Al-Powered Cyber Threat Prediction Using Graph Neural Networks (GNNs)

Ambika Kunj[1], Dr. Kushal Pal Singh[2]

[1]Research Scholar, Department of Mathematics, JS University, Shikohabad, Uttar Pradesh
[2]Supervisor, Department of Mathematics, JS University, Shikohabad, Uttar Pradesh

ABSTRACT

On account of the ever-evolving nature of online threats, it is difficult for businesses to keep up with the latest developments. It is necessary to make use of contemporary methods in order to identify threats in a timely and accurate manner. The majority of security solutions are based on rule-based or signature-based techniques, which have difficulty identifying threats that are not just complicated but also covert or dynamic. As a result of the deep links that they are able to describe and evaluate in data from networks, system logs, and patterns of user activity, Graph Neural Networks (GNNs) have the potential to be an exceptional artificial intelligence tool for the purpose of cyber threat forecasting, as shown by this study. In the GNN-based strategy that has been described, nodes are used to represent devices, users, IP addresses, and network flows. Meanwhile, edges are used to represent the interactions between these nodes. It is essential to be able to recognize both structural relationships and temporal patterns in order to discover new hazards, and this delivers both of those things.

Rich embeddings are generated by the system via the use of the GNN's recurring message-passing features in conjunction with data from nodes and edges. These characteristics include connection frequency, user data access patterns, information on prior security flaws, and behavioral factors. Because of this, the system is able to identify potentially malicious acts, such as gaining access to greater privileges, moving from one level of authority to another, or launching distributed denial of service attacks, all of which might be indicators of hacking by the system. With attention-based GNN versions as well, different relevance weights are applied to the connections that are being considered. Because of this, the model is able to reduce the amount of false positives and concentrate on the behaviors that pose the most danger.

With the use of real-world network datasets and safety requirements, we put the proposed model through its paces and evaluate it in comparison to well-known machine learning approaches such as Support Vector Machines, Random Forests, and ordinary neural networks. When it comes to identifying both old and new hazards, the GNN-based system surpasses the competition in terms of precision, accuracy, memory, and F1-score. This is something that was determined by the findings. Not only that, but the model is able to function effectively even when it is confronted with busy or missing data, which is something that often occurs in real-world computing networks.

According to the findings of this study, graph-based deep learning is an approach that is both efficient and successful when it comes to danger prediction in the field of cybersecurity. Through the use of interpretable embeddings and attention processes that emphasize crucial attack lines, the suggested strategy improves the capacities of both discovery and decision-making. When these considerations are taken into account, merging artificial intelligence with general neural networks (GNNs) is an ideal technique for designing next-generation security systems that are adaptable and capable of dealing with increasingly sophisticated threats in dynamic network situations.

Keywords: Graph Neural Networks (GNNs), Cyber Threat Prediction, Network Security Analytics, Anomaly Detection

1. INTRODUCTION

The widespread circulation of misleading information has become a serious concern as a result of the quick diffusion of information via social media and other platforms that are almost identical. It is especially true in important fields like public health, where the situation is critical. In May of 2024, there were around 4.8 billion people using social media platforms all throughout the globe. When compared to January 2021 and January 2022, when the corresponding populations were 4.2 and 4.62 billion, this is a huge growth [1]. The exponential development of social media is

evidence that it continues to have an impact on the manner in which people all over the world connect with one another and share information amongst themselves. There is still widespread dissemination of misleading information about health. A significant number of individuals continue to use the term "misinformation" to refer to information that is either biased or erroneous, meaning that it does not fit with what doctors normally know or presents facts in a biased manner.

In the majority of instances, this false information spreads at a pace that is far faster than the genuine scientific knowledge. A typical source of this is the widespread diffusion of statements that are not supported by evidence or the absence of definitive scientific proof. In today's world, everyone utilizes social media, which has resulted in a change in the way information is sent and received. It is unfortunate that the ease with which information can be communicated has also made it easier for incorrect information to spread quickly, especially in regards to topics pertaining to health. It is possible that the dissemination of false information about health on social media might lead to a variety of problems, such as public health catastrophes, reluctance to get vaccines, and treatments that are either hazardous or ineffective.

Therefore, we need to quickly come up with effective methods of identifying and minimizing the number of health-related misinformation that are spread on these platforms. It is of the utmost importance to detect instances of health misinformation on social media platforms in order to avoid the widespread dissemination of remarks that are either harmful or erroneous. AI-based techniques that have the ability to solve this problem include data analysis, machine learning, and natural language processing (NLP) elements [2]. The term "misinformation" encompasses a wide range of concepts, including but not limited to the following: untrustworthy, spam, gossip, disinformation, information disorder, and so on.

Uncertainty, results that contradict one another, and evidence that develops over time are all examples of various types of data that might be harmful to one's health.

Furthermore, despite the fact that communication is of the utmost importance, it is not enough to solve these problems that include blatantly wrong information [3]. Figure 1 illustrates a number of different types of disinformation, some of which include rumors, lies, false information, and information that is not accurate.Due to the fact that it is possible for anybody to submit content on social networking sites without being subject to any external restrictions, it may be challenging to assess whether or not a source is reliable [4, 5]. It is possible for anybody to disseminate inaccurate information, and those who receive it or assist in correcting it will have their own personal goals in mind [6].
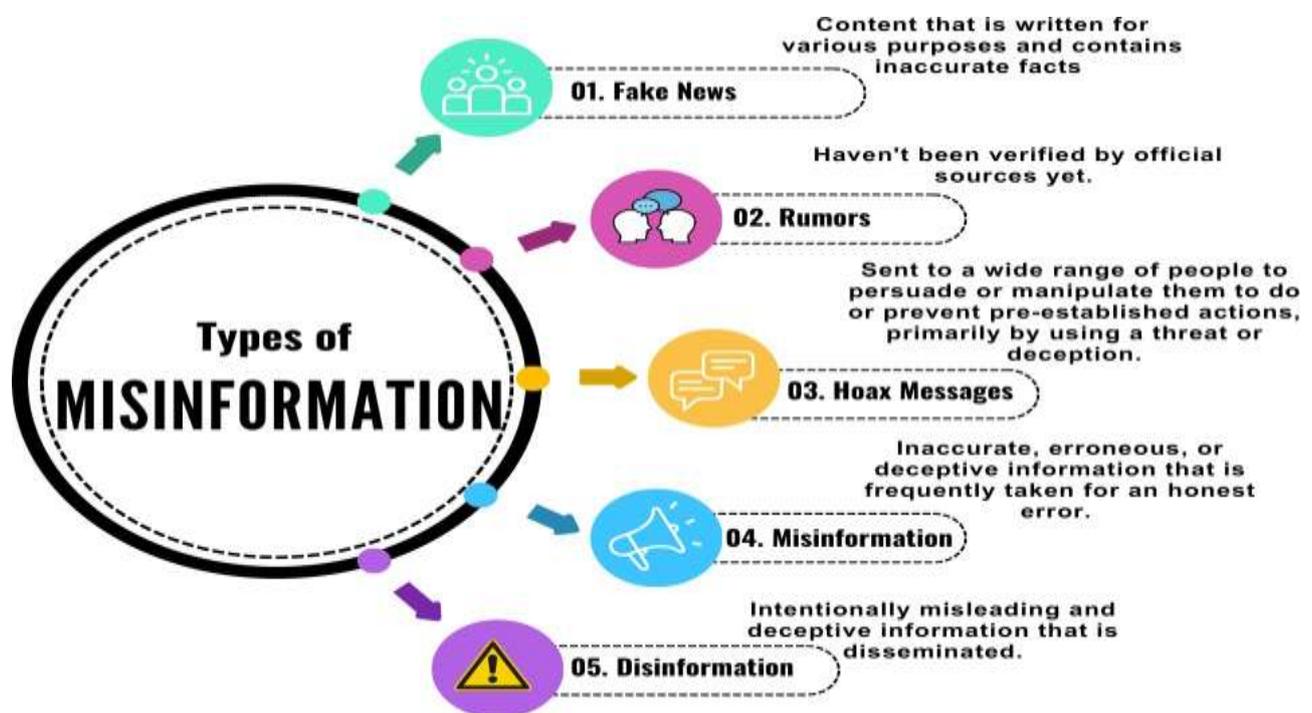

Figure 1.1: Types of Misinformation

A. **Detection Obstacles in the Detection of Health Misinformation**

There are billions of people who are members of online social networks, and they interact with material in real time and share it with others. This is a rapid and scalable phenomenon. When it comes to health-related disinformation, which may become viral in a matter of hours, it is difficult to maintain control and prevent its rapid spread.

Hold in check. As a result of the rapid dissemination of incorrect information, the impact of this information is exacerbated, making it more difficult to safeguard the health and safety of the general people.

Restrictions imposed by the currently available detection methods: Traditional methods of machine learning and rule-based techniques are used in a significant number of the ongoing initiatives to identify health misinformation. Disinformation is always changing, and these tactics have a difficult time keeping up with it. Considering that they are unable to adapt, it is possible that as time passes, they will become less efficient at detecting new or subtle kinds of deceit. In addition, many of the methodologies that are now in use disregard the relational and contextual data that is available in social media in favor of content-based analysis [16].

The spread of false information about health may result in the development of unhealthy behaviors, poor choices, and a loss of confidence in healthcare institutions. It is for this reason that strategies that are successful in identifying and reducing health misconceptions are necessary quickly. Generalized Convolutional Neural Networks (GCNs) are a kind of neural network that use mathematical structures known as graphs, which are composed of nodes and edges, to recognize health misinformation presented in textual data. The use of GCNs has proven effective in a number of different domains, including computer vision and social network research. Recently, researchers that are interested in natural language processing (NLP) applications such as entity identification, sentiment analysis, and text classification have noticed that GCNs have garnered their attention [17].

Quoted as [18]. In addition to providing an overview of illness prediction applications that make use of a variety of Graph Machine Learning (GML) algorithms, the study goes into a wide range of approaches that are used within graph machine learning on a variety of different levels. This work provides significant insights into the potential of machine learning in illness prediction using electronic health data by conducting a comprehensive literature search and trend analysis [19]. These findings are supported by the findings of the study. There is a rising worry over the dissemination of false health information [20], despite the fact that quick information interchange has significant advantages. There is the potential for health information that is either untrue or misleading to have serious repercussions, including the encouragement of actions that are detrimental to one's health or the dissemination of wrong information about procedures and treatments. As a result of the quick dissemination of information on social media, this issue has become much more severe, which is why there is an immediate need for effective measures to identify and combat health misinformation [21].

II. LITERATURE REVIEW

Due to the widespread availability of inaccurate information on health, it is especially difficult for public health professionals to carry out their responsibilities and make decisions in this digital age. Identifying and preventing fake information is an essential step in the process of providing information that can be relied upon. In natural language processing (NLP) applications such as text categorization, sentiment analysis, and information discovery, there is a great deal of promise for machine learning and deep learning techniques [22].

By using these methods, we are able to search for wrong information by swiftly scanning and analyzing huge amounts of text using data-driven methodology. This allows us to find material that is not accurate. As a result of its capacity to show complicated patterns and relationships, GNNs have gained a significant amount of attention as prospective means for finding deceptive information in situations impacting health. In this literature review [23], we take a look at recent research and developments that make use of GNNs to uncover health myths.

The data determines which of the various AI-based tactics is used in order to get the best possible results. At the moment, there is a significant amount of false information that is being disseminated via the internet and on social media [40]. Due to the volume of information that is both correct and erroneous that is accessible online, it is difficult for users to quickly identify what they need. The process of determining whether or not data is real is referred to as identifying misinformation. Several different machine learning algorithms have been developed by researchers in an effort to detect bogus news; nevertheless, the outcomes of these attempts have been insufficient. RNNs, SVMs, KNNs, LRs, RFs, CNNs, sentiment-aware multimodal embedding, hierarchical attention networks, and dEFEND are some examples of fundamental machine learning models that are used to categorize data in accordance with their principles. Additional examples include CNNs. [41] [41] This is the [42].

A viewpoint known as the Elaboration Likelihood Model (ELM) is used in order to analyze the two types of health fraud that occur online. These are referred to as the core and secondary. Propaganda, advertising, information that is not useful, and information that is wrong are the four primary types of misleading information that may be discovered in online health forums [43]. For the purpose of developing a model for the finding of health news, data on language, topics, feelings, and behaviors were merged. For the purpose of evaluating the features of the suggested model, a dataset from the actual world was used, and the results revealed that around 85 percent of health lies were accurate [44, 45].

When it comes to discovering health lying in online health communities, behavioral criteria, as opposed to linguistic ones, are more useful [46]. These individuals have developed three reliable methods for identifying stuff that is not true. A number of other models, including KNN, MC-CNN, and LSTM, which is a form of RNN, are now under consideration [37]. For the purpose of carrying out the simulations, the CoAID disinformation dataset was used. This dataset is comprised of a collection of misinformation about healthcare [47]. Along with the Naïve Bayes approach, gradient boosting algorithm, support vector machine, multi-layer perceptron, and decision tree, we use three main types of data. It was found that the SVM had the best level of accuracy among all of them [48]. We went with the LR model since it offers a simple cost function and an equation for categorizing problems into two or more categories. This was a major factor in our decision. A set of new choices for displaying documents is one of the most important changes that has been implemented.

III METHODOLOGY

There are a lot of misleading health claims that are spread over social networking sites. There have been a number of studies that have shown that false information on health care may spread wider and more quickly than the truth. Examples of significant areas that have been harmed include unsubstantiated claims for vaccines, errors associated with COVID-19, and unproven treatments for a variety of ailments. Social media platforms are ideal for spreading disinformation such as this because of the vast user bases they have and the ease with which they can communicate with users. To provide an example: According to the findings of a study conducted by the World Health Organization, a significant amount of false information on the COVID-19 pandemic and its treatment received widespread dissemination. Because of this, there was a "infodemic" that made it more difficult for public health professionals to deliver support to those in need.

Social media platforms such as Facebook and Twitter have been responsible for the dissemination of false information on vaccines, which has resulted in a reduction in the number of people who have had vaccinations or even a total refusal to get vaccinations, as shown by study. It has had a negative impact on the health of the general populace. In the event that customers continue to wait for care or seek treatments that are not essential as a result of deceptive promises, the healthcare system may wind up costing the economy more money. On a frequent basis, being exposed to material that is alarming or inaccurate may have a detrimental influence on people's mental health by causing them to experience worry, anxiety, and bewilderment. People who are not well educated are more likely to make bad decisions about their health, such as not getting vaccinated, using treatments that have not been fully examined, or neglecting to take preventive measures. In the worst possible situation, more diseases will spread, which would result in more fatalities.

A. Neural Network for Graphs

A GNN, which is a special kind of neural network, has the potential to handle data that is organized in a graph in a more efficient manner. In a manner similar to social networks, chemical structures, and knowledge graphs, graphed neural networks (GNNs) are the way to go when dealing with easily graphed real-world data. When given input that is organized in a grid, such as pictures and patterns, concurrent neural networks (CNNs) and recurrent neural networks (RNNs) perform better than other types of neural networks.

The purpose of strong neural networks, of which GNNs are a subset, is to handle data that is provided in a graph structure. It has been shown that they are highly useful in a number of settings due to the fact that they may exhibit links and linkages. The mathematical theory, essential structures, training procedures, and applications of GNNs are comprehensively discussed in this lengthy exposition, with a focus on text data applications. The exposition also includes a great deal of information. Graph Neural Networks, also known as GNNs, are a kind of deep learning model that was developed expressly for the purpose of processing data that is structured in the form of a graph. Neural Networks for Graphs is another name for these types of networks. Data is often shown in GNNs using nodes, which represent vertices, and edges, which represent relationships. The operation of typical neural networks is based on vectors, pictures, or sets of specified sizes; this function is separate from such neural networks. For this reason, they shine when applied to real-world problems that include interactions and connections that are of critical importance.

The usual notation for defining a graph is G=(V,E), where V represents a collection of nodes and E represents the edge that links any two of them. In other words, G is used to describe a graph. It's possible that every node and edge has certain traits in common with one another. Generalized neural networks (GNNs) are able to learn to generate meaningful representations at the node, edge, or graph levels by using both structural and feature data. Neighborhood aggregation, also known as message sharing, is the fundamental concept upon which Google Neural Networks are founded. Throughout the course of the process, every node is responsible for updating its own image by gathering information from other nodes. After collecting signals from their neighbors and adding them up using an aggregation function like as max, sum, or mean, nodes carry out a neural update at each network layer. This occurs after the nodes have received this information. It is possible for nodes to understand patterns in the structure on both a local and a global scale if this method is repeated. As a result, data is able to travel across the network.
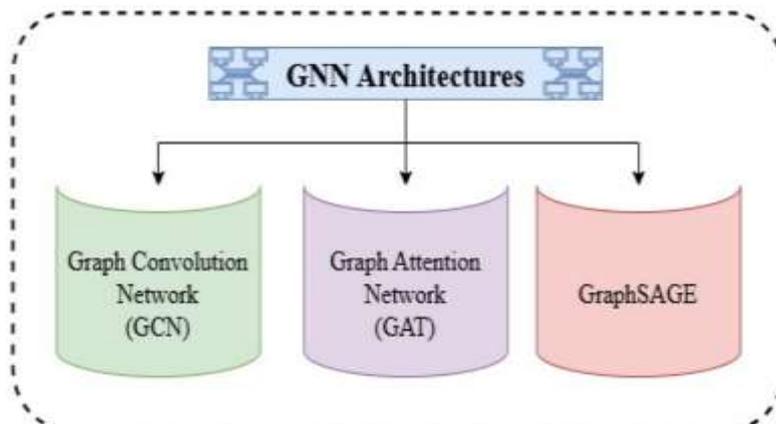
B. GNN Architectures

Figure 3.1: GNN Architecture

Patterns of Graph Neural Networks (GNNs) reflect the collection, movement, and change of information in graph-structured data. These patterns are used to generate neural networks. Throughout the years, a number of different designs have been created in order to address many different concerns, including expressiveness, scalability, and adaptability to different kinds of graphs. The manner in which these systems handle different kinds of graphs, in which they prioritize nodes, and in which they collect area data varies.

A well-known and well-established kind of neural network is known as the Graph Convolutional Network (GCN). The notion of convolution is transferred from grid-structured data to graphs via the use of GCNs. This is accomplished by adding together the weights of every node's neighbors, including the node itself. It is via the use of normalized adjacency matrices that they are able to avoid feature growth and keep the learning rate constant. When it comes to semi-supervised node classification tasks, GCNs perform very well, provided that the graph topology remains consistent and the graph size is relatively modest. It is possible that GCNs will have difficulty adequately portraying the surrounding nodes of complex networks since they operate on the premise that all nodes are of similar importance.

GATs, which are Graph Attention Networks, are able to avoid this problem since they include attention processes into GNNs. The GATs assist the model in concentrating on the most important connections by requesting that nodes provide their partners with varying degrees of attention throughout the process. When working with mixed graphs, in which friends contribute varying amounts, this customizable ranking makes the process of working with mixed graphs simpler. The capacity to concentrate on more than one activity at the same time makes it easier to describe oneself and ensures that learning is maintained consistently. GATs may be challenging to employ on large graphs because to the pairwise attention calculations that they need, despite the fact that they provide these advantages.

The GraphSAGE (Graph Sample and Aggregate) design was another key design that was developed to facilitate natural learning. GraphSAGE is able to generalize to nodes and graphs that it has never seen before because it learns aggregation functions rather than fixed embeddings. This allows it to acquire new capabilities. Contrast this with models that are based on transduction, such as GCNs. The features of a certain number of neighbors in close proximity are compiled by it via the use of techniques such as mean, max-pooling, or aggregators based on LSTM architecture. As a result of the fact that this sampling approach makes it easier to add more nodes, GraphSAGE performs very well with big graphs, similar to social networks.

Message passing neural networks, often known as MPNNs, serve as the basis for a broad range of generic neural networks (GNNs). There is a constant flow of messages between the nodes in MPNNs and their neighbors, which causes the nodes' statuses to vary over time. An effective structure for molecular graph analysis is provided by this structure. Atoms are represented by the nodes in this diagram, and the lines connecting them are the chemical bonds that exist between them. The message and update functions available with MPNNs have the potential to be versatile, which enables the incorporation of edge characteristics and information that is relevant to a domain.

The inadequacies of prior GNNs in terms of descriptiveness were the impetus for the creation of Graph Isomorphism Networks, often known as GINs. In order to enhance their capacity to differentiate between different graph designs, GINs are meant to be as resilient as the Weisfeiler-Lehman graph isomorphism test. The use of multilayer perceptrons after sum-based aggregation is an efficient method for addressing problems related to graph categorization.

Designs that are more advanced, such as Heterogeneous GNNs, Relational GNNs, and Dynamic GNNs, are able to handle graphs that include a variety of node and edge types, in addition to topologies that continuously change over time. These models improve upon the typical GNN architectures in order to accomplish the goal of accurately representing the complexity of reality.

C. Graph Convolutional Networks (GCNs)

GCNs are a kind of neural network architecture that are used for the purpose of processing and evaluating data that is structured and organized in a graph. GCNs are well suited for tasks that include graph-structured data, such as those found in social networks, recommendation systems, reference networks, and other applications that are connected to these types of networks. The use of these networks might be a complement to the conventional CNNs. Examples of applications that may benefit from these include those that need data to be organized in a grid, such as photographs. CNNs exclusively use the convolution method on grids, despite the fact that GCNs also use it on graphs using the convolution method. GCNs, or graph convolutional networks, are convolutional algorithms that are used to handle graph data. By taking into consideration the distinctive characteristics of each node in addition to those of its neighbors, it is possible for data to be gathered and distributed across the nodes that make up a network.

In contrast to GNNs, CNNs are especially built to deal with ordered (Euclidean) data, such as grid-based photo collections. This is one of the most important differences between the two types of networks. GNNs, on the other hand, improve CNN's ability to deal with non-Euclidean data, which is often asymmetrical owing to the fact that the ordering of nodes and the kinds of connections involved are not uniform. A generalized convolutional neural network (GCN) uses many layers of convolution and aggregation to enhance the representation of the graph's nodes. The ability of GCNs to infer complex connections and patterns within the structure of the graph is made possible by the technique of stacking layers.
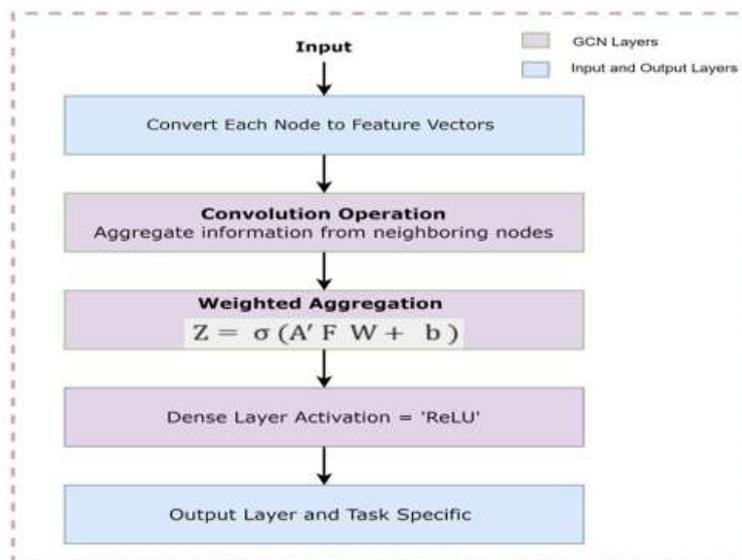

Figure 3.2: Working of GCN

IV DISCUSSION

About the discovery of health information that is not accurate. The primary objectives of this research are to investigate the efficacy of various modeling techniques and to determine the effects that mixed methods have on recognition performance. For each model, a comprehensive analysis of its accuracy, precision, recall, F1-score, and ROC-AUC was presented in the chapter that came before this one. In the sections that are to come, we will discuss the usefulness of graph-based techniques, the operation of mixed approaches, and the differences in performance that exist across different datasets.

A dataset having information about the COVID-19 pandemic and another dataset providing information about the original health are both used in this example to demonstrate how to apply health misinformation detection to two different datasets. There were 15,641 tweets included in the COVID-19 dataset, and only 57% of them were considered reliable, while 43% were considered to be untrustworthy. For the duration of the outbreak, computers are instructed to differentiate between genuine and fabricated material via the use of these tweets. In order to deliver accurate predictions for incoming text inputs, we reviewed the models that were deployed, which included GNN-based

techniques, and illustrated how they functioned by using a live Streamlit interface. By comparing our mixed model against leading techniques, which obtained remarkable accuracy rates of 96.16% on the health dataset and 89.23% on the COVID-19 dataset, we were able to illustrate the effectiveness of our mixed model. Additional topics that were discussed in this part included the complexity of time and space, mathematical models, as well as the advantages and disadvantages of each combination strategy. The flexibility and reliability of a number of different techniques, as well as the computing requirements of graph-based models, were explored.

CONCLUSION

The project that aimed to build systems driven by artificial intelligence to identify health-related misinformation on social media platforms produced positive findings over its course of investigation. When integrated with a variety of graph neural network (GNN) models, text feature extraction methods such as TFIDF, Cosine Similarity, and PMI produced results that were much better than before. The results of this comparison study suggest that the most advanced approaches, such as Graph Neural Networks, are effective in identifying instances of health fraud that occur on social media platforms. The process of fine-tuning the settings has a significant impact on the efficiency of these operations. We came to the conclusion that GraphSAGE was the most effective tool that was available.

As can be seen in Tables 4.4 and 4.8, mixed models, in particular those that make use of Graph Neural Networks (GNNs) like GraphSAGE (GSAGE), perform much better than deep learning and standard machine learning models. This is the case across the board. Support Vector Machines (SVMs), Random Forests (RFs), and Decision Trees (DTs) are examples of machine learning algorithms that perform very well in terms of accuracy, precision, and F1-scores. Use Random Forest to get the best possible outcomes. All things considered, the GNN-based mixed models emerge victorious when ROC-AUC and other performance measures are taken into consideration. As a result of the fact that deep learning algorithms like CNN and LSTM perform badly on the COVID-19 lies dataset, especially in terms of accuracy, it is obvious that these models are unable to handle the complexity of the dataset.

Performance-wise, Hybrid Models 2 and 3 consistently surpass GSAGE, which is comprised of the other hybrid models. Their ROC-AUC values are really close to being perfect, and their F1 scores are significantly highest. It has been shown via these studies that GNNs are capable of successfully recognizing complicated text links when combined with other embedding or similarity methods. On account of this, they are well suited for professions that involve the identification of incorrect information about health. According to the results, mixed models are a good choice for difficult text sorting problems since they considerably boost performance. This is especially true when mixed models are paired with GSAGE modeling in particular.

REFERENCES

[1]. DataReportal, "DataReportal," accessed Aug. 28, 2024. [Online]. Available: https://datareportal.com/.
[2]. D. Boovitha, M. Abirami, S. Gunavathi, N. Revathi, and S. Rubavarshini, "Fake Media Detection Based on Natural Language Processing and Blockchain Approaches," South Asian J. Eng. Technol., vol. 13, no. 1, pp. 69–82, 2023.
[3]. W. Y. S. Chou, A. Gaysynsky, and J. N. Cappella, "Where we go from here: Health misinformation on social media," Am. J. Public Health, vol. 110, pp. S273–S275, 2020, doi: 10.2105/AJPH.2020.305905.
[4]. B. Swire-Thompson and D. Lazer, "Public health and online misinformation: Chal- lenges and recommendations," Annu. Rev. Public Health, vol. 41, pp. 433–451, 2019, doi: 10.1146/annurev-publhealth-040119-094127.
[5]. S.-Y. C. C.-J. Y. &Fang Yu, "XFlag: Explainable Fake News Detec- tion Model on Social Media," Int. J. Hum. Comput. Interact., 2022, doi: https://doi.org/10.1080/10447318.2022.2062113.
[6]. Vraga, Emily K., and Leticia Bode. "Correction as a solution for health misinformation on social media." American Journal of Public Health 110, no. S3 (2020): S278-S280.
[7]. S. Zhang, F. Ma, Y. Liu, and W. Pian, "Identifying features of health misinformation on social media sites: an exploratory analysis," Libr. Hi Tech, no. 71420107026, 2021, doi: 10.1108/LHT-09-2020-0242.
[8]. Suarez-Lledo, Victor, and Javier Alvarez-Galvez. "Prevalence of health misinformation on social media: systematic review." Journal of medical Internet research 23, no. 1 (2021): e17187.
[9]. L. Cui, H. Seo, M. Tabar, F. Ma, S. Wang, and D. Lee, "DETERRENT: Knowl- edge Guided Graph Attention Network for Detecting Healthcare Misinformation," Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., pp. 492–502, 2020, doi: 10.1145/3394486.3403092.
[10]. S. G. Paul, A. Saha, M. Z. Hasan, S. R. H. Noori, and A. Moustafa, "A systematic review of graph neural network in healthcare-based applications: recent advances, trends, and future directions," IEEE Access, 2024.
[11]. H. T. Phan, N. T. Nguyen, and D. Hwang, "Fake news detection: A survey of graph neural network methods," Appl. Soft Comput., p. 110235, 2023.
[12]. T. Liu et al., "Rumor Detection with a novel graph neural network approach," arXiv Prepr. arXiv2403.16206, 2024.
[13]. B. Khemani, S. Patil, K. Kotecha, and S. Tanwar, "A review of graph neural networks: concepts, architectures,

techniques, challenges, datasets, applications, and future di- rections," J. Big Data, vol. 11, no. 1, p. 18, 2024.

[14]. J. Zhou et al., "Graph neural networks: A review of methods and applications," AI Open, vol. 1, no. January, pp. 57–81, 2020, doi: 10.1016/j.aiopen.2021.01.001.

[15]. E. Min and S. Ananiadou, "PESTO: A Post-User Fusion Network for Rumour De- tection on Social Media," in Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, 2023, pp. 1–10.

[16]. A. D'Ulizia, M. C. Caschera, F. Ferri, and P. Grifoni, "Fake news detection: a survey of evaluation datasets," PeerJ Comput. Sci., vol. 7, p. e518, 2021.

[17]. M. N. Alenezi and Z. M. Alqenaei, "Machine learning in detecting covid-19 misinfor- mation on twitter," Futur. Internet, vol. 13, no. 10, p. 244, 2021.

[18]. S. V. Mahadevkar et al., "A Review on Machine Learning Styles in Computer Vision - Techniques and Future Directions," IEEE Access, vol. 10, no. September, pp. 107293– 107329, 2022, doi: 10.1109/ACCESS.2022.3209825.

[19]. H. O. Boll et al., "Graph neural networks for clinical risk prediction based on electronic health records: A survey," J. Biomed. Inform., p. 104616, 2024.

[20]. H. Gao, Z. Wang, and S. Ji, "Large-scale learnable graph convolutional networks," in Proceedings of the 24th ACM SIGKDD international conference on knowledge discov- ery & data mining, 2018, pp. 1416–1424.

[21]. P. Vyas, G. Vyas, and J. Liu, Proliferation of health misinformation on social media platforms: a systematic literature review., Issues Inf. Syst., vol. 22, no. 3, 2021.

[22]. B. Khemani, S. Patil, D. K. Kotecha, and D. Vora, "Detecting Health Misinformation: A Comparative Analysis of Machine Learning and Graph Convolutional Networks in Classification Tasks," Available SSRN 4715918.