International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7191
***Article Received**: 23 April 2023 **Revised:** 11 June 2023 **Accepted:** 27 June 2023*

Deep Learning based Densenet Convolution Neural Network for Community Detection in Online Social Networks

M. Selvakumar[1*], Dr. A. Vijaya Kathiravan[2]
[1]*Research Scholar, Department of Computer Science,
Periyar University,
Salem – 636011, India
mselvaamca@gmail.com
[2]Assistant Professor and Head, Department of Computer Applications,
Sri Meenakshi Government Arts College for Women,
Madurai – 625002, India
vijayakathiravan@gmail.com

Abstract: Online Social Networks (OSNs) have become increasingly popular, with hundreds of millions of users in recent years. A community in a social network is a virtual group with shared interests and activities that they want to communicate. OSN and the growing number of users have also increased the need for communities. Community structure is an important topological property of OSN and plays an essential role in various dynamic processes, including the diffusion of information within the network. All networks have a community format, and one of the most continually addressed research issues is the finding of communities. However, traditional techniques didn't do a better community of discovering user interests. As a result, these methods cannot detect active communities. To tackle this issues, in this paper presents Densenet Convolution Neural Network (DnetCNN) approach for community detection. Initially, we gather dataset from Kaggle repository. Then preprocessing the dataset to remove inconsistent and missing values. In addition to User Behavior Impact Rate (UBIR) technique to identify the user URL access, key term and page access. After that, Web Crawling Prone Factor Rate (WCPFR) technique is used find the malicious activity random forest and decision method. Furthermore, Spider Web Cluster Community based Feature Selection (SWC2FS) algorithm is used to choose finest attributes in the dataset. Based on the attributes, to find the community group using Densenet Convolution Neural Network (DnetCNN) approach. Thus, the experimental result produce better performance than other methods.

Keywords: Online Social Networks (OSNs); community detection; user interest; feature selection; preprocessing.

I. Introduction

The term social media refers to computer-based technologies that allow people to share thoughts, ideas, and information via virtual networks and communities. Based on the Internet, social media provides a way for users to quickly exchange content such as personal information, documents, videos, and photos electronically. Users interact with social media through web-based software or apps, computers, tablets, or smartphones. Social media can take the form of many different types of technologically enabled activities. These activities include photo sharing, blogging, social gaming, social networking, video sharing, business networking, virtual worlds, reviews, and more. Governments and even politicians use social media to communicate with their voters and their constituents. Facebook is the largest social media platform in the world, it has the same audience as other social media like Twitter and Instagram, but has a distinct advantage over them.

Social media sites like Facebook, Instagram, and Twitter etc. Those social networks will deal with stock markets. Either it can buy new shares or you can sell the ones you already have. They are networks not only in the field of technology but also in our everyday social life, we deal with many networks. Communities are property of multiple networks in that a given network can have multiple communities, meaning that nodes within a community are densely connected. Nodes in many communities may overlap.

It is necessary to analyze different networks and discover the communities within them. A major application of social discovery technology in social media platforms is to identify people with common interests and leverage its data. Groups with similar characteristics are used to exclude groups for social detection purposes in machine learning. This technique can be used to detect manipulation groups in social networks or stock market.

The literature uses graph partition-based methods to partition graphs into components such that components have specific connections. The kernighan-Line14 graph partition algorithm is one of the earliest graph partition techniques. Minimize the sum of the costs of all cut edges by dividing the vertices of the graph into subsets of a certain

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7191
***Article Received**: 23 April 2023 **Revised:** 11 June 2023 **Accepted:** 27 June 2023*

size and computing the costs on the edges. However, the main drawback of this method is that the lot of groups must be pre-defined.

The proposed method computes the main interest in community detection of groups or cohesive subgroups. Clustering serves as the foundation of many social detection algorithms. An edge median-based segmentation algorithm can detect data for graphs with undirected and unweighted edges. The algorithm focuses on the edges between communities and between communities by removing these edges from the original graph.

An adaptive heuristic search algorithm is a genetic algorithm (GA) designed to find the optimal solution for a given situation. Chromosomes that start with a set of solutions are called Ru Genetic Algorithm Chromosomes. Then exercise calculates the activity for the chromosomes. If a solution with a higher fitness than the mutation operator is obtained in the current solution set, the other solution with a more random intersection is terminated to obtain a new solution set. It would be optimal to choose an objective function that captures the intuition that internal connectivity is better than external connectivity as a social detection optimization problem for communities.

Our contributions can be summarized as follows:

- Initially, we gather dataset from Kaggle repository. Then preprocessing the dataset to remove inconsistent and missing values.
- In addition to User Behavior Impact Rate (UBIR) technique to identify the user URL access, key term and page access.
- After that, Web Crawling Prone Factor Rate (WCPFR) technique is used find the malicious activity random forest and decision method.
- Furthermore, Spider Web Cluster Community based Feature Selection (SWC2FS) algorithm is used to choose finest attributes in the dataset.
- Based on the attributes, to find the community group using Densenet Convolution Neural Network (DnetCNN) approach.
- Thus the implementation result provides higher community detection accuracy performance, sensitivity, and specificity. It produces less false classification with low time complexity results.

II. Literature Review

E. D. Raj et al, (2021), the author proposes a method by which communities within OSN can be detected. A new social detection algorithm called Grain-Based Social Detection (GBCD) is developed based on the Rough Set Granular Social Network (RGSN) model. The granular social factor and the objective social factor use two dimensions. Four real-world datasets were used to evaluate the datasets and were generated by computer.

R. Aktunc et al, (2022), Compared with previous solutions in the literature, it is easier to analyze the event detection performance on real-world and standard datasets using the authors' proposed method. The author proposes a new method, CN-NEW, based on overall social structure. Experimental results show that the proposed method achieves higher event detection accuracy than the baseline method. This method is used to analyze large amounts of communication data and measure it.

K. Chakraborty et al, (2020), the author proposes a method to provide a multifaceted view of the evolution of sentiment analysis by exploding the vast amount of data on the web. Describes the process of extracting data from social media to detect similarity of users in social networks based on similar choices. Classification techniques are also used in this method to analyse user data. The data presented in different formats analysed as part of the survey.

W. Luo et al, (2022), the author states that the social structure of multiple networks can be detected by two methods, Fast Modular Algorithm (Fast Modular) and Label Propagation Algorithm (LPA). Ghost Modular and CLPA are used to discover community structure in multiple networks to provide. CofastModular and CoLPA methods help to identify social structures effectively.

G. Li et al, (2019), The author proposes a new type of procedural object-oriented behavioural control models based on mining techniques to develop insights from occurrence data in silent media. Process models are developed based on real-life data to describe user behaviour patterns. This method is used to detect deviations and disturbances on the stock exchange website in question and answer process.

A. Sakor et al, The author proposes a method for retrieving contextually relevant posts by focusing on a specific topic. A knowledge-based framework, PINYON provides a method for efficiently retrieving relevant posts. PINYON is used to implement a two-fold pipeline. It encodes the corpus of records and the input post into a graph. Posts to existing knowledge maps are annotated with organizations and linked based on similarity.

X. Li et al, (2019), the author develops a model based on influence correlation index of multi-layer path length metrics for multi-layer networks. The method works as a local social detection model combining direct influence

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7191
***Article Received**: 23 April 2023 **Revised:** 11 June 2023 **Accepted:** 27 June 2023*

and indirect relationships based on a multi-layered network influence relationship (IMLC).

E. Kafeza et al, (2020), The author introduces a Twitter Personality-Based Communicative Community Extraction (T-PCCE) system to identify the most connected communities in the Twitter network graph by considering user personalities. This technique works by modelling the performance Twitter graph considering the individuality factor of communication intensity of the extracted community. This method emphasizes multiple indicators to quantify the strength of communication within each community.

B. A. H. Murshed et al, (2022), the author proposes a DEA-RNN method to detect CB in Twitter social media networks using a hybrid deep learning model. A new method DEA-RNN model is developed that combines an optimized dolphin echolocation algorithm (DEA) with an Elman-type recurrent neural network (RNN). It is used to reduce the training time and fine tune the parameters of Elman RNN.

C. He et al, (2022), the authors propose the NMFGAAE method based on the Non-negative Matrix Factorization (NMF) based Graph Attention Auto Encoder (GAAE) method. The method was developed to improve the performance of NMF-based neural network community detection methods with deep clustering. An attention mechanism is introduced to enable GAAE to drive through NMF-based social detection while NMF focuses on revealing these representations.

Z. Xu et al, (2020), the author proposes a 5W model of what, where, when, who and why to identify and understand a real-time urban emergency event. Targeting of user crowd-sourced social media. Real-time event detection in social media is used to extract spatial and temporal information. GIS-based annotation of urban emergency event is detected and known.

S. Zhang et al, (2022), the author proposes a new label propagation algorithm (LPA) (NOHLPA) method to combine overlap and historical label similarity in a multi-layer neighbourhood. Considers both label selection rules and node update order. We cited label entropy and predicted the most suitable label selection rules as the basis for the node update sequence predicting multi-layer neighbourhood and historical label similarity.

R. Ren et al, (2022), the author proposes a new method to detect ranked and intersecting communities based on the Cumulative Opinion Distance (COD). Standard fitness is known to be different from classical algorithms that rely on measurements. Deploys asynchronous connectivity across the network. The detection limit of algorithms increases efficiency in random networks with results related to the speed of consensus convergence by estimating the eigenvectors of adjacency matrices.

M. Qiao et al, (2019), the author introduces network analysis of urban space from social media and the micro-subjective division of human interaction. Cut points for exponential distributions are determined by fitting probability distribution functions to normal function modes. Given the importance of node gravity in designing weight-based urban spaces, gravity models are developed by incorporating hierarchical traditional spatial networks.

F. K. Sufi et al, (2022), the authors propose anti-vax and pro-vax methods to ethically identify and deal with surveilled artificial intelligence (AI)-based social groups. Named Institutional Recognition (NER) uses AI-based sentiment analysis to allow political scientists to assess the influence and power of social groups to policymakers. During the period of surveillance, the pro-Vax social movement is known to have an average of negative sentiment in COVID-19-related posts globally.

A. Rodriguez et al, (2022), A method was introduced by the author to provide data analysis processing in natural language for media providers to understand the spread of hate in social media. Page Reach's Recent Posts analyses comments using sentiment and sentiment analysis algorithms. A clustering algorithm evaluates posts suspected of containing inhumane words.

D. Stiawan et al, (2021), the author proposes a new dataset IDS that can be used to identify the best-fitting selected features as critical features. A method for developing an optimally integrated IDS to achieve this goal is developedIt is used to select six parameters namely Information Gain (IG), Gain Ratio (GR), Symmetric Uncertainty (SU), Relief-F (R-F), One-R (OR) and Chi-Square (CS).

H. Rong et al, (2022), the author proposes K-means algorithm to be used for segmentation of beautiful images. The uncertainty in the number of clusters and the inherent sensitivity to randomness in the initialization phase of the initial cluster centers of the K-means algorithm can be quantified using the new method.

N. Hussain et al, (2020), Spam Review Detection Methods The author proposes a new method for two different jobs. Spam Review Detection (SRD-BM) calculates a review spam score using two methods of thirteen different spammer behavioural features. Spammers are used to recognise spam comments. Comment spam

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7191
***Article Received**: 23 April 2023 **Revised:** 11 June 2023 **Accepted:** 27 June 2023*

detection is performed using a linguistic method (SRD-LM) for feature selection on the comment content.

S. Salloum et al, (2022), Research field Machine learning algorithm used for phishing email detection the author proposes a phishing email detection method that uses NLP in phishing emails. Text features, datasets, and sources used in phishing emails Machine learning algorithms improve the criteria for evaluating phishing emails. Feature extraction is a major research area in phishing detection research, followed by phishing email detection.

D. Zhang et al, (2020), the author proposes a search process under a regional search paradigm under a deep reinforcement learning framework. This method is used to learn the agent by creating a search process to learn the search agent step by step. Pseudo-holistic object regions and corresponding local discriminants are developed to extract object regions and learn such search agents under weak supervision.

C. Jiménez-Mesa et al, (2020), The authors propose to use data from the International Challenge for Automatic Prediction of MCI from MRI Data for multiclass classification problems. This method solves the problem of outlier detection using a novel multi-class classification approach. The pairwise t-test feature-selection partial least-squares method projects the extraction of selected features onto the multiclass subspace. Error correction works on the output code classification.

W. Ai et al, (2022), the author proposes a two-channel method to create a Chinese enterprise automatic summary. The Enterprise Component Channel typically releases candidate summaries in a two-channel manner. The method selects features in single-character channel irregular summaries for irregular candidate summaries to improve the processing effect.

S. Nasim et al, (2022), the main goal of the authors is to use advanced machine-learning techniques to predict polycystic ovary syndrome. This dataset will be used to develop a research model based on the clinical and physical parameters of the women. A new feature selection method based on the Augmented Chi-Square (CS-PCOS) mechanism has been developed. The performance of the 10 hyper Para metalized models is improved compared to the machine learning model.

III. Implementation of the proposed methodology

This module explain detail description of community detection in Online Social Networks (OSNs). Figure 2 illustrates the overview of community detection using Densenet Convolution Neural Network (DnetCNN) approach. Initially, we gather dataset from Kaggle repository. Then preprocessing the dataset to remove inconsistent and missing values. In addition to User Behavior Impact Rate (UBIR) technique to identify the user URL access, key term and page access.

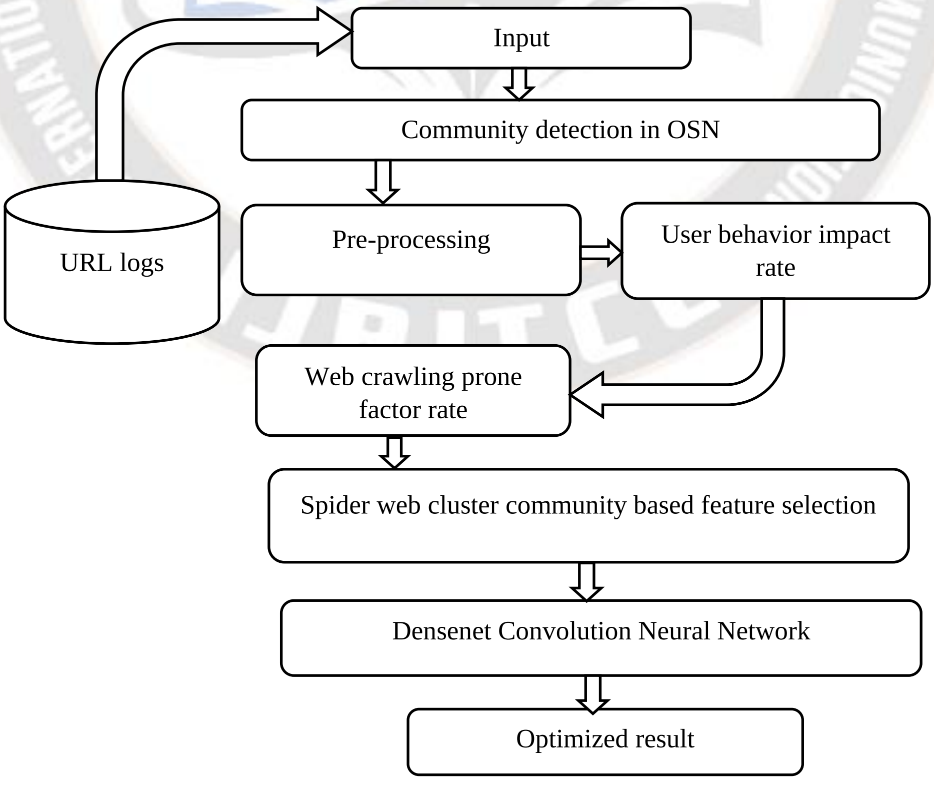

Figure 1: Block diagram for Community detection using DnetCNN

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7191
***Article Received**: 23 April 2023 **Revised:** 11 June 2023 **Accepted:** 27 June 2023*

After that, Web Crawling Prone Factor Rate (WCPFR) technique is used find the malicious activity based on random forest and decision method. Furthermore, Spider Web Cluster Community based Feature Selection (SWC2FS) algorithm is used to choose finest attributes in the dataset. Based on the attributes, to find the community group using Densenet Convolution Neural Network (DnetCNN) approach. Thus the implementation result provides higher community detection accuracy performance, sensitivity, and specificity. It produces less false classification with low time complexity results.

3.1 Preprocessing

In this stage, preprocess was carried out from the input logs to prepare the dataset for feature analysis. First the records are verified to presence of all the data values. Verification and indexing records was carried to fill, remove, cleaning the noise data. All the verification are done, each logs carried out with defined marginal values. To check the presence of all the logs values related to the attribute nature. All the records in the dataset are prepared to reduce the dimension of the dataset in intrusion detection.

Mean $(M^n) = \sum_{i=1}^{n} \frac{O_b}{T_R}$ (1)

Equation 1 depicts mean values of given dataset. O_b Denotes observed values and T_R presents total records in the dataset. Mean value is used to analysis the average values of dataset.

Standard deviation $(s^d) = \sqrt{\sum_{i=1}^{n} \frac{(O_b - M^n)^2}{T_R}}$ (2)

Equation 2 depicts to identify the standard deviation based on mean values and observed values.

Input: URL dataset In_d

Output: Preprocessed datasetN_D

Begin procedure

Import URL dataset In_d

$\forall\, In_d\ do\ n\ then$:

Check noisy and null values (N_L) in the dataset

$$N_L = \sum \frac{I_d - M^n}{s^d}$$

Eliminate the records (r_i)

Rearrange the normalized dataset N_D

$$N_D = \sum \frac{N_L - minimum(f_v)}{Maximum(f_v) - minimum(f_v)}$$

Update records in the dataset.

End for

End for

Return preprocessed dataset← N_D

Stop procedure

The proposed algorithm is efficiently remove null and noisy records in the air quality dataset. Let us assume f_v feature values, and n is the number of times to check null and noisy values in the dataset.

3.2 User Behavior Impact Rate

In this stage, we uses User Behavior Impact Rate (UBIR) technique to identify the user interest, key term and web page access of the processed dataset. In this step, the proposed algorithm calculates the degree of interaction between each pair of users connected to the social network. This technique found that the importance of interactions varied based on user behaviour. First, we calculate the likes and comments based on the probabilistic process among original and different users in social networks.

$T_{like} = [Tags] * p_1$ (3)

$T_{comment} = [Tags] * p_2$ (4)

Expression 3 and 4 is used to estimate user's tag like $T_{like}$and comments $T_{comment}$on web social networks. Where, p_1 and p_2 is the probability coefficient.

$K_{term}(P_{cn}, key_w, P_b) = \frac{t_f(key_w, P_{ci})}{\sum_{key_w \in P_b} t_f(key_w, P_{ci})} X log\left(\frac{|P_b|}{|\{key_w \in P_b\}|}\right)$ (5)

Expression 5 is used to evaluate the user's specific key term K_{term}in the SN. Let us assume P_{ci} is the preprocessing set whole collection of nth user's blogs. key_w Denotes keyword, P_b is the all blogs in the dataset. $t_f(key_w, P_{ci})$ Represents amount of times key_w appeared in the datasetP_{cn}.

$U_{Interst} = \alpha U_n + (1-\alpha)\sum_{n=1} K_{term}\, Com_n$ (6)

Expression 6 is used to estimate the predict user interest$U_{Interst}$. Let us assume n is the number of users U. Com_n represents the community in SN, α denotes adjustable value of own interest.

$U_{activity} = T_{like} * w1 + T_{comment} * w2 + U_{Interst} * w1$ (7)

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7191
Article Received: 23 April 2023 Revised: 11 June 2023 Accepted: 27 June 2023

Expression 7 defines user activity in social networks $U_{activity}$ based on tag like T_{like}, comments $T_{comment}$ and user interest $U_{Interst}$. Where, $w1$, $w2$ and $w3$ is the weight factor to identify the user behavior in the network. This section is proficiently analysis the user's activity in the social networks.

3.3 Web Crawling Prone Factor Rate

In this section we apply the web crawling prone factor rate to find malicious activity in the social network. This method worked with random forest and decision methods to detect web user malicious activity. Random Forest algorithm makes multiple decision trees integrated for more accurate identification of activities. The forest selects the user malicious activity detection with the highest number of votes. When using a random forest for regression, the forest chooses the average of all tree outputs.

$$U_{a,b} = C_{a,b}^{1} X w^{1} + C_{a,b}^{2} X w^{2} + \cdots C_{a,b}^{ty} X w^{ty} + \beta \quad (8)$$

The above expression is estimate the two users a and b relative interaction in SN. Let us assume, $\{C^1, C^2, \dots C^{ty}\}$ denotes user interaction, ty presents interaction types between users a and b in OSN and Weights $\{w^1, w^2, \dots w^{ty}\}$ respectively. β denotes the positive important of interaction.

$$M_{a,b}^{u} = \sum_{n=1}^{g} U_{a,b}^{M^u} \quad (9)$$

The above equation is evaluate the group activity $M_{a,b}^{u}$. Here this expression computes each pair of users connected by one or more common neighbors and their interaction with those common neighbors to calculate the group activity of teams of users. Assuming that, $M^u = \{m1, m2, \ldots mg\}$ denotes common users of a and b, g defines number of common neighbors.

$$URL_{a,b} = \frac{|URL_a \cap URL_b|}{|URL_a \cup URL_b|} \quad (10)$$

The above expression is used to estimate URL sharing in social networks between users a, b. assuming that, URL_a, URL_b are a set of URLs is shared by social network users.

$$SI_{a,b} = \frac{|N(U_a) \cap N(U_b)|}{\sqrt{|N(U_a)| * N(U_b)}} \quad (11)$$

The above expression is used to estimate social interaction similarity $SI_{a,b}$ among neighboring N users a and b.

3.4 Spider Web Cluster Community based Feature Selection

In this phase, we apply Spider Web Cluster Community based Feature Selection (SWCFS) method for select optimal features of community. The Social web community is a metaheuristic approach that emulates social spiders living together, finding food and exchanging necessary information. Commonly, the social spider population is split into two parties (females and males, 50-90% of the spider population), and these parties form a web for searching for prey. The SWCFS solution is represented by the location of each spider in the web and translates important features the zone of the prey and each spider to the other spiders. This important feature is represented by its vibrations as it moves from one place to another in the spider web.

$$F_t = \frac{w_{inter}^{C}}{w_{inter}^{C} + w_{exter}^{C}} \quad (12)$$

The above equation is used to evaluate the fitness function F_t of spider in the web. Assuming that, C is the user interaction, w_{inter}^{C} is the internal weights and w_{exter}^{C} is the external weights in the web.

$$Vib_{xy} = F_t e^{-D_{xy}^{2}} \quad (13)$$

The above expression identify the vibration Vib_{xy} of spider x and spider y. Here, D is the distance between two spiders. The distance defined by,

$$D_{xy} = \left\| z_x - z_y \right\| \quad (14)$$

Here z is the spider's position in the web.

Algorithm steps

Input: Crawling Prone Factor Rate $SI_{a,b}$, male spider, female spider

Output: Finest features (F)

Begin procedure

Import spider population in the web (L_n)
$//L_n$ spider population

Generate female spiders F_{spider}

$$F_{spider} = [(0.9) - r_n X L_n]$$

Generate male spiders M_{spider}

$$M_{spider} = L_n - F_{spider}$$

Estimate the F_{spider} and M_{spider} upper (h_{bou}) and lower bounds (l_{bou})

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7191
Article Received: 23 April 2023 Revised: 11 June 2023 Accepted: 27 June 2023

$F_{spider}(x,y) = F_{spideri(l_{bou})} + r_n\left(F_{spideri(h_{bou})} - F_{spideri(l_{bou})}\right)$

$M_{spider}(x,y) = M_{spideri(l_{bou})} + r_n\left(M_{spideri(h_{bou})} - M_{spideri(l_{bou})}\right)$

For each male spider M_{spider} do n then

For each female spider F_{spider} do n then

Estimate the features dimension f_d

$$f_d = \left[\frac{Dim(SI_{a,b})*i}{maximum\ (i)}\right]$$

Calculate feature dimension upper limit L_{upper} and lower limit L_{lower}

$$R_{LU} = \sum_{M_{spider}(x,y)}^{F_{spider}(x,y)} \frac{\left(f_d{}^{L_{upper}} - f_d{}^{L_{lower}}\right)}{2}$$

Spider x move to y^{th} spider based on their vibration S^{move}

$$S^{move} = \begin{cases} S_i^{move} + \alpha Vib_{xi}(I_n - S_i^{move}) + \beta Vib_{xi}(I_n - S_i^{move}) + \delta(r_n - 0.5) \\ S_i^{move} + \alpha Vib_{yi}(I_n - S_i^{move}) - \beta Vib_{yi}(I_n - S_i^{move}) + \delta(r_n - 0.5) \end{cases}$$

End for each

End for each

Return F← S^{move}

Stop procedure

The above algorithm steps is analysis the optimal features of social community based on spiders behaviors in the web. Let us assume r_n is the random numbers among [0, 1], Dim is a dimension in the dataset, i denotes current iteration, $maximum\,(i)$ is present total iteration in the dataset, R_{LU} is a feature dimension identification based on lower limit and lower limit, α, β, δ is a positive integers between [0, 1] and I_n integer adjacent to the superior spider.

3.5 Densenet Convolutional Neural Network

This section applied Densenet Convolutional Neural Network (DnetCNN) algorithm for classify the community in the social networks. Hence, matrices are used for the proposed approach to decrease the computational burden. This Densenet CNN algorithm contains classical ingestion, Convolutional, pooling and fully connected layers. The Ingestion layer handles the generation of the adjacency matrix from the CNN approach. This matrix is fed into a convolutional layer that runs a series of filters to extract salient features. This layer outputs a set of matrices (feature maps) on which a maximum pooling operation is achieved to minimize the issue solution. The previous layer's output is fed as input to the fully connected layer, whose objective is to define the probability distribution of each node in K classes.

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix} \quad (15)$$

The above expression is input of ingestion layer.

Algorithm steps

Input: Finest features (F)

Output: Optimized result (O_R)

Begin procedure

Import Finest features (F)

For each F do n then

Calculate Ingestion layer process Ing_L

$$Ing_L = exp^{\sigma(1 - D_{jk} * w_{jk})}$$

//Two generic nodes j and k

Compute Convolutional Layer process Con_L

$$Con_{L\,(jk)} = ReLu\left(bias + \sum_{F=1}^{n} w_{jk} * Ing_L\right)$$

Compute maximum pooling layer process Poo_L

$$Poo_{L(jk)} = ReLu\left(\left[w_{jk} * Con_{L\,(jk)}\right] + bias\right)$$

Compute fully connected layer $Full_L$

$$Full_L = ReLu\left(\left[w_{jk} * Poo_{L(jk)}\right] + bias\right)$$

Evaluate the specific community classification

$$O_R = \frac{1}{2}\sum_n (Full_L - loss)^2$$

End for each

Stop procedure

This algorithm steps efficiently classifies the community detection O_R in the social networks. Assuming that j, k are the adjacent nodes, σ Attenuation factor, $ReLu$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7191
***Article Received**: 23 April 2023 **Revised:** 11 June 2023 **Accepted:** 27 June 2023*

is the activation function, $loss$ denotes loss function to detect community in the social networks.

IV. Result and discussion

This section describes the proposed implementation comparing with other techniques using confusion matrix. The Web crawler detection using subset pattern feature analysis based on Multi-Perceptron neural network uses Phishing Dataset collected from the Kaggle repository. The comparison method are such as Grain-Based Social Detection (GBCD), Label Propagation Algorithm (LPA) and Subset pattern feature analysis Fuzzy inference approach based on Multi-Perceptron Neural Network (SPFAFI-MPNN) techniques. The result performance metrics are evaluated through confusion matrix performed n python evaluation.

Table 1 Simulation parameter settings

Parameters used	**Values**
Dataset name	Ebbu2017 Phishing Dataset
Tool	Anaconda/Jupiter notebook
Language	Python
Test result	Confusion matrix
Number of features	20
Number of class	3(High medium low)

The proposed carry enforcement generates higher detection rates by classifying results according to the class order. A system configuration with 4GB of RAM with the i3 Intel processor. Simulation parameters settings are illustrated in table 1.

This paper defines following parameters are classification accuracy, sensitivity, specificity, time complexity and false classification performance.

$$\text{Classification accuracy} = \frac{T_n+T_p}{(T_p+F_p+F_n+T_n)}$$

T_n is true negative, T_p is true positive, F_p is a false positive and T_n is a true negative.

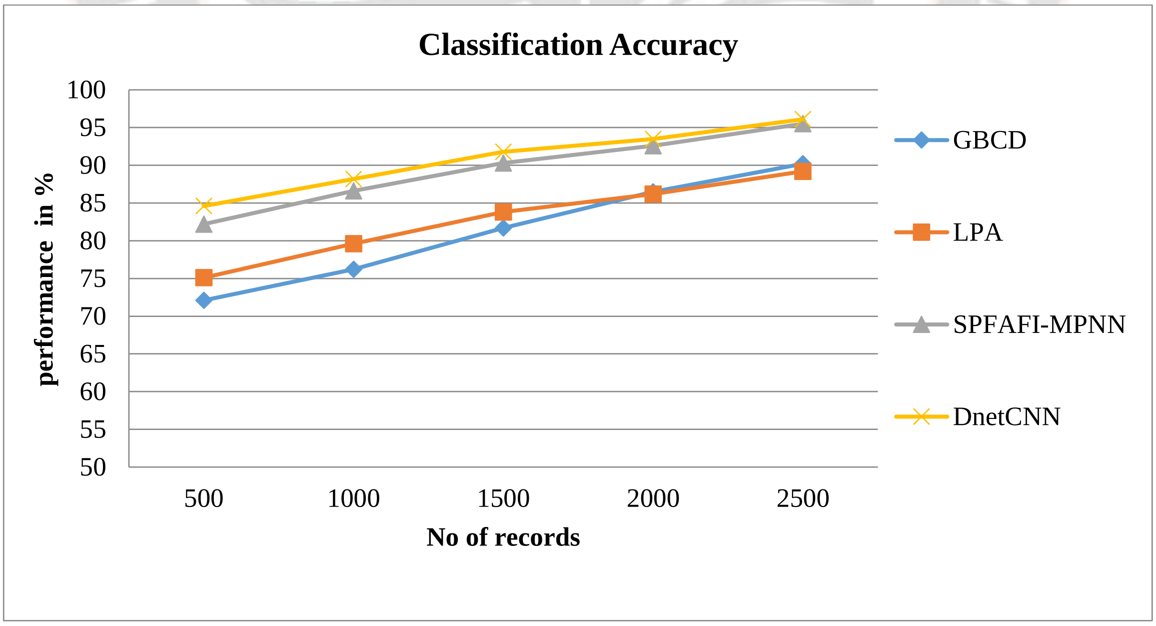

Figure 2 Analysis of classification accuracy performance

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7191
***Article Received**: 23 April 2023 **Revised:** 11 June 2023 **Accepted:** 27 June 2023*

Classification defines the sensitivity and uniqueness of frequent measurements predicted by fit/recall generated by positive values and shows absolute results based on type class. Figure 2 shows the accuracy of the classification.

Table 2 Impact of classification accuracy

Performance in %				
No of records/ Methods	GBCD	LPA	SPFAFI-MPNN	DnetCNN
500	72.1	75.1	82.2	84.6
1000	76.2	79.6	86.6	88.2
1500	81.7	83.8	90.3	91.8
2000	86.5	86.2	92.6	93.5
2500	90.2	89.2	95.5	96.1

Table 2 shows the classification accuracy compared to the previous approaches. The proposed method has the best performance in detecting malicious websites. Additional valid positive correlations are obtained with the actual value of the negative range.

$$\text{Sensitivity/recall } = \frac{T_p}{T_p+F_n}$$

The sensitivity estimation is done on phishing datasets. The online malicious web dataset, for the SPFAFI-MPNN produces 96.1% sensitivity, and the LPA achieves 87.1 % sensitivity, yet the GBCD classifier achieves 86.5 % sensitivity. The proposed SPFAFI-MPNN system has a higher impact on sensitivity.

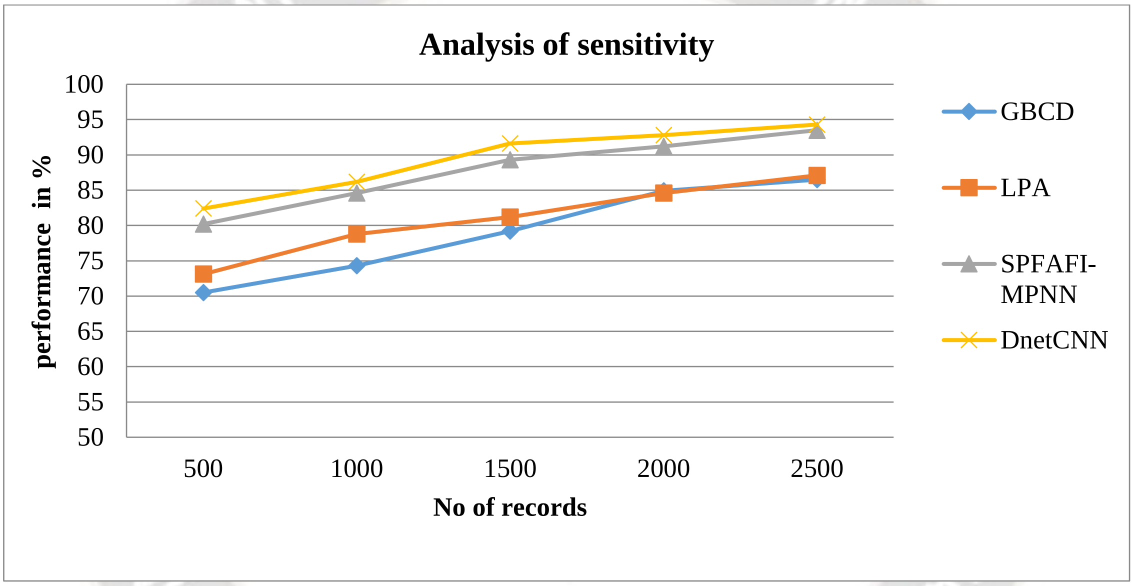

Figure 3 Analysis of sensitivity performance

Figure 3 defines analysis of sensitivity performance using phishing dataset collected from kaggle. The proposed SPFAFI-MPNN system accomplishes higher performance than the existing methods.

Table 3 Analysis of sensitivity performance

Sensitivity performance in %				
No of records/ Methods	GBCD	LPA	SPFAFI-MPNN	DnetCNN
500	70.5	73.1	80.2	82.4
1000	74.3	78.8	84.6	86.2
1500	79.2	81.2	89.3	91.6
2000	84.9	84.6	91.2	92.8
2500	86.5	87.1	93.5	94.3

Table 3 reviews the analysis of sensitivity performance with different records and previous techniques. The specificity is calculated as follows.

$$\text{Specificity} = \frac{T_n}{T_n+F_p} \quad (17)$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7191
***Article Received**: 23 April 2023 **Revised:** 11 June 2023 **Accepted:** 27 June 2023*

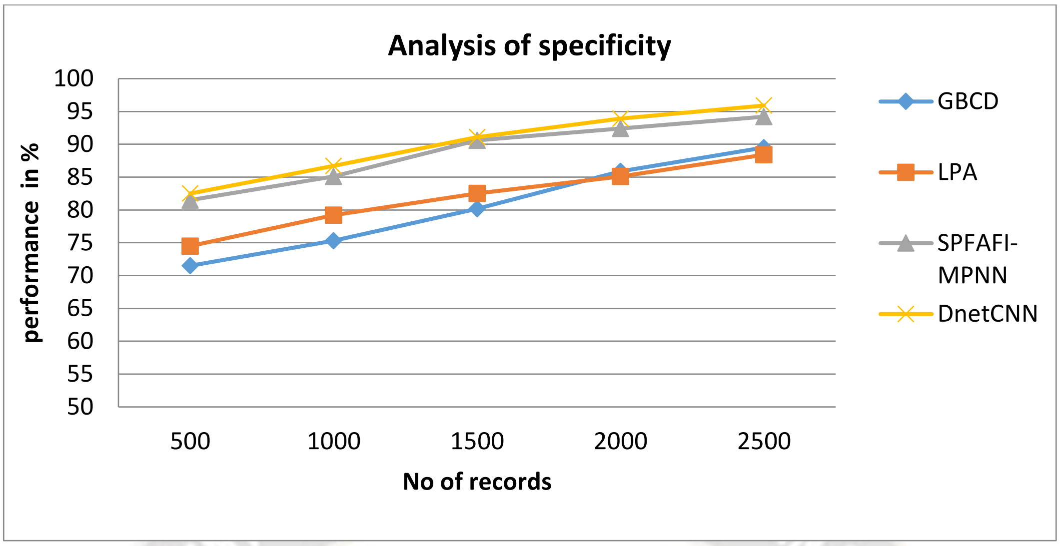

Figure 4 Analysis of specificity performance

Figure 4 shows variations of the so-called integrity generated by different approaches, where the SPFAFI-MPNN method produces higher performance than the additional methods.

Table 4 Impact of specificity performance

Specificity performance in %				
No of records/ Methods	GBCD	LPA	SPFAFI-MPNN	DnetCNN
500	71.5	74.5	81.5	82.5
1000	75.3	79.2	85.1	86.7
1500	80.2	82.5	90.6	91.1
2000	85.9	85.1	92.4	93.9
2500	89.5	88.4	94.2	95.9

The criterion of accuracy, depending on the percentage of valid values in the database, represents the coherent representation that avoids false positives and negatives. Table 4 shows the different methods of analyzed accuracy to indicate the ratio. The proposed approach emphasizes higher efficiency compared to other methods.

The expression of this expression is identified as follows:

False Extraction Ratio (Fer) = $\sum_{k=0}^{k=n} \times \frac{TotalDataset\ FailedtoClassify\ (Fer)}{TotalnoofData(Fr)}$

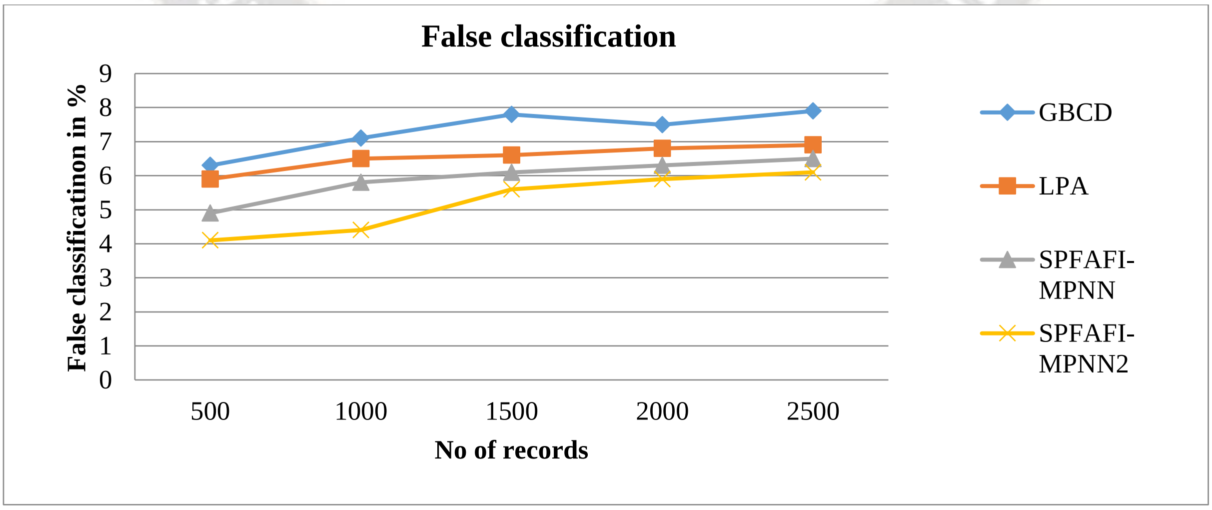

Figure 5 Effect of false classification

Figure 5 shows the variance of false ratios generated by different approaches, and the predicted SPFAFI-MPNN method has fewer erroneous classifications than the rest of the other approaches.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7191
***Article Received**: 23 April 2023 **Revised:** 11 June 2023 **Accepted:** 27 June 2023*

Table 5 Analysis of false classification

False classification performance in %				
No of records/ Methods	GBCD	LPA	SPFAFI-MPNN	SPFAFI-MPNN
500	6.3	5.9	4.9	4.1
1000	7.1	6.5	5.8	4.4
1500	7.8	6.6	6.1	5.6
2000	7.5	6.8	6.3	5.9
2500	7.9	6.9	6.5	6.1

Table 5 above explains that the misdiagnosis failed classification is classified as a misclassified class, regardless of whether the data were identified as a trained class or an attack. The misclassification detection class has the lowest predictive value for the proposed SPFAFI-MPNN system compared to other systems.

Time complexity (Tc)

$$=\sum_{k=0}^{k=n} \times \frac{\text{Total Features Handeled to Process in Dataset}}{\text{Time Taken(Ts)}}$$

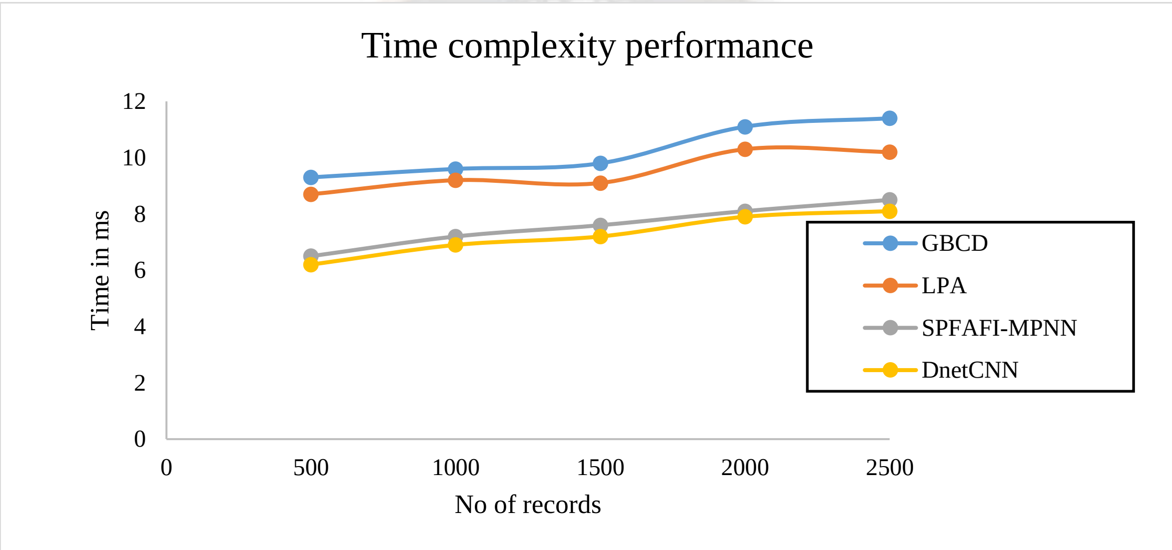

Figure 6 Impact of time complexity

Detection accuracy is designed based on the time it takes to process. Compare detection accuracy in different ways. O (n) will take time for better detection by processing all records based on the type definition of the defined type. The proposed system SPFAFI-MPNN produces 9.1 (ms) higher performance than all previous systems, as shown in Figure 6.

Table 6 Analysis of time complexity

Time complexity in ms				
No. of records/ Methods	GBCD	LPA	SPFAFI-MPNN	DnetCNN
500	9.3	8.7	6.5	6.2
1000	9.6	9.2	7.2	6.9
1500	9.8	9.1	7.6	7.2
2000	11.1	10.3	8.1	7.9
2500	11.4	10.2	8.5	8.1

Table 6 shows an estimate of the time problem generated by multiple approaches, and the predicted method caused the least time problem. The time lag is the total time it takes to load a database and execute feature selection and classification within a specified time. . The temple complex is calculated every millisecond. The asymptotic code O (n) on the front page is the lower limit when executing instructions. Using the time calculation and the worst form, the mean upper limit g (n) and the mean middle limit finite time f (n) calculates the difference in average time, which can measure the Length of time it can take.

V. Conclusion

This paper carried out Dense-net Convolution Neural Network (DnetCNN) approach for community detection in social networks. So in this proposed approach pre-processed the dataset was successfully removed irrelevant records. Then, we find the user interest based on key term and page access using User Behavior Impact Rate (UBIR) technique. Based on user interests our proposed method identified malicious activity in the gathered dataset using Web crawling prone factor rate. Afterwards, finest features selecting using Spider web cluster community based feature selection. Lately we detect community in the social networks using Densenet Convolutional Neural Network (DnetCNN) algorithm. Therefore, the proposed simulation result are classification accuracy rate is 96.1%, sensitivity is 94.3%, specificity is 95.9%, false classification rate is 6.1% and community detection time complexity result is 8.1%. Hence the proposed method performs better for community detection using python language than other methods.

Reference

[1] E. D. Raj, G. Manogaran, G. Srivastava and Y. Wu, "Information Granulation-Based Community Detection for Social Networks," in IEEE Transactions on Computational Social Systems, vol. 8, no. 1, pp. 122-133, Feb. 2021, doi: 10.1109/TCSS.2019.2963247.

[2] R. Aktunc, P. Karagoz and I. H. Toroslu, "Event Detection via Tracking the Change in Community Structure and Communication Trends," in IEEE Access, vol. 10, pp. 109712-109728, 2022, doi: 10.1109/ACCESS.2022.3214543.

[3] K. Chakraborty, S. Bhattacharyya and R. Bag, "A Survey of Sentiment Analysis from Social Media Data," in IEEE Transactions on Computational Social Systems, vol. 7, no. 2, pp. 450-464, April 2020, doi: 10.1109/TCSS.2019.2956957.

[4] W. Luo, B. Duan, L. Ni and Y. Liu, "Collaborative Detection of Community Structure in Multiple Private Networks," in IEEE Transactions on Computational Social Systems, vol. 9, no. 2, pp. 612-623, April 2022, doi: 10.1109/TCSS.2021.3089535.

[5] G. Li and R. M. de Carvalho, "Process Mining in Social Media: Applying Object-Centric Behavioral Constraint Models," in IEEE Access, vol. 7, pp. 84360-84373, 2019, doi: 10.1109/ACCESS.2019.2925105.

[6] A. Sakor, K. Singh and M. -E. Vidal, "Resorting to Context-Aware Background Knowledge for Unveiling Semantically Related Social Media Posts," in IEEE Access, vol. 10, pp. 115351-115371, 2022, doi: 10.1109/ACCESS.2022.3217492.

[7] X. Li, G. Xu, W. Lian, H. Xian, L. Jiao and Y. Huang, "Multi-Layer Network Local Community Detection Based on Influence Relation," in IEEE Access, vol. 7, pp. 89051-89062, 2019, doi: 10.1109/ACCESS.2019.2921571.

[8] E. Kafeza, A. Kanavos, C. Makris, G. Pispirigos and P. Vikatos, "T-PCCE: Twitter Personality based Communicative Communities Extraction System for Big Data," in IEEE Transactions on Knowledge and Data Engineering, vol. 32, no. 8, pp. 1625-1638, 1 Aug. 2020, doi: 10.1109/TKDE.2019.2906197.

[9] B. A. H. Murshed, J. Abawajy, S. Mallappa, M. A. N. Saif and H. D. E. Al-Ariki, "DEA-RNN: A Hybrid Deep Learning Approach for Cyberbullying Detection in Twitter Social Media Platform," in IEEE Access, vol. 10, pp. 25857-25871, 2022, doi: 10.1109/ACCESS.2022.3153675.

[10] C. He, Y. Zheng, X. Fei, H. Li, Z. Hu and Y. Tang, "Boosting Nonnegative Matrix Factorization Based Community Detection With Graph Attention Auto-Encoder," in IEEE Transactions on Big Data, vol. 8, no. 4, pp. 968-981, 1 Aug. 2022, doi: 10.1109/TBDATA.2021.3103213.

[11] Z. Xu et al., "Crowdsourcing Based Description of Urban Emergency Events Using Social Media Big Data," in IEEE Transactions on Cloud Computing, vol. 8, no. 2, pp. 387-397, 1 April-June 2020, doi: 10.1109/TCC.2016.2517638.

[12] S. Zhang, S. Yu, X. E, R. Huo and Z. Sui, "Label Propagation Algorithm Joint Multilayer Neighborhood Overlap and Historic Label Similarity for Community Detection," in IEEE Systems Journal, vol. 16, no. 2, pp. 2626-2634, June 2022, doi: 10.1109/JSYST.2021.3113826.

[13] R. Ren, J. Shao, Y. Cheng and X. Wang, "Detecting Hierarchical and Overlapping Network Communities Based on Opinion Dynamics," in IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 6, pp. 2696-2710, 1 June 2022, doi: 10.1109/TKDE.2020.3014329.

[14] M. Qiao, Y. Wang, S. Wu, A. Luo, S. Ruan and Y. Gu, "Fine-Grained Subjective Partitioning of Urban Space Using Human Interactions From Social Media Data," in IEEE Access, vol. 7, pp. 52085-52094, 2019, doi: 10.1109/ACCESS.2019.2911664.

[15] F. K. Sufi, I. Razzak and I. Khalil, "Tracking Anti-Vax Social Movement Using AI-Based Social Media Monitoring," in IEEE Transactions on Technology and Society, vol. 3, no. 4, pp. 290-299, Dec. 2022, doi: 10.1109/TTS.2022.3192757.

[16] A. Rodriguez, Y. -L. Chen and C. Argueta, "FADOHS: Framework for Detection and Integration of Unstructured Data of Hate Speech on Facebook Using Sentiment and Emotion Analysis," in IEEE Access, vol. 10, pp. 22400-22419, 2022, doi: 10.1109/ACCESS.2022.3151098.

[17] D. Stiawan et al., "An Approach for Optimizing Ensemble Intrusion Detection Systems," in IEEE Access, vol. 9, pp. 6930-6947, 2021, doi: 10.1109/ACCESS.2020.3046246.

[18] W. Han et al., "Improving Training Instance Quality in Aerial Image Object Detection With a Sampling-Balance-Based Multistage Network," in IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 12, pp. 10575-10589, Dec. 2021, doi: 10.1109/TGRS.2020.3038803.

[19] Joseph Miller, Peter Thomas, Maria Hernandez, Juan González,. Adaptive Decision Making using Reinforcement Learning in Decision Science. Kuwait Journal of Machine Learning, 2(3). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/204

[20] H. Rong, A. Ramirez-Serrano, L. Guan and Y. Gao, "Image Object Extraction Based on Semantic Detection and Improved K-Means Algorithm," in IEEE Access, vol. 8, pp. 171129-171139, 2020, doi: 10.1109/ACCESS.2020.3025193.

[21] N. Hussain, H. Turab Mirza, I. Hussain, F. Iqbal and I. Memon, "Spam Review Detection Using the Linguistic and Spammer Behavioral Methods," in IEEE Access, vol. 8, pp. 53801-53816, 2020, doi: 10.1109/ACCESS.2020.2979226.

[22] S. Salloum, T. Gaber, S. Vadera and K. Shaalan, "A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques," in IEEE Access, vol. 10, pp. 65703-65727, 2022, doi: 10.1109/ACCESS.2022.3183083.

[23] D. Zhang, J. Han, L. Zhao and T. Zhao, "From Discriminant to Complete: Reinforcement Searching-Agent Learning for Weakly Supervised Object Detection," in IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 12, pp. 5549-5560, Dec. 2020, doi: 10.1109/TNNLS.2020.2969483.

[24] C. Jiménez-Mesa et al., "Optimized One vs One Approach in Multiclass Classification for Early Alzheimer’s Disease and Mild Cognitive Impairment Diagnosis," in IEEE Access, vol. 8, pp. 96981-96993, 2020, doi: 10.1109/ACCESS.2020.2997736.

[25] W. Ai, H. Shao, J. Xu, T. Meng and K. Li, "A Two-Channel Chinese Enterprise Abbreviation Generation Method Based on an Enterprise Component and Single-Character Strategy," in IEEE Access, vol. 10, pp. 81123-81133, 2022, doi: 10.1109/ACCESS.2022.3192965.

[26] S. Nasim, M. S. Almutairi, K. Munir, A. Raza and F. Younas, "A Novel Approach for Polycystic Ovary Syndrome Prediction Using Machine Learning in Bioinformatics," in IEEE Access, vol. 10, pp. 97610-97624, 2022, doi: 10.1109/ACCESS.2022.3205587.