International Journal of Sciences

Research Article Volume 13 – September 2024 (09)

Examining the Role of Machine Learning Algorithms in Flood Monitoring Systems

Junmin Quan[1]

[1]Trinity Grammar School, 119 Prospect Road, Summer Hill, Sydney, New South Wales, 2131, Australia

Abstract: The unpredictable nature of floods often entails the process of monitoring them to be highly sophisticated. Issues with a country's wellbeing and economic stability often ensue, necessitating newer machine learning algorithms like graphical neural networks (GNNs) and logistic regression model. The main purpose of this paper is to compare the role and effectiveness of the two machine learning algorithms in combating flooding. By exploring on the Yangtze River in China and the Banten province in Jakarta, it proposed positive impacts the GNN model and logistic regression model can have in enhancing flood monitoring systems worldwide. Through systematic selection and examination of existing literature, it found that a GNN model operated at 80-98% accuracy in flood prediction,[7] while the logistic regression at 85.05% - 94.39%.[5] Still, while there is room for the sphere of influence and accuracy of these models to be improved, they both have the capacity to contribute positively to existing flood monitoring systems due to the unique benefits they each provide. Overall, the paper seeks to add value into existing literature by making a personalized and substantiated judgement regarding the applicability of the GNN and logistic regression model to flood monitoring systems and their potential for future success.

Keywords: Robotics and Intelligent Systems, Machine Learning, Graphical Neural Networks, Logistic Regression, Mathematics, Yangtze River, Banten Province

Introduction

Background and Context

Floods are one of the most catastrophic phenomena that can occur in the territory of a country; buildings are destroyed, populations are flooded, and enormous economic losses are incurred.[9] The exacerbating frequency of flooding worldwide is primarily attributable to climate change. This process refers to the ongoing increase in global temperatures and the volatility of weather events. Throughout 1998, the flooding at the Yangtze River resulted in the deaths of 1,562 locals and a national economic loss of 255 billion RMB.[6] Similarly, just under 1,000 floods occurred in Jakarta in 2017, causing large-scale population displacement and forced migration internally.[5] Yet, the ubiquitous presence of interconnected data entails the continual advancement of machine learning algorithms to further reduce the human and monetary losses incurred. Contemporarily, several approaches which are believed by existing literature to improve the accuracy and efficiency flood monitoring system have been proposed, such as graphical neural networks (GNN) and logistic regression models.[9] The GNN models the connections between features in a hydrological system,[4,8] while the logistic regression model analyses a large quantity of casual factors of flooding to predict the probability of its occurrence within a geographical area.[1] Although the application of these machine learning algorithms will yield more reliable early flood monitoring, they still comprise issues of being overly complex, reliance on large amounts of data, or an overly simplified prediction.[6] Holistically, the significance of this paper lies in its ability to suggest further improvements to the existing machine learning models used in flood monitoring today, which plays a crucial role in safeguarding a country's socioeconomic security.

Aims and Objectives

The scope of this study pertains to the field of computer science, particularly the theoretical and mathematical aspect of machine learning algorithms. Through the process of a literature review, this paper aims to evaluate the advantages and disadvantages of the GNN and logistic regression model in flood monitoring using empirical case studies and mathematics, discern which of the two models are more accurate and discuss their respective potential for future improvements. This will be achieved through scrutiny of journal articles, credible websites, and books. Additionally, it will make an informed judgement on which of the two models is more reliable comprehensively, and whether have positive prospects. However, due to the broadness of the scope of research and the high quantity of data needed to thoroughly explore it, this paper will only focus on only the two machine learning algorithms mentioned above. Ultimately, it seeks to provide a concise evaluation and comparison of the two machine learning algorithms' effectiveness in flood monitoring and offer an informed insight on the potential rooms for advancements the machine learning models each have.

Search Strategies and Inclusion Criteria

This paper will follow the process of a literature review. The academic literature which will be reviewed will undergo a rigorous and systematic selection process to ensure their relevance to

Published at: http://www.ijsciences.com/pub/issue/2024-09/
DOI: 10.18483/ijSci.2799; Online ISSN: 2305-3925; Print ISSN: 2410-4477

Junmin Quan (Correspondence)
+

computer science, specifically machine learning. Sources used for preliminary research and evaluation of the two machine learning algorithms will strictly be extracted from academic databases and credible journals. However, websites will be occasionally used for relevant diagrams related to the two machine learning algorithms, which does not steer this paper away from its academic focus. To ensure that information is not out-of-date and accurate, all literature reviewed throughout this paper will be published within a decade of its publication. Additionally, the key words: “Machine learning”, “Graphical neural networks”, “AI”, “Logistic regression models”, “Flood monitoring”, “Positive social impact”, “Negative social impact”, “Climate Change”, “Technology” and “Mathematics” are used to assist with obtaining relevant literature in the research process.

Data Extraction, Synthesis Method and Quality Assessment

Data from selected works of literature, including authorial information, data of publication, title, journal name, volume, issue number and DOI will be extracted to properly cite them. Additionally, key qualitative findings, figures and statistics, mathematical explanations and quantitative data like statistics and empirical results that are relevant to the role of GNN or logistic regression model in flood monitoring systems will be extracted for the analysis and evaluation of the two machine learning algorithms. This information will be synthesized through the process of thematic analysis, in which information relevant to each machine learning algorithm will be organized together as a chapter. Additionally, the quality of data will be assessed through its accuracy and reliability. This will be achieved through the process of cross-validation, where key details from a piece of literature will be compared with another of the same topic. As no tools will be used to implement this process, the quality assessment may be prone to the bias and subjectivity of the author.

Discussion

What is a Graphical Neural Network (GNN)?

GNNs are neural networks specifically designed to handle the data structure in graphical format.[3] The primary difference between GNNs and traditional neural networks is its usage of a message-passing mechanism.[4] This mechanism allows it to aggregate complex information from nodes, which are data points that store information about factors that affect the likelihood of a flooding event, including topography, elevation and height.[2] These nodes then connect with other nodes to form mathematical relationships known as edges.[8] Nodes and edges aggregate information through message-passing in a graph structure, whose computational representation resembles a tree diagram (Figure 1).[3] The nodes, upon undergoing message-passing in an edge formation, maps out the spatial dependencies over a geographical region whose data was collected by learning from one another.[3]

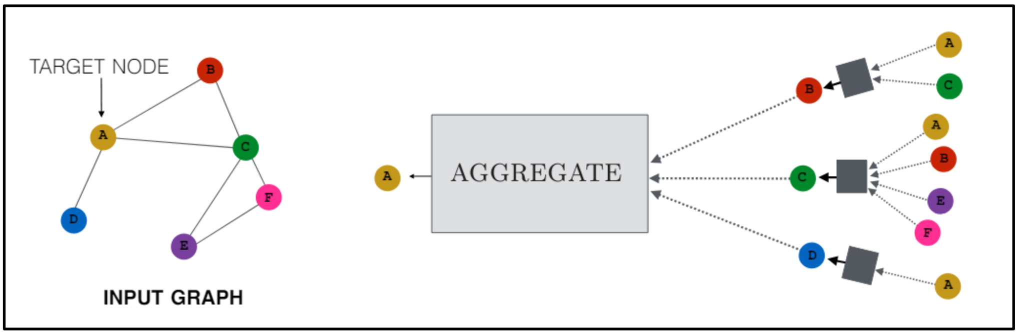

Figure 1: Diagram illustrating the graphical neighborhood of a GNN and how a node aggregates information from neighboring nodes in an edge during an iteration.[3]

In Figure 2, it is evident that after each iteration of the message-passing mechanism, every node in an edge will gain a significant amount of information compared to before, further accentuating the efficiency of the GNN in acquiring information.[3] It is especially important for flood monitoring because of the model’s capacity to manage multiple spatial interactions and co-dependencies in different geographical areas, which becomes increasingly accurate over time as more data are collected between nodes from one another.[2]

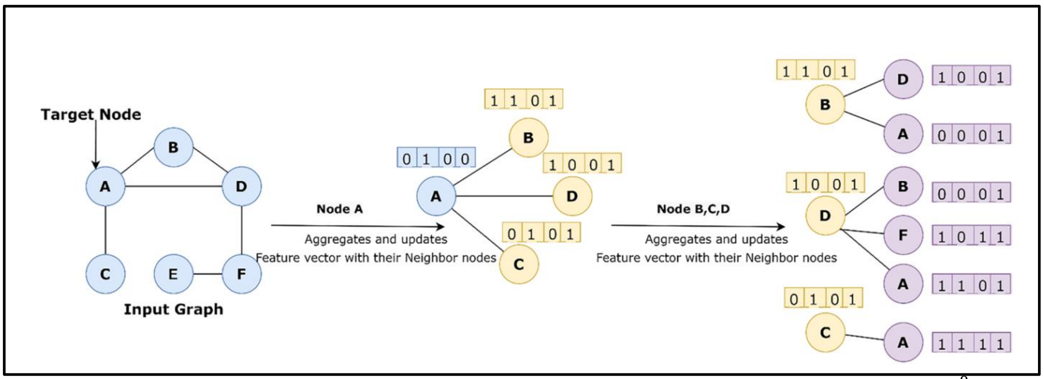

Figure 2: Visualization showcasing the process of message aggregation in a GNN.[8]

Mathematical Explanation of the GNN Model

GNNs operate by propagating information through nodes and edges in a graphical network using its message-passing mechanism.[4] The mathematical expression representing the information aggregated to a node u from its neighboring node v after k iterations by an GNN can be represented as:

$$h_u^{(k+1)} = UPDATE^{(k)}\left(h_u^{(k)}, AGGREGATE^{(k)}(\{h_v^{(k)}, \forall v \in \mathcal{N}(u)\})\right) \quad \text{(Equation 1)}$$

In Equation 1, $h_u^{(k+1)}$ is the updated version of node u after k iterations, $h_u^{(k)}$ and $h_v^{(k)}$ are nodes u and v prior to aggregation, $UPDATE^{(k)}$ and $AGGREGATE^{(k)}$ are neural network functions responsible for message-passing and $\mathcal{N}(u)$ is the graph neighborhood of node u.[4] More specifically, $UPDATE^{(k)}$ is responsible for updating the message of node u after an iteration, whereas the $AGGREGATE^{(k)}$ function is responsible for the aggregation of information to node u from neighboring nodes. $AGGREGATE^{(k)}\left(\left\{h_v^{(k)}, \forall v \in \mathcal{N}(u)\right\}\right)$, which represents the message aggregated to node u from its graph neighborhood $\mathcal{N}(u)$ after k iterations, is commonly expressed as $m_{\mathcal{N}(u)}^{(k)}$ throughout literature for simplicity.[3]

$$h_u^{(k+1)} = UPDATE^{(k)}\left(h_u^{(k)}, m_{\mathcal{N}(u)}^{(k)}\right)$$
(Equation 2)

From Equation 2, the variables $h_u^{(k)}$ and $h_{\mathcal{N}(u)}^{(k)}$ are comprised within the $UPDATE^{(k)}$ function. This means that during each iteration, the GNN synthesizes existing features of node u with messages aggregated from its neighboring nodes.[4] As nodes neighboring node u within the same edge will also undergo a similar process, they collectively acquire more information each iteration. This means that all nodes in a graphical network will aggregate increasing amounts of information the more iterations that occur, making the relationship between interactions and co-dependencies in a geographical area more accurate each time.[3] In the context of flood monitoring, these nodes, after a sufficient number of iterations, will ultimately yield a percentage result predicting a flood's likelihood of occurrence.

Case study of the Yangtze River Basin, China

A study conducted by Ha et al investigated the existing machine learning neural networks currently used in the Yangtze River basin, China, a landform adjacent to over 400 million Chinese inhabitants.[4,6] Over the past three decades, flooding in the Yangtze River is responsible for thousands of fatalities and China's annual loss of 100 billion RMB.[4] Existing hydrological and artificial neural network models used for flood monitoring in the region can merely provide forecasts a month in advance and have a high margin of error[4]. However, albeit not currently being used, the GNN can effectively function by collecting past geographical data related to the rainfall, water level and topography near the river and represent them as nodes.[8] By considering all spatial and temporal dependencies affecting flooding at the river, repeated iterations from the GNN will eventually yield reliable data.[3] Additionally, due to GNN's high scalability, it is a viable and effective option for forecasting flood in a large geographical region like the Yangtze River. Furthermore, as climate conditions vary across different areas of the Yangtze River, separate edges can be formed by the GNN for separate predictions to be made for these areas.[4] Overall, the adoption of the GNN algorithm would greatly mitigate the socioeconomic impacts of flooding of a country and increase the disaster-preparedness of its populace and government. The effectiveness of the GNN is substantiated through a study conducted by Kazadi et al, which showed that over 5 trials, the GNN performed at an accuracy between 80% - 98% in flood monitoring, while a variation of it performed at an accuracy of 63% - 95%, rounded to the nearest percentage.[7]

Advantages of using GNNs for Flood Monitoring

GNNs are better at flood monitoring compared to other machine learning models systems in aspects such meta-learning, scalability, and real-time observation.[8] Unlike existing networks like the convolutional neural networks (CNNs) and recurrent neural networks (RNNs) which cannot handle data structured in graphs, GNNs can perceive and interrelate complex hierarchal spatial relations, and train itself progressively over time after iterations. It can link factors that do not directly relate to one another together and form a relationship from it, which cannot be done by many existing machine learning models.[3] This is beneficial as it allows the precise and timely prediction of flooding, which is crucial for the administration of resources in emergency situations.[8] Moreover, GNN's are capable of handling large volumes of data, considering a multitude of features in a landform influencing the likelihood of flood using its graph structure.[8] Thus, it can process considerable meteorological, hydrological, and topographical data at a large scale, making it scalable and suitable for flood monitoring across large geographical areas.[3] In addition, the utilization of real-time data feeds from local sensors, weather stations, and satellite images, which collect real-time data of geographic regions. This, coupled with the repeated iteration of the GNN, will greatly enhance the promptness of real-time flood detection. Ultimately, the usage of the GNN in flood monitoring allows the local populace of flood-prone regions to become better prepared for the disastrous effects of floods.[3]

Drawbacks of using GNNs for Flood Monitoring

As the GNN relies heavily on defining various nodes to obtain a graph structure, its functionality is contingent on the availability of large and high-quality data, which is highly time-consuming to collect, particularly in developing nations.[8] Its high dependency on a large quantity of data that has undergone cleansing makes them less suitable in lower-income nations who are unable to afford sufficiently advanced technology to run the GNN and struggle with collecting of a sufficient sample of data.[8] Secondly, the effective aggregation of nodes demands considerable labor expertise and capital to implement and maintain due to its complex nature, making it highly costly to operate.[3] Lastly, although the GNN's accuracy increases in an iterative manner, the accuracy of the output data is directly contingent ion whether the original data input accurate or not.[3] If the initial data collection stage is not carefully monitored, the GNN's flood prediction would become highly prone to errors over time, limiting the ability for governments to receive accurate flood forecasts during this period.[3] These drawbacks can be mitigated by first implementing several iterations to the GNN and verifying its accuracy through observation before relying on its predictions allows it to adequately capture the spatial and temporal dependencies of flood-prone regions.

What is logistic regression?

Logistic regression is a statistical tool and parametric classification model belonging to various linear models.[1] It is involved in solving classification problems that involve a dichotomous outcome: division into two contradictory outcomes.[1] It is defined as the ratio of the probability of an event occurrence over its non-occurrence.[1] In the context of flood monitoring, the dependent variable, flood occurrence, must be a binary variable, meaning whether flood is forecasted to occur is dependent on the calculated probability value from 0 to 1, where 0 represents a 'false' outcome and 1 represents a 'true' outcome.[9] Commonly, a result higher than 0.5 concludes that an event is expected to occur and vice versa. This value is later rounded for a binary classification of an event's occurrence to be determined. As seen in Figure 3, the graphic representation of the logistic regression follows a sigmoid shape, which resembles an S-shaped curve.[10] This shape combines various inputs affecting an event's occurrence, and allows the result produced to be directly interpreted as an output value eligible for binary classification to investigate an event's occurrence.[1]

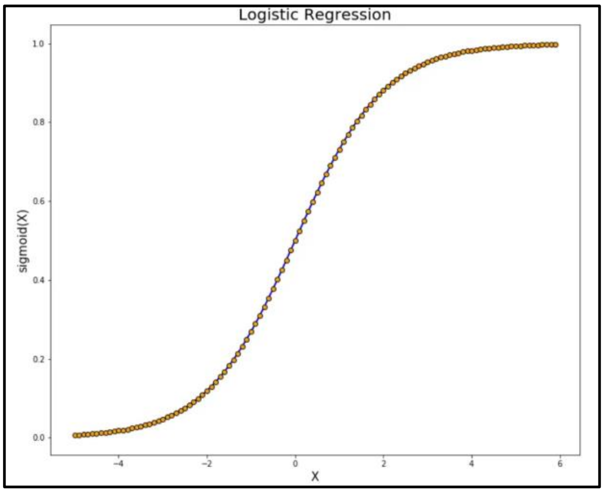

Figure 3: Diagram featuring the graphic representation of the logistic regression model curve, where the vertical axis is the sigmoid and the horizontal axis is any independent variable.[10]

This machine learning algorithm is used in flood warning systems due to its ability to consider multiple independent and predictor variables, which are casual factors such as precipitation, drainage and runoff efficiency.[1] Thus, it is evident that the logistic regression model functions in a similar manner to GNN.[1] Despite the analysis of flood-risk factors and the corresponding contributions using logistic regression models being less sophisticated than GNNs, they are still sufficiently accurate and informative.

Mathematical Explanation of the Logistic Regression Model

According to Al-Juaidi et al,[1] the mathematical relationship linking the probability of event occurrence (p) with independent variables affecting it is known as the logistic function (Equation 3).[1] For n casual factors for flooding, p can be expressed as:

$$p = \frac{1}{1+e^{-Z}}; Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots \dots + \beta_n X_n$$

(Equation 3)

Where Z is the linear combination of casual flood-causing factors, β_0 is the intercept of the mode, β are parameters to the predictor variables X .[1] The predictor variables represent casual factors of flooding and contain a numerical value that is predetermined.[5] As p is a probability, it cannot ever exceed 1 or become negative.[1] This means that theoretically, an infinite number of casual flood-causing factors can be considered in the flood forecasting process, if their respective predictor variables are accurately determined.[5] Thus, it is evident that the logistic regression model allows all spatial and temporal dependencies causing flood to be factored in, making it an effective flood forecasting model.

Case Study: Logistic Regression Prediction of Flood Risk in Banten, Jakarta

Jati et al investigated the use of the logistic regression method in the province of Banten, located westmost of Jakarta, is its special capital region.[5] Throughout history, 46 out of the 250 total occurrences of flooding occurred in this region, leading to mass population displacement and economic losses.[5] This necessitates immediate improvements to flood monitoring to take place. This led to the Indonesian government's utilization of logistic regression, a quick and simple model that could be set up in a short period of time.[1] After modelling 70% of the total number of random flooding events that occurred in this region for a total of 500 times, the predictor variables X of each casual factor at the limits 5%, 50% and 95% were obtained (Figure 4).[5] Using these parameters, the literature concluded that the logistic regression model functioned with 85.05% - 94.39% accuracy.[5] Thus, it is an accurate and reliable flood-monitoring model.[5]

Parameter	Coefficient of Logistic Regression		
	5%	50%	95%
(Interception)	5.38328627721561	8.92077080302076	14.885488352785
Precipitation	-0.00858129471319676	0.0156654046354833	0.0450427170912787
Elevation	-0.00720598863261365	-0.00529532261788892	-0.00416033344926504
Slope	-0.302567439717362	-0.16851258016166	-0.0945001391429277
Coefficient Runoff	-9.09726512880789	-1.27434994746956	4.36197681012406
DNS	-0.00804182415433407	-0.00277991447925938	0.00263320067939322
Flow Accumulation	3.15997403560425e^{-06}	0.000130278891448994	0.00158313585774173

Figure 4: Table of secondary experimental results for the coefficient of logistic regression (β) obtained for each casual factor affecting flood occurrence in Banten.[5]

Advantages of using Logistic Regression for Flood Monitoring

One advantage of the logistic regression model is its intrinsic simplicity and high interpretability. As it ultimately leads to a binary result, it is clear to governments whether flood prevention precautions will take place that day.[5] Additionally, the logistic regression model is highly cost-effective, as it comprises considerably fewer computing demands than the GNN. This makes it viable for use in areas where the technological base is relatively weak.[5] Moreover, another advantage of this machine learning algorithm is its ability to consider unrelated flood-causing casual factors like the GNN given their predictive variables are determined appropriately,[1] making it highly suitable for monitoring floods that are commonly caused by many factors quickly. Together, these factors make the machine learning model highly scalable. Furthermore, using the logistic regression models helps estimate significant factors regarding flood occurrences simultaneously, which allows it to monitor flooding sufficiently accurately, albeit being unsophisticated in nature.[1] Thus, through consideration of these advantages, it can be concluded that the adoption of this model helps improve the disaster preparedness of various countries worldwide.

Drawbacks of using Logistic Regression for Flood Monitoring

One drawback of the logistic regression models is that they might oversimplify the relations between casual factor variables. The model assumes that variables like β and X are, which is not necessarily the case due to the intricate interplay of factors that influence flood occurrence.[5] This leads to unreliable forecasts because it ignores complex interactions between variables, such as the correlation of spatial and temporal dependencies considered by the GNN. As seen from the Jati's Banten case study, over 500 repeated runs of the GNN were conducted to calculate the predictor variables for that geographical area.[5] This shows that by nature, logistic regression models are overly reliant on the quality and completeness of the input data.[5] Therefore, the vigorous demand for data quality may cause the process of implementing the logistic regression to be highly time-consuming, decreasing the lead time of forecast.[5] Additionally, the model mandates a higher sample size of data as it does not undergo meta-learning like the GNN, meaning it is unable to make significant improvements to its forecasting accuracy progressively like that can.[5] However, a mitigation strategy to this is the usage of more advanced computing systems in accordance with the logistic regression model, increasing its efficiency. For instance, it can be used with other neural networks that incorporate advanced computing systems, leading to faster and more accurate forecasts in the future.[2]

Comparison of the GNN and Logistic Regression Models

Table 1: Table featuring the literature values for the accuracies of the GNN and logistic regression models in flood monitoring.[5,7]

Machine Learning Model	Accuracy	Midrange value of accuracy
FloodGNN - Original	80% - 98%	89%
FloodGNN - Variation	63% - 95%,	79%
Logistic regression model	85.05% - 94.39%	89.72%

As seen in Table 1, the results demonstrate that the accuracy of the logistic regression model is a lot less volatile and generally higher than the GNN models.[5] This is reflected through its minimum accuracy of 85.05% and a midrange accuracy value of 89.72%, which is 0.72% more accurate than the GNN model.[7] However, both variations of the GNN model have a higher maximum recorded accuracy, meaning that the GNN is occasionally more accurate than the logistic regression model. Nonetheless, it is believed that the midrange value of accuracy is the most reliable indicator of a machine learning model's overall accuracy. Therefore, it is believed that the logistic regression model, which is shown to be more stable and has a lower margin of error, is generally better for flood prediction. However, a potential limitation of this dataset is that the literature results for the GNNs were only repeated 5 times,[7] making it less reliable compared to the results for the logistic regression models, which were repeated 500 times.[7] Still, it is believed that the two models are both deeply valuable in contributing to the increased accuracy of flood monitoring systems worldwide.[5,7] Through review of literature, it is concluded that the logistic regression model is a more accurate and financially viable machine learning algorithm for flood monitoring than the GNN.

Room for Future Improvement of the Machine Learning Algorithms

The effectiveness of both the GNN and logistic regression depends on the sample size of data they have access to, and the accuracy of the predictions they make with a given sample.[2] By better incorporating AI and Internet of Things (IoT) technology, GNNs and logistic regression models can become better connected with data-collecting devices.[11] The ability of the IoT to exchange allow sensor data to be transmitted immediately online to one another leads to the collection of a larger sample size of real-time data in a shorter span of time, ultimately improving the predictive accuracy of these models.[11] Despite its high short-term expense for emerging economies, the effectiveness of the IoT sensors in data collection is believed to provide them long-term financial benefits by increasing their hazard preparedness.[11] The IoT can also be used in accordance with these machine learning algorithms in smart cities: it helps them better achieve their sustainability-orientated objective by minimizing loss from extreme events like flooding.[11] Moreover, the two algorithms can work with existing flood monitoring models currently used to further validate the reliability of their initial data pertaining to flood occurrence. Lastly, the accuracy of these two models can also be improved through the process of data cleansing, which involves removing inconsistencies and duplications of a data set to improve its quality. By removing outliers and anomalies from the initial data set, the forecasting accuracy of the two machine learning models can be improved.[1]

Conclusion

In conclusion, this paper compared the effectiveness and accuracy of the application of GNN and logistic regression models in flood monitoring systems qualitatively and quantitatively. Examining a vast array of literature accentuated that a geographical region's susceptibility to flooding events are dependent on various factors, which require a of a large sample space to fully evaluate. Through the case studies of Yangtze River, China and Banten, Jakarta, it is discovered that the applicability of the two models is dependent on a country's socioeconomic context, and the size of the geographical region under investigation. Although the GNN model has advantages like meta-learning, scalability and real-time observation, which allow them to generate a structure-aware representation of a geographical area, they are prone to oversimplification, require a large sample size of data to properly function, and will be inaccurate if the initial data is incorrect. Moreover, while the logistic regression model is simpler, cheaper to use and thoroughly considers all casual factors contributing to flooding, they require a larger sample size of data than the GNN, which can establish trends over time alone, to forecast accurately. Through comparison of the two machine learning models, it is concluded that the logistic regression model is a more reliable algorithm for flood prediction due to its lower margins of error and higher midrange accuracy value.

Nonetheless, these findings collectively show that both the GNN and logistic regression models can effectively benefit the modern flood monitoring systems soon. The paper's suggestion of future improvements for these models contributes to existing literature by adding a personalized perspective into this field of study. Additionally, the quantitative findings from existing literature summarized in this journal offer a valuable insight to the different accuracies of these algorithms to be considered by nations and urban planners when establishing and enhancing flood monitoring systems. However, a major difference between the journals reviewed is the greater availability of journals pertaining to GNN than to logistic regression, which led to a slight imbalance in the exploration of these machine learning algorithms.

One realistic improvement to this investigation is the comparison of the two modes using specific individual metrics, which would allow a more focused and informed evaluation to take place regarding the better machine learning model for flood monitoring. An area of research which could be explored to add to the existing scope of research is the IoT. For example, devices like IoT sensors can yield more reliable real-time data when incorporated with the GNN and logistic regression, which can effectively interpret and analyze its intrinsically complex data. Additionally, the research can also investigate collecting first-hand data. This would mitigate the lack of reliability for the accuracy of GNN results. Overall, the research compared and evaluated the GNN and logistic regression models in improving existing flood monitoring methods today, and the continued potential for greater impact they each have, meeting all its objectives.

An evident limitation of this paper is its lack of primary experimental results, which limits the exclusivity and innovativeness of the claims presented. This, coupled with its highly theoretical nature, limits the ability for the paper's findings to be compared with different existing literature quantitatively, which can often yield more substantiated results. Additionally, a limitation in the paper's research method is the lack of sufficient literature comprising quantitative data surrounding this area of study. This hinders the quality and consistency of data compared, limiting its ability to offer insightful contributions to this field. Personal insights may therefore be prone to personal bias. This decreases the depth and reliability of the exploration.

Acknowledgements

I express my extreme gratitude to my mentor Maher Uddin — Master student at Cambridge University's Information and Computer Engineering program. His specialty in statistical signal analysis and machine learning provided me with invaluable guidance in writing this paper. I would also like to thank Indigo Research for providing me with the opportunity to investigate my field of interest. Additionally, I would like to thank my family for their continued support throughout the course of this study.

References

1. Al-Juaidi, A. E. M, Nassar, A.M, Al-Juaidi, O. E. M. Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors. *Arabian Journal of Geosciences,* 11, 1-10 (2018). https://doi.org/10.1007/s12517-018-4095-0
2. Bentivoglio, R, Isufi, E, Jonkman, S. N, Taormina, Riccardo. Rapid spatio-temporal flood modelling via hydraulics-based graph neural networks. *Hydrology and Earth System Sciences.* 27, 4227-4246 (2023). https://doi.org/10.5194/hess-27-4227-2023
3. Hamilton, W. L. Graph Representation Learning. *Morgan & Claypool.* 14, 1-159 (2020). https://www.cs.mcgill.ca/~wlh/grl_book/
4. Ha, S, Liu, D, Mu, L. Prediction of Yangtze River streamflow based on deep learning neural network with El Niño–Southern Oscillation. *Scientific Reports.* 11, 1-23 (2021). https://doi.org/10.1038/s41598-021-90964-3
5. Jati, M. I. H, Suroso, Santoso, P. B. Prediction of flood areas using the logistic regression method (Case study of the provinces Banten, DKI Jakarta, and West Java). *Journal of Physics: Conference Series.* 1367, 1-14 (2019). https://iopscience.iop.org/article/10.1088/1742-6596/1367/1/012087
6. Ke, W, Ouyang, C, Li, Y, Chen, M, Ma, J, An, H, Zhou, S. Reflections on the Catastrophic 2020 Yangtze River Basin Flooding in Southern China. *Innovation.* 1, 1-2 (2020). 10.1016/j.xinn.2020.100038
7. Kazadi, A, Doss-Gollin, J, Sebastian, A, Silva, A. Flood Prediction with Graphical Neural Networks. Rice University. (2022). https://cs.rice.edu/~al110/pubs/ccai-neurips22.pdf
8. Khemani, B, Patil, S, Kotecha, K, Tanwar, S. A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions. *Journal of Big Data.* 11, 1-43 (2024). https://doi.org/10.1186/s40537-023-00876-4
9. Mane, P, Katti, M, Nidgunde, P, Suve, A. Early Flood Detection and Alarming System Using Machine Learning Techniques. *International Journal of Research in Engineering, Science and Management.* 3, 29-32 (2020). https://journal.ijresm.com/index.php/ijresm/article/view/330
10. Thorn, J. Logistic Regression Explained. Medium. (2020). https://towardsdatascience.com/logistic-regression-explained-9ee73cede081
11. Kumar, Sachin, Tiwari, Prayag, Zymbler, Mikhail. Internet of Things is a revolutionary approach for future technology enhancement: a review. *Journal of Big Data.* 6, 111 (2019). https://doi.org/10.1186/s40537-019-0268-2