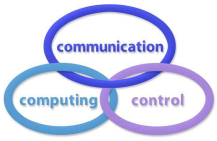
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 18, Issue: 1, Month: Feburary, Year: 2023
Article Number: 4478, https://doi.org/10.15837/ijccc.2023.1.4478



CCC Publications

Semantic Graph Based Convolutional Neural Network for Spam e-mail Classification in Cybercrime Applications

S. Rahmath Nisha, S.Muthurajkumar

S. Rahmath Nisha
Department of Computer Science and Engineering, K. Ramakrishnan College of Technology,
Samayapuram, Trichy-621112, Tamilnadu, India
Corresponding author: rahmathnishas.cse@krct.ac.in

S. Muthurajkumar
Department of Computer Technology, Madras Institute of Technology(MIT) Campus,
Anna University, Chrompet, Chennai-600044, Tamilnadu, India
muthurajkumarss@gmail.com

Abstract

Spam is characterized as unnecessary and garbage E-mails. Due to the increasing number of unsolicited E-mails, it is becoming more and more crucial for mail users to utilize a trustworthy spam E-mail filter. The shortcomings of spam classifier are defined by their increasing inability to manage large amounts of relevant messages and to effectively detect and effectively detect spam messages. Numerous characteristics in spam classifications are problematic. Given that selecting features is one of the most often used and successful techniques for feature reduction, it is a crucial duty in the identification of keyword content. As a result, features that are unnecessary and pointless yet potentially harm efficiency would be removed. In this study, we present SGNN-CNN (Semantic Graph Neural Network With CNN) as a solution to tackle the difficult task of mail identification. By projections E-mails onto a graph and by using the SGNN-CNN model for classifications, this technique transforms the E-mail classification issue into a graph classification challenge. There is no need to integrate the word into a representation since the E-mail characteristics are produced from the semantic network. On several open databases, the technique's effectiveness is evaluated. Some few public databases were used in experiments to demonstrate the high accuracy of the proposed approach for classifying E-mails. In term of spam classification, the performance is superior to state-of-the-art deep learning-based methods.

***Keywords*:** Spam E-mail classification, Convolutional Neural Network, Semantic Graph, Graph Neural Network

1 Introduction

Unwanted spam messages have grown to be a major issue on the Internet in recent times. Spam messages not just to eat up a lot of available bandwidth, but they also take up users' processing time. Spyware programmed that collect sensitive information and send it to marketers as well as other third - party may be included in certain spam messages. Therefore, there is a critical need for the creation of more effective filters that can recognize these messages autonomously.

The term "feature" refers to a set of characteristics that measure certain elements of an E-mail user's activity or behavior [25], [26], [27], [28]. To create precise and effective classifiers, it is crucial to extract and choose essential characteristics for E-mail classification. The "bag of words" model, in which each location in the input feature vector corresponds to a specific word or phrase, was employed by researchers to classify E-mails. For instance, the use of the term "free" may be a helpful characteristic in identifying spam E-mail. As a result, well-chosen features may significantly increase classification accuracy while also lowering the volume of data needed to achieve the necessary performance. The following features are the most popular ones:

1. **E-mail Header Features:** From the header of an E-mail, E-mail header features are retrieved and chosen. The from, to, bcc, and cc fields are found in a header. For instance, the subject line of E-mails with phrases like "banking," "debt," "Fwd.: Re:" and "confirm" are often used to identify phishing E-mails. Other instances include the subject's length in character, words, and from field words, as well as the sender's E-mail addresses non-model subdomain.
2. **E - mail Content Features:** The E-mail body segment, which includes the E-mail's primary content, is where E-mail body features are chosen from. Examples of E-mail body characteristics used to identify phishing E-mails include HTML content, HTML forms, the word "dear," the number of letters and utterances, word structure (such as "lending," "select," "record," "recognize," and "information," etc.), the word "suspended," and the term "verify your account."

Various tactics for classifying E-mail spam have been suggested by certain academics. Decision Tree, Support Vector Machine, and Naïve Bayes [1], [2], [3] approaches are a few of these approaches. The mails used in these conventional approaches frequently need to have characteristics manually extracted as embedding vectors before being fed into in the classification model. Additionally, the Convolutional Neural Network (CNN) [4] was used in recent studies to classify spam E-mails ([5-8]).

There is no requirement to individually feature extracted from the E-mails since extracted features and classifications in CNN models is done automatically throughout the entire model. The embedded vectors are used as inputs both by the conventional and CNN algorithms. The major contributions of the paper is follows:

1. The E-mail classifying issue is transformed into a graph classification problem by the technique we provide in this study. In contrast to other approaches, our technique skips the step of incorporating the E-mail content into the numeric column vector.
2. Instead, the approach converts the word into a graph and classifies the spam E-mail using Graph Neural Network (GNN) and Convolutional Neural Network (CNN). When compared to some few public datasets, the suggested architecture had a greater accuracy for E-mail classification tests.

The structure of this essay is as follows. Problem statement is described in Section 2. The related work involving rule-based technique and deep learning-based approach is described in Section 3. Section 4 goes through our suggested method for using the GNN to categorise spam E-mails. Section 5 goes into further detail about the studies, which include preprocessing, building and using a graph neural net, testing that on samples, and evaluating its effectiveness. Section 6 serves as the paper's conclusion.

2 Problem Statement

For the spam E-mail classification, we take that e ∈E a predetermined set of subclasses and a high-dimensional E-mailing spaces (E) model of an E-mails C ={c_1, c_2, …, c_i}. There are just two different classes in this assignment: ham and garbage. A training set T of labelled E-mails is provided to us $<e,\ c>$, where $<e, c> \in E \times C$, For example:

$$<e,\ c> = <\mathit{Congratulations},\ \mathit{claim\ your\ free}\ \$100\ \mathit{gift\ card},\ \mathit{spam}> \tag{1}$$

We want to create a classifier model that translates messages to labels using machine learning supervised learning.

$$\text{œ: E} \rightarrow \text{C(2)} \tag{2}$$

The supervised approach to machine learning is indicated $\mathcal{L}$ by and write $\mathcal{L}(\text{T}) = \sigma$. the approach of supervised machine learning L provides the figured out classifier model after taking the trained set T as inputs.

3 Relate Work

This section provides a thorough introduction to related research on E-mail classifications. Rule-base technique and deep learning-based approach are our two ways for summarizing the relevant work.

3.1 Rule-based Method

Many academics have explored rule-based mail recognition methods that are based on SVM and Naive Bayes technologies in order to efficiently address the danger presented by spam E-mails. For E-mail classification, Rathod et al. suggested the Naive Bayes method. The suggested solution employs characters using ham and malware to determine if a letter is likely to be spam or not [9]. Open-source E-mail spam filters make extensive use of the Naive Bayes approach [10]. It is not sensitive to superfluous characteristics. Because Nave Bayes often requires less quick evaluation and training to recognise and screen a phishing E-mail, this is the cause. Rusland et al. employ the WEKA [11] programme, which is based on the Spambase and Spam datasets, to evaluate the Nave Bayes algorithm for classifying E-mail spam. The results of the experiment demonstrated that the Nave Bayes' effectiveness was influenced by the dataset's example count and E-mail type [12]. Over the years, Support Vector Machines have also established themselves as one of the best Classification techniques. A Naive Bayes filtering approach that is based on Support Machine Vector was suggested by Feng et al. The interdependence assumptions between the features taken from the initial training dataset is intended to be disproved when the Naïve Bayesian approach is used. According to experiment results, this technique can identify spam more accurately and with a quicker classification speed [13]. In order to filter spam, Vishagini et al. suggested using a scaled Support Vector Machine using the weight variables discovered by the KFCM technique. The relevance of various categories is reflected in the weight variable. The misinterpretation of E-mail is decreased by increasing the weight value. The classification of E-mails may be decreased by increase in weight value. Studies reveal that there is still room for improvement in the precision and accuracy of the spam detection system's performance [14]. The methodologies of ACO and SVM were combined and used to create the spam categorisation approach that Karhika et al. disclosed. The suggested approach uses a hybrid approach. The selection of the characteristics is crucial to the models. terms of precision, accuracy, and recall, the investigation demonstrates that the proposed method was preferable to a number of the most cutting-edge classification approaches [15]. The approach of Support Vector Machines has the benefit of great precision. This approach often takes longer than other approaches.

3.2 Deep Learning-based Method

The machine learning applications of object detection, object tracking, and picture classifications have subsequently shown CNN to be quite successful. To address the spam detection issue, several researchers have used CNN. Bagui et al. suggested a technique that uses deep learning technology to identify the intrinsic characteristics of E-mail in classifying E-mails either spoofing or non-phishing. Researchers categorize E-mails utilizing a deep-learning algorithm and do deep semantic evaluation using one-hot encoding with and without words. Additionally, they evaluated the precision of several deep and machine learning techniques both with and without phrases[16]. Seth et al. employ a CNN to analyse the complete content (i.e., text and photos), then run it via a separate classifier to determine if the mail is spam or ham. They suggested two hybrid multi-modal designs. In order to distinguish between spam and legitimate E-mail, the architectures gathering the input from those two separate models integrated the output data. Studies reveal that the proposed technique performs the classification job with higher accuracy than the standalone image and textual classifications [17]. Alghoul et al. provide an artificial neural networks model for mail identification. To use a feedforward backpropagation technique, the model has been trained. Then, provided the data for this model. This study demonstrates how artificial neural networks may be used to classify E-mails [6]. Another spam identification model dubbed THEMIS was introduced by Soni et al. They employed an imbalanced data set with a decent mix of legitimate and phishing E-mails to evaluate THEMIS' suitability. The THEMIS model produced good results in the trials [7]. A scalable and robust content-based malware detection network dubbed DeepSpamNet was presented by Srinivasan et al. as a framework for network threat positional awareness. Because there are no stages involved in feature extraction, deep learning allows for quick change of the different nature of spammer. Research demonstrates that deep learning models perform better than traditional machine-learning classifiers [8]. The CNN's self-learning capabilities and dependable fault - tolerant make it valuable.

4 System Model

In this study, we projected E-mails onto a network to transform the E-mail classification issue into a graph classification task. There is no need to integrate the word into a numeric input vector since the E-mail characteristics are produced from the semantic network. The proposed methodology transforms the classification issue for spam E-mails into a graph classification problem. The proposed solution is divided into four main steps, including data prepressing, graph creation, training and evaluation of graph neural networks, and graph classification, as illustrated in Fig. 1. The dataset has to be manually cleaned using data preparation methods since it is noisy and imbalanced. Then, we construct a sizable graph made up of word and E-mail document nodes. Each node has embedding vectors depending on the characteristics of its neighbours. After creating the graphs, we input it to the GNN so it can learn high-dimensional characteristics. Lastly, we use a neural network convolutional to the E-mail classification issue to create a graph classification based mostly on E-mail message and words graphs.

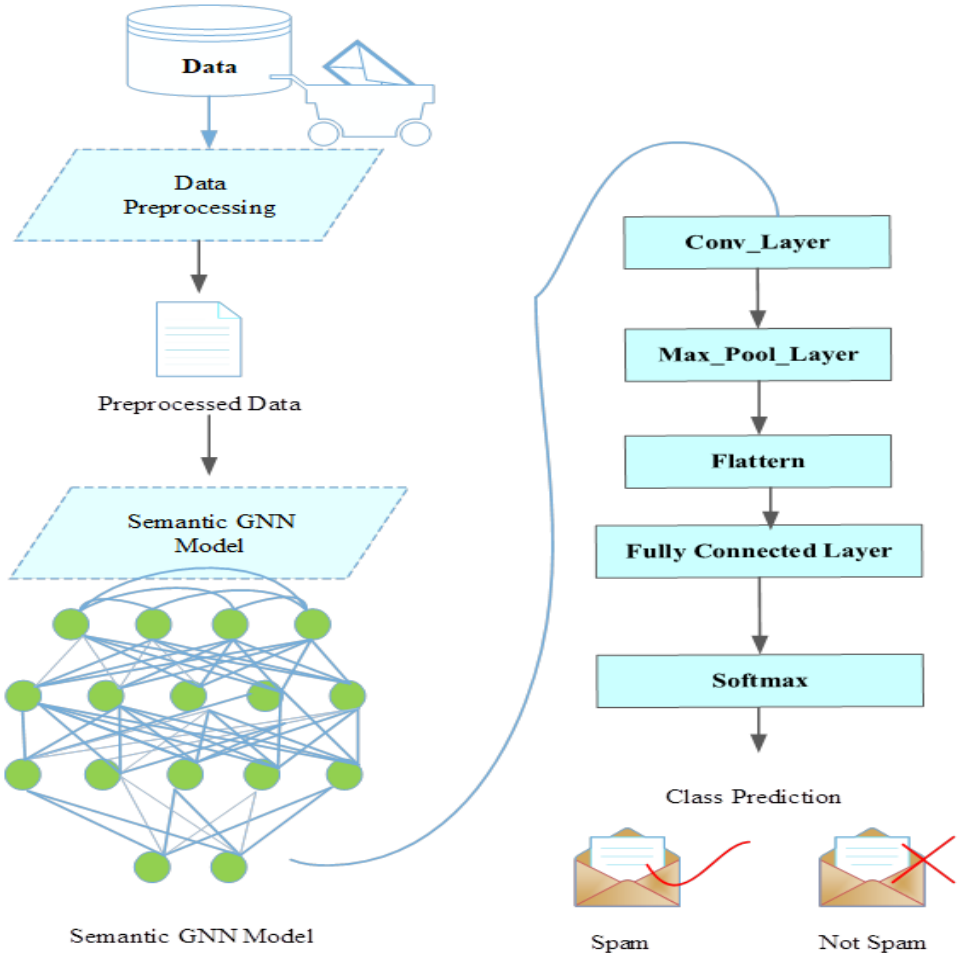


Figure 1: System Architecture SGNN-CNN

4.1 Data Prepressing

Preparing the input is required to convert E-mails from humans to machines usable formats for the further analysis. Stop - word, special characters, and all symbols are changed to lowercase letters as part of a sequence of procedures which we carry out.

The E-mail collection was assigned to either the spam or ham class in accordance with the standards given by the E-mail spam classification. Our three pieces of information are as follows: 70% of the information will be trained. 10% of the data will be data set, and 20% will be the verification set.

4.2 Building Graph

We construct an e - mail text network including E-mail word nodes, subject nodes and documents nodes, in classifying phishing messages. As stated as in graph:

$$\mathrm{G} = (V, E) \tag{3}$$

$$\mathrm{V} = \{word \mid text \mid topic\} \tag{4}$$

$$\mathrm{E} = \left\{ \begin{array}{c} e_{ij} \mid e_{id} \mid e_{jd}, iisword, \\ jistext, disdomaintopic \end{array} \right\} \tag{5}$$

where V stands for the node set(s). E indicates groups of edges. Word nodes, E-mail text nodes, and subject nodes are the three different types of nodes. Utilize the Linear Discriminant analysis Allocation (LDA) model to deduce the domain subject from the E-mail documents. The semantic similarity architecture of the corpus may be clustered using the generating probabilistic model LDA. We use LDA to assist us in automatically finding subjects that are present in textual information.

LDA generates a topic term joint distribution for each subject d. The mutual information of issue mixing, a set of N words, a set of N topics, and the parameters, and, are given by:

$$\rho(\theta, z, \omega \mid \alpha, \beta) = \rho(\theta \mid \alpha) \prod_{n=1}^{N} \rho(z_n \mid \theta) \rho(\omega_n \mid z_n, \beta) \tag{6}$$

The word-word vertices, word-text line segments, topic-word edges, and topic-text corners make up the graph's corners. The LDA topic model learns the weights of the topic-word vertices as well as the topic-text vertices. We use the Point - wise Mutual Knowledge to compute the term weights (PMI). The purpose of PMI is to measure the chance that two terms will appear together. Strong meaning connection between words is indicated by a high PMI value, and weak semantically connection is shown by a low PMI scoring system. The PMI calculation involves,

$$\mathrm{PMI}(a, b) = log2^{\frac{\rho(a,b)}{\rho(a)\rho(b)}} \tag{7}$$

$$\rho(a) = \frac{W(a)}{|W|} \tag{8}$$

$$\rho(a, b) = \frac{W(a, b)}{|W|} \tag{9}$$

In which a and b are two words in a couple. The quantity W(a,b) is the amount of sliding window frames that have the letters a and b in them. W(a) is the quantity of slide window in the corpora that solely contain the letter a. We do not include connections with low PMI values; rather, we exclusively maintain vertices with high PMI values. To determine the weight of the connection here between word and the text, we use the BM25 method [19]. A series of papers are sorted using the bag-of-words search feature (BM25) in accordance with the search keywords. The document d's BM25 score, given a query term, is:

$$\mathrm{relevant}_{\mathrm{score}(w,\ doc)} = \sum_{i=1}^{n} IDF(q_i) \times \frac{TF(q_i) \times (k_1+1)}{TF(q_i) + k_1 \times \left(1 - b + b \times \frac{|doc|}{ave\_len}\right)} \tag{10}$$

Where $IDF(q_i)$ is q_i 'the document's inverse document regularity. Divide the overall amount of documents by the amount of words in the document that include the phrase to get the appears in a document frequency. It is a quantitative statistic that indicates whether a phrase is frequent or uncommon in a certain corpus of documents. $\mathrm{TF}(q_i)$ is q_i 's term frequency.

The number of instances a word occurs in a particular text is indicated by the frequency distribution. $|doc|$ is the length of the document doc. ave_len is the length of the article on averages. b and k_1 are free parameters.

4.3 SGNN-CNN Mechanism

Through projected the E-mail content onto the graph using the created graph model, we change the mail classification issue into a graph node classification problem. Furthermore, it was shown that Graph Neural Networks performed convincingly on such a challenge [20–21]. It is suggested to use Graph Neural Network to cooperatively aggregate data from graph structure. GNNs maintain a state that, in contrast to conventional neural nets, may represent data from its neighbourhood with any depth. h v, which is a state encoding, is what the GNN is designed to study.

$$h_v = f\left(x_v, x_{co[v]}, h_{ne[v]}, x_{ne[v]}\right) \tag{11}$$

$$\mathrm{Y} = \mathrm{softmax}\left(\mathrm{H}\sigma\left(H\sigma\left(HXW^{(1)}\right)W^{(2)}\right)W^{(3)}\right) \tag{12}$$

where σ is the activation function $Relu\left(x_i\right) = max\left(0, x\right)$. Y represents the classification engine's outcome. $W^{(1)}$, $W^{(2)}$, and $W^{(3)}$ are learned using gradient descent on weight matrix. $H = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ denotes Laplacian Matrix, $\tilde{A} = A + I.A$ is an identity matrix, whereas is a mathematical expression. $\tilde{D}$ is the degree matrix of $\tilde{A}$.

Minimizing the cross-entropy loss between both the preceding label and ground truth labeling is the goal of training. The following definition of the gradient descent:

$$loss = -\sum_{l \in Y_L}\sum_{f=1}^{F} Y_{lf} ln Z_{lf} \tag{13}$$

where F is the output feature's dimension, Y_L is the true indicator Y the matrix of label indicators; and Z represents the matrix for output.

5 Experiments

Under this section, we run a number of tests on 3 different datasets TREC Spam, Spambase, and Enron-Spam Datasets—to assess the performance of our proposed SGNN-CNN.

1. **Enron-Spam Data:** The Fed Energy Regulator got the Enron-Spam set of data when looking into the demise of Enron. There are over 500,000 mails produced by Enron workers in it [22].
2. **The Spambase Dataset** concentrates on classifying spam-like E-mails or not spam by keyword or word frequencies. The data has 58 characteristics and 4601 instances. For predictions, it has the variables "Spam" and "Not Spam." It is a multidimensional, actual dataset that is mostly employed for feature identification. George Forman provided the database, which was created at Hewlett-Packard Labs [23].
3. **TREC E-mail Data:** Every E-mail in this set of data is classified as spam or not spam using a temporal indexing. There are 92,189 E-mail messages in the database. 39,399 messages are classified as ham, whereas 52,790 are classified as spam [24].

Every experiment requires dynamically dividing the input graph's nodes into train, verification, and testing set, each of which has the same total number of nodes. The comparison between our method and the most sophisticated deep learning-based E-mail classification model is shown in Table 1. It is evident that our model routinely surpasses the most recent model by a margin of much more than three percentage points. The major factor for SGNN-CNN's success is its ability to record both word-to-word and word-to-topic relationships. Another benefit of our approach is how straightforward and reliable our text-to-graph projection is in practice. Even more so, users are not required to carry out laborious data preparation. The empirical SGNN-CNN findings demonstrate that the word-topic data may enhance the classifying impact of E-mail.

The amount of positive class classification instances (TP), clear negative class instances (TN), fake correctly classified cases (FP), and false negative class cases may be used to calculate the accuracy of the classifier (FN). In the instance of binary classification issues, these numbers constitute a confusion matrix, as seen in Fig. 3. TP is the percentage of instances that are accurately identified as belonging to the "SPAM" class but are actually projected to be "NOT SPAM." TN is the percentage of accurately identified occurrences that fall into the "NOT SPAM" class and are anticipated to be "NOT SPAM." FP is the percentage of occurrences that are erroneously assigned to the "SPAM" class but are really anticipated to be "NOT SPAM." The percentage of occurrences that should have been classed as "NOT SPAM" but were instead misclassified as "SPAM" is known as the "FN" rate.

On several test dataset, the classification performance of our model at various feature levels is shown. The high-performing for the Enron-Spam, Spambase, and TREC Spam Datasets are shown in Figures 4, 5, and 6 respectively. We describe the outcomes of three dataset's worth of mail classification tasks using feature dimensions ranging from 32 to 1024. On all three databases, evaluate the accuracy rises as feature size does. The findings demonstrate that SGNN-CNN is stable if the feature size exceeds 256. Additionally, it was shown that various dataset with varied amounts of characteristics had various classifications outcomes.

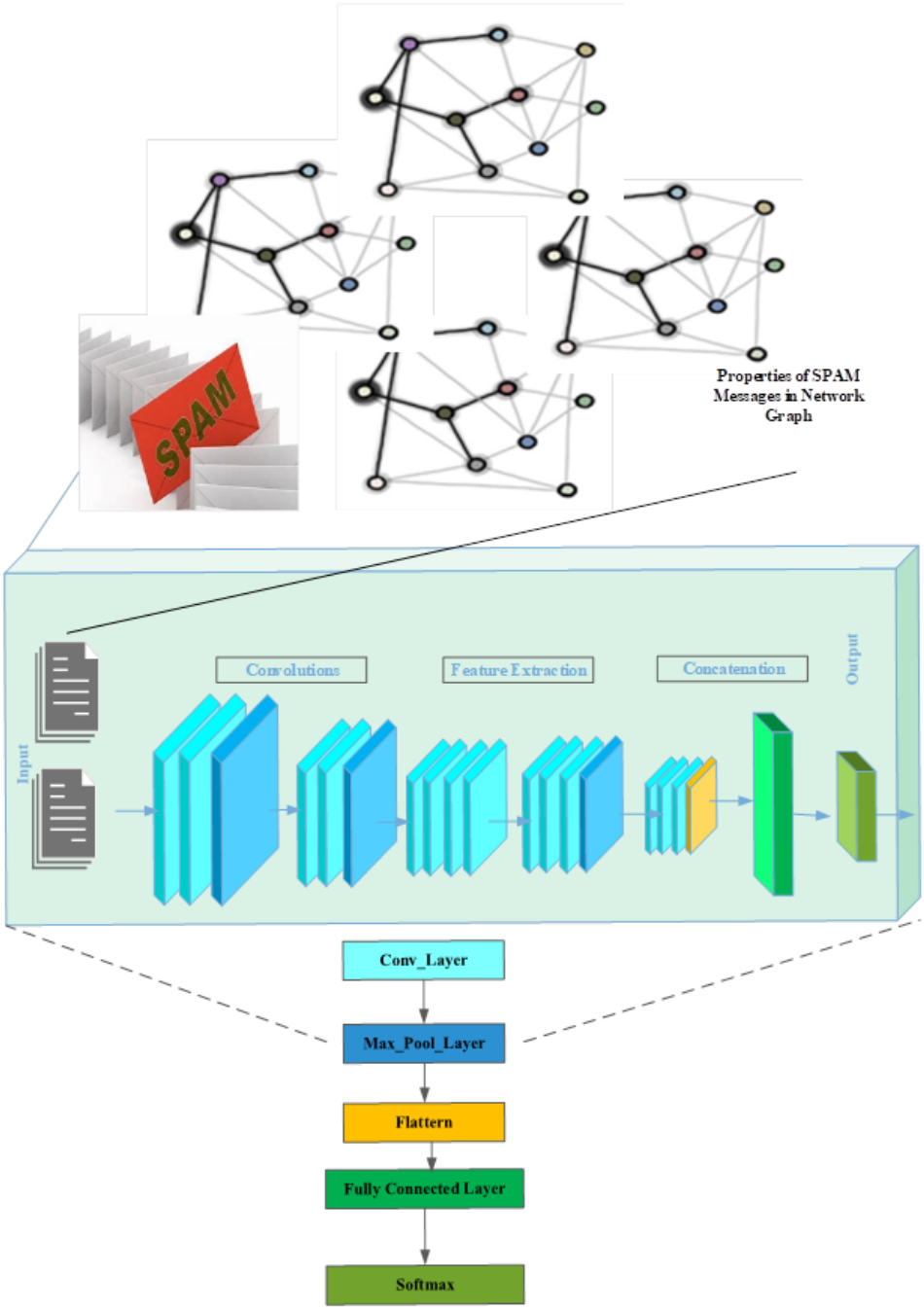


Figure 2: Semantic GNN with CNN Architecture

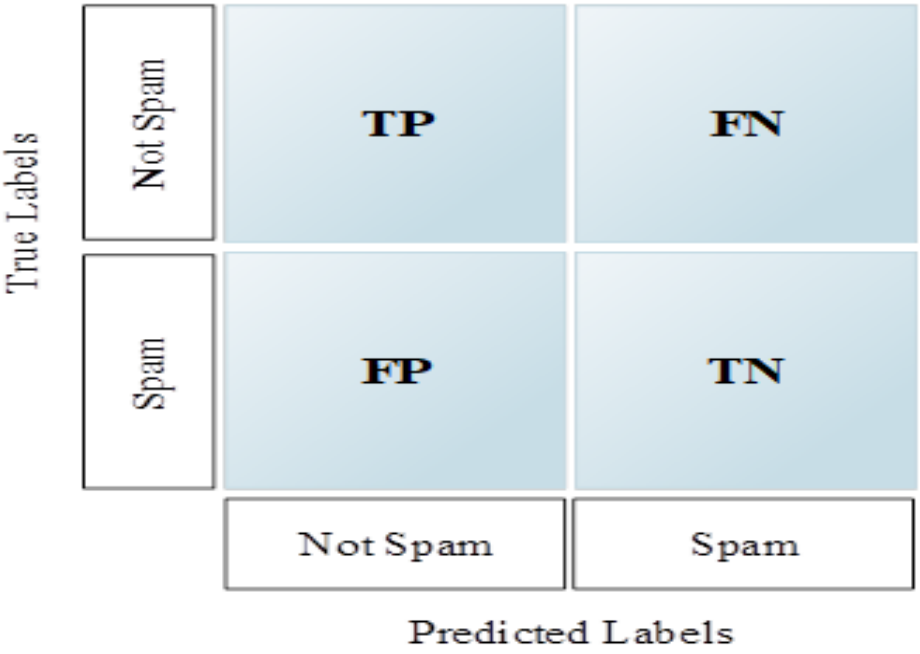


Figure 3: Confusion Matrix for Proposed Work

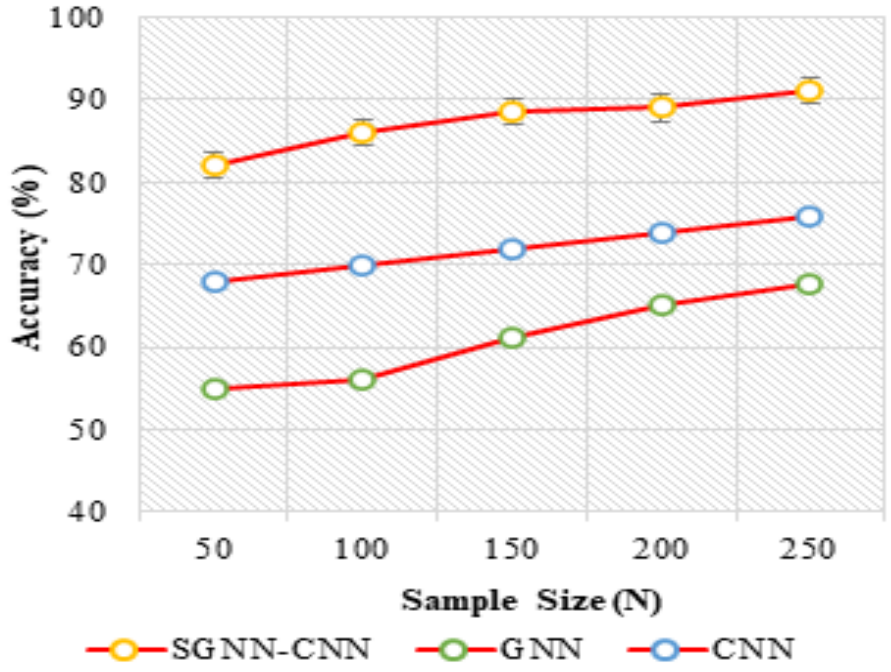


Figure 4: Evaluation of Accuracy for TREC Spam Dataset

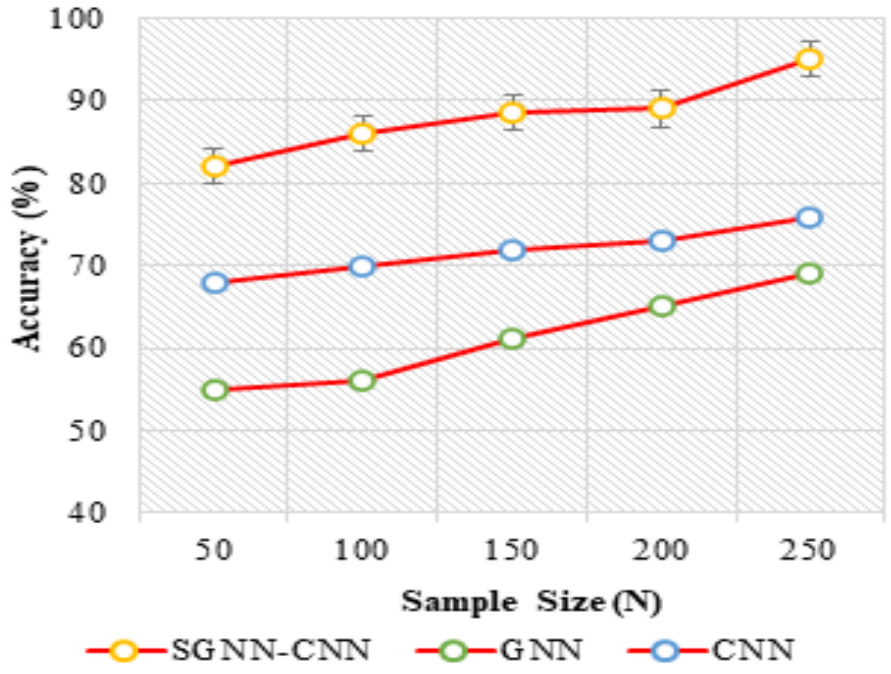


Figure 5: Evaluation of Accuracy for Spambase Dataset

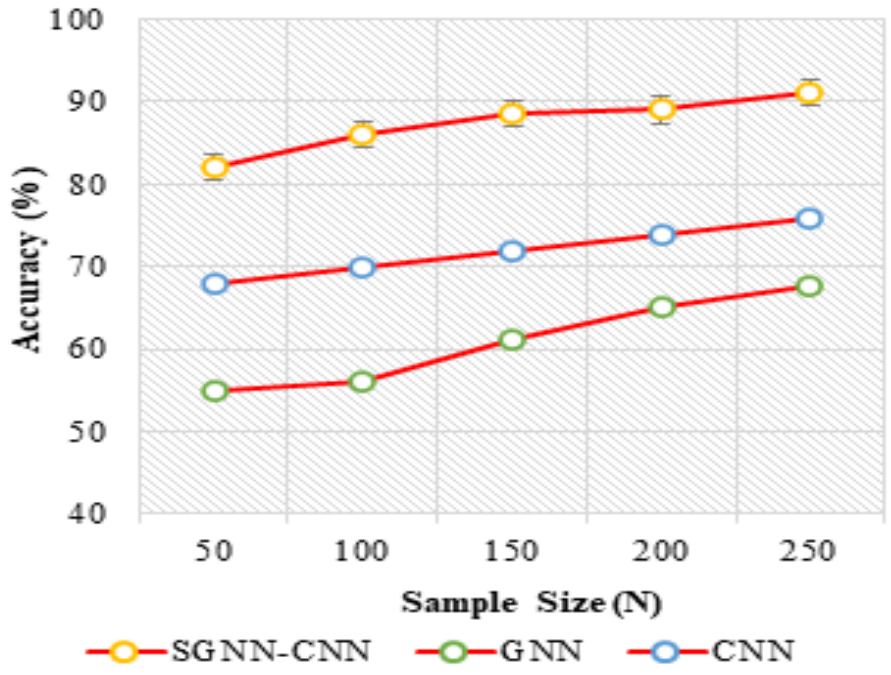


Figure 6: Evaluation of Accuracy for Enron-Spam Dataset

6 Conclusion and Future Work

In this study, we provide an SGNN-CNN approach for E-mails classification. The mail identification issue is transformed into a graph classification task, and the GNN framework is then used to categories the mail. The GNN algorithm extracts the E-mail's properties in an attractive manner. Researchers have used many open datasets to evaluate our methodology. The testing findings shown that, throughout term of spam identification, our effectiveness is superior to that of the most advanced deep learning-based system. To further increase the precision of the proposed approach, we may use a variety of preprocessing techniques including word clarification as well as other approaches in future study. Purely text-based mail keep spamming may currently be detected using presented approach. In the future, we want to expand our strategy and adapt it to spam detection using hybrid deep learning techniques.

References

[1] Harisinghaney, Anirudh, et al. "Text and image based spam E-mail classification using KNN, Naïve Bayes and Reverse DBSCAN algorithm." International Conference on Reliability Optimization and Information Technology (ICROIT). IEEE. 2014.

[2] Sharaff, Aakanksha, and Harshil Gupta.. "Extra-tree classifier with metaheuristics approach for E-mail classification." Advances in Computer Communication and Computational Sciences. Springer, Singapore, 2019. 189-197.

[3] Bouguila, Nizar, and Ola Amayri. "A discrete mixture-based kernel for SVMs: application to spam and image categorization." Information processing & management 45.6 : 631-642.

[4] Derhab, Abdelouahid, et al. 2020 "Intrusion Detection System for Internet of Things Based on Temporal Convolution Neural Network and Efficient Feature Engineering." Wireless Communications and Mobile Computing . 2009.

[5] Cao, Yukun, Xiaofeng Liao, and Yunfeng Li. "An e-mail filtering approach using neural network." International Symposium on Neural Networks. Springer, Berlin, Heidelberg. 2004.

[6] Alghoul, Ahmed, et al. "E-mail Classification Using Artificial Neural Network." International Journal of Academic Engineering Research, Vol. 2 Issue 11. 2018.

[7] Soni, Ankit Narendrakumar. " Spam e-mail detection using advanced deep convolution neural network algorithms." Journal for innovative development in pharmaceutical and technical science 2.5 : 74-80. 2019.

[8] Srinivasan, Sriram, et al. "Spam E-mails Detection Based on Distributed Word Embedding with Deep Learning." Machine Intelligence and Big Data Analytics for Cybersecurity Applications. Springer, Cham,. 161-189. 2021.

[9] Rathod, Sunil B., and Tareek M. Pattewar."Content based spam detection in E-mail using Bayesian classifier." International Conference on Communications and Signal Processing (ICCSP). IEEE. 2015

[10] Androutsopoulos, Ion, Georgios Paliouras, and Eirinaios Michelakis. Learning to filter unsolicited commercial e-mail. " DEMOKRITOS", National Center for Scientific Research. 2004.

[11] Hall, Mark, et al. "The WEKA data mining software: an update." ACM SIGKDD explorations newsletter 11.1 : 10-18. 2009.

[12] Rusland, Nurul Fitriah, et al. "Analysis of Naïve Bayes algorithm for E-mail spam filtering across multiple datasets." Proceedings of the IOP Conference Series: Materials Science and Engineering. 2017.

[13] Feng, Weimiao, et al. "A support vector machine based naive Bayes algorithm for spam filtering." 2016 IEEE 35th International Performance Computing and Communications Conference (IPCCC). IEEE. 2016.

[14] Vishagini, V., and Archana K. Rajan. "An improved spam detection method with weighted support vector machine." International Conference on Data Science and Engineering (ICDSE). IEEE. 2018.

[15] Karthika, R., and P. Visalakshi. "A hybrid ACO based feature selection method for E-mail spam classification." WSEAS Trans. Comput 14 : 171-177. 2015

[16] Bagui, Sikha, et al. "Classifying Phishing E-mail Using Machine Learning and Deep Learning." International Conference on Cyber Security and Protection of Digital Services (Cyber Security). IEEE. 2019

[17] Seth, S., & Biswas, S. , "Multimodal spam classification using deep learning techniques." In 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems. IEEE. 2017

[18] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." the Journal of machine Learning research.3 : 993-1022. 2003

[19] Trotman, Andrew, Antti Puurula, and Blake Burgess. "Improvements to BM25 and language models examined." Proceedings of the 2014 Australasian Document Computing Symposium.. 2014.

[20] Karthik, A., MazherIqbal, J.L. Efficient Speech Enhancement Using Recurrent Convolution Encoder and Decoder. Wireless Pers Commun 119, 1959–1973 (2021). https://doi.org/10.1007/s11277-021-08313-6.

[21] Zhou, Jie, et al. "Graph neural networks: A review of methods and applications." arXiv preprint arXiv:1812.08434 . 2018.

[22] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y.. A comprehensive survey on graph neural networks. IEEE transactions on neural networks and learning systems, 32(1), 2020.pp4-24.

[23] Douzi, Samira, et al. "Hybrid E-mail spam detection model using artificial intelligence." Int. J. Mach. Learn. Comput 10.2 : 316-322. 2020.

[24] Saini, Divyanjali, and Monalisa Meena. "Hybrid Forecasting Scheme for Enhance Prediction Accuracy of Spambase Dataset." Proceedings of International Conference on Communication and Computational Technologies. Springer, Singapore.. 2021

[25] Kim, Ji-hye, and Ok-ran Jeong. "Knowledge Graph-based Korean New Words Detection Mechanism for Spam Filtering." Journal of Internet Computing and Services 21.1 : 79-85. 2020

[26] Tran, M., Elsisi, M., & Liu, M. (2021). Effective feature selection with fuzzy entropy and similarity classifier for chatter vibration diagnosis. Measurement, 184, 109962.

[27] D. N. V. S. L. S. Indira, Rajendra Kumar Ganiya, P. Ashok Babu, A. Jasmine Xavier, L. Kavisankar, S. Hemalatha, V. Senthilkumar, T. Kavitha, A. Rajaram, Karthik Annam, Alazar Yeshitla, "Improved Artificial Neural Network with State Order Dataset Estimation for Brain Cancer Cell Diagnosis", BioMed Research International, vol. 2022, Article ID 7799812, 10 pages, 2022. https://doi.org/10.1155/2022/7799812.

[28] Tran, M., Liu, M., & Elsisi, M. (2021). Effective multi-sensor data fusion for chatter detection in milling process. ISA transactions.

[29] Tran, M., Elsisi, M., Mahmoud, K., Liu, M., Lehtonen, M., & Darwish, M.M. (2021). Experimental Setup for Online Fault Diagnosis of Induction Machines via Promising IoT and Machine Learning: Towards Industry 4.0 Empowerment. IEEE Access, 9, 115429-115441.

[30] Elsisi, M., Tran, M., Mahmoud, K., Mansour, D.A., Lehtonen, M., & Darwish, M.M. (2021). Towards Secured Online Monitoring for Digitalized GIS Against Cyber-Attacks Based on IoT and Machine Learning. IEEE Access, 9, 78415-78427.

Cite this paper as:

Rahmath Nisha, S.; Muthurajkumar, S. (2023). Semantic Graph Based Convolutional Neural Network for Spam e-mail Classification in Cybercrime Applications, *International Journal of Computers Communications & Control*, 18(1), 4478, 2023.

https://doi.org/10.15837/ijccc.2023.1.4478