic Algorithm, and

An optimized Gated Recurrent Unit (OGRU) model

**Training** is the process of teaching a neural network to recognize patterns and make predictions by its internal parameters based on input data.

**Testing** involves evaluating the performance of a trained neural network on new, unseen data to access its generalization ability and accuracy.

**1.6 Optimization Techniques**

Optimization techniques are the techniques used to discover the best solution out of all the possible solutions available under the constraints present. **Genetic Algorithms (GA)** will be used in this research.

**1.7 Performance Indicators / Performance Evaluation**

To illustrate the effectiveness of the prediction model for network traffic, the following performance Evaluation metrics or Standard Metrics are introduced to measure the prediction accuracy of the prediction model

**EXPECTED CONTRIBUTION TO KNOWLEDGE**

The research is expected to:

a. Establish a more accurate prediction model that functions based on deep learning and optimization technique.

b. Provides a reference for the planning of university network traffic

Organization of the Thesis