Detecting Image Spam on Social Media Platforms Using Deep Learning Techniques

Himani Jain[1, 2], Amit Dixit[3], Aditi Sharma[4,*]

[1]Quantum University, Roorkee, Uttarakhand Ph.D. Scholar, India
[2]Department of MCA, ABES Engineering College, Ghaziabad, Uttar Pradesh, India
[3]Dean Research Quantum University, Roorkee, Uttarakhand, India
[4]Department of Computer Sc. and Eng., Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India
Email: himanijain1987ap@gmail.com; dixitamit777@gmail.com; aditi.sharma@ieee.org

Abstract

Image spam involves the practice of concealing text within an image. Various machine-learning techniques are used to categories image spam, utilizing a wide range of features extracted from the images. Convolutional neural networks (CNNs) are commonly used for image classification and feature extraction tasks because of their outstanding performance. In this study, our focus is to analyses image spam using a CNN model that incorporates deep learning techniques. This model has been meticulously fine-tuned and optimized to deliver exceptional performance in both feature extraction and classification tasks. In addition, we performed comparative evaluations of our model on different image spam datasets that were specifically created to make the classification task more challenging. The results we obtained show a significant improvement in classification accuracy compared to other methods used on the same datasets.

Keywords: Deep Learning; Image Spam; Spam Detection; CNN; Social Sites

1. Introduction

Nowadays, individuals across all age groups are actively engaging on online social networking platforms. In addition, their usage has skyrocketed in recent years. In the past, individuals' primary motivation for signing up for OSN was to increase their social circle by meeting new people. On the other hand, now, more and more people are turning to OSN as a platform to boost their profile and increase their popularity. They produce reels, short clips, and videos, then upload them to OSN [1] to share with others. Their videos are viewed, and people choose whether to follow them based on their preferences. However, some people utilize the OSN platform improperly by sharing content with others that violates ethical standards. People that misuse personal data frequently obtain personal information from other people who are unaware of their actions.

Some individuals and organizations transfer data in order to gather personally identifying information about users for the intention of engaging in illegal activity [2]. Sending unwanted communications in a variety of formats, including trending hashtags, photographs, videos, Bollywood news, and explicit content is, in essence, how they want to turn a profit. People who participate in these kinds of illegal operations frequently use accounts that are known to be fake [3]. The identification and categorization of image spam are extremely fascinating academic fields that are also constantly developing. In spite of the many layers of protection and security, spammers continue to devise novel

DOI: https://doi.org/10.54216/JCIM.150106
Received: January 29, 2024 Revised: April 28, 2024 Accepted: July 24, 2024

approaches to reach people and target them with their messages. Users frequently receive a variety of types of communication, including email notifications from their banks, text messages from news outlets, and more. On the other hand, it might be challenging to differentiate between information that is authentic and that which is spam. Competitors are exceptionally bright in terms of IQ. Images are used as a bait to attract visitors who are not paying attention, and once they have their attention, they are sent to websites selling fake goods. On the other hand, users frequently ignore or neglect text messages, but they show a significant interest in the information that is communicated through images on OSN. In the past, people who transmitted unwanted messages known as spam were considered a nuisance. However, users were able to regain control of the situation by configuring filters in their mailboxes to identify and remove messages that were determined to be spam. However, the filter that detects photos does not successfully recognize spam images [4].

Different machine learning techniques, such as support vector machines, Naive Bayes trees, and Random Forests, are a few examples of the potential solutions that have been proposed to the problem of identifying spam. These spam detectors, on the other hand, place their primary emphasis on text- and graph-based features for classification, hence ignoring the detection of spam in images. A number of knowledgeable individuals carried out an in-depth investigation and then had a roundtable discussion about the benefits and drawbacks of using these classifiers. They also investigated several methods for improving these algorithms and observed that the circulation of photos has been gradually increasing in contrast to the circulation of text messages. In OSN, working with photos and obtaining their content can be a difficult and time-consuming process. Therefore, it is necessary for us to locate a solution that will enable us to recognize spam photographs and filter them out. In the beginning, the OCR approach is employed in order to identify image spam [5]. On the other hand, deep learning strategies have been shown to be helpful in overcoming OCR's limitations, which were discovered in the system's earlier iterations. The fundamental objective of this research is to construct a model that is able to determine with a high degree of precision whether photographs that are published on social media platforms are genuine or have been modified. As a result, we develop a very effective methodology for the detection of spam that greatly enhances the accuracy of spam detection in comparison to existing methods that are considered state-of-the-art. People have a tremendous interest in learning about the personal lives of Bollywood celebrities and actors in today's society due to the widespread curiosity surrounding this topic. When users on OSN click on specific links or bogus hashtags, they may be sent to unethical websites or apps where their personal information can be stolen. These websites and apps may also contain malware. Spammers frequently employ this method in their work. In the past, spammers were only able to deceive and manipulate information using textual data [6-7]. However, in recent years, people with malevolent intentions have been using fake photos, videos, URLs, and voices in order to deceive others. This practice has become more widespread. Consequently, it is necessary for us to identify the aforementioned kinds of activities that take place on social media; this is an important research subject that is gaining steam. For the purpose of this investigation, we carry out a number of experiments that make use of CNN models to categories image spam [8]. We describe the image spam detection with the help of flow chart in figure 1.

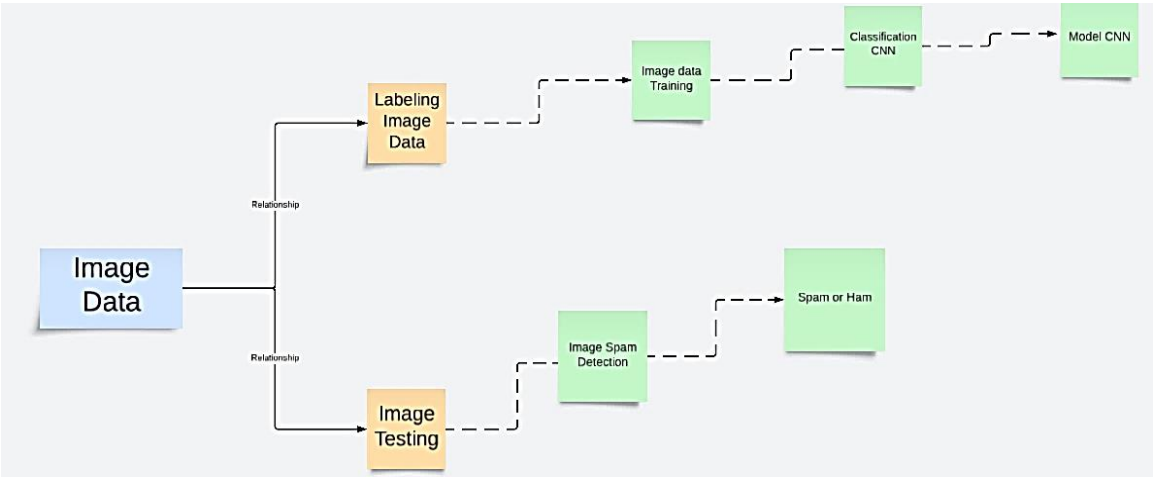

Figure 1. Flow chart for Image Spam Detection

DOI: https://doi.org/10.54216/JCIM.150106
Received: January 29, 2024 Revised: April 28, 2024 Accepted: July 24, 2024

The paper is structured into five distinct sections. In Section 2, we will deliver into the background of image spam. Section 3 discusses various related studies on identifying spammers. In Section 4, various image-processing techniques are discussed. Section 5 covers the methodology for collecting the dataset, extracting features, defining the spammer detection model, conducting experiments, and evaluating the results. Lastly, Section 6 covered the Result discussion, conclusion and future work.

2. Related Work

There are two primary approaches to detecting image spam: content-based and characteristics-based methods. In the content-based approach, the first step involves extracting text from images that could potentially be spam. Following that, the text undergoes analysis to ascertain whether it is spam or not. The same process can be used when dealing with spam that does not contain images. However, the second approach relies on the inherent characteristics of image files to identify spam content. In their study, Sharmin et al. [9] introduced three machine learning techniques: support vector machines (SVM), multilayer perceptron's (MLP), and convolutional neural networks (CNN), along with different combinations of image features. The feature extraction process displayed impressive efficiency, especially when compared to methods that rely on a diverse set of image features. The model proposed by Imam et.al.[10] focuses on identifying spam images on Twitter that include Arabic text by utilizing an end-to-end approach. The text detection and recognition were performed using the Accurate Scene Text Detector (EAST) and Convolutional Recurrent Neural Network (CRNN) models. The findings of this study demonstrate that developing a text recognition model from scratch necessitates a substantial amount of high-quality training data. When working with various languages during model training, it can be challenging to differentiate between characters that share similar shapes. A new technique was proposed [11] that combines content-based and graph-based features to identify spammers' profiles on the Twitter platform. The author solely relied on text-based features to detect spammers, have not explored the use of image spam for detecting spam on social media platforms. The work in [12] conducted a study titled "Image Spam Analysis and Detection Utilizing Two Methodologies." One approach involves utilizing PCA (Principal Component Analysis), while the other method focuses on extracting 21 features using SVM (Support Vector Machine). The subset of 13 features produced the most positive results. Furthermore, the authors have developed a one-of-a-kind dataset referred to as the "enhanced" dataset. It is important to note that this dataset cannot be identified using PCA and SVM techniques. An investigation was carried out on identifying image spam on Instagram through the utilization of Convolutional Neural Networks. The author [13] discussed four different architectures, including a 3-layer model, a 5-layer model, the AlexNet, and the VGG16. The team of experts used a web crawler to collect images from Instagram. The VGG16 model demonstrated superior accuracy compared to other architectures, achieving an impressive accuracy score of 0.842. It is worth noting that one particular architectural choice had a longer execution time compared to the others. In their work [14], the authors introduced a new method for detecting image spam called "Deep Learning and Data Augmentation." They used Convolutional Neural Networks (CNN) to address the issue of image spam in emails. For their study, a dataset was collected that consisted of 6,000 spam images and 2,313 non-spam images. The results demonstrated impressive performance metrics: an F1-Score of 88% when using CNN and an F1-Score of 82% when using a Support Vector Machine (SVM) for comparison.

3. Background

Spam refers to the unwanted sending of electronic mail and information to people, as well as the sharing of untrusted links, messages, photos, and videos. During this section, we covered picture spam and delved into relevant studies in the field.

3.1. Image Spam

Text spam usually involves receiving unsolicited messages in the form of text. On the other hand, the message is concealed within an image in Image Spam. In the past, individuals would exclusively send unsolicited text messages to users. However, in recent times, spammers have begun to transmit spam information using images Many users find it more effective to convey information through visual elements. So, hackers benefit from this. Image-based spam is frequently employed to deceive individuals into disclosing their personal information, disseminating harmful software, and advertising products. It can be quite challenging to differentiate between spam messages that contain

DOI: https://doi.org/10.54216/JCIM.150106
Received: January 29, 2024 Revised: April 28, 2024 Accepted: July 24, 2024

images and those that contain only text, due to the different techniques used to create these images. Several methods are utilized to create image spam, such as employing image-based obfuscation, incorporating multiple images, and introducing noise to the images.

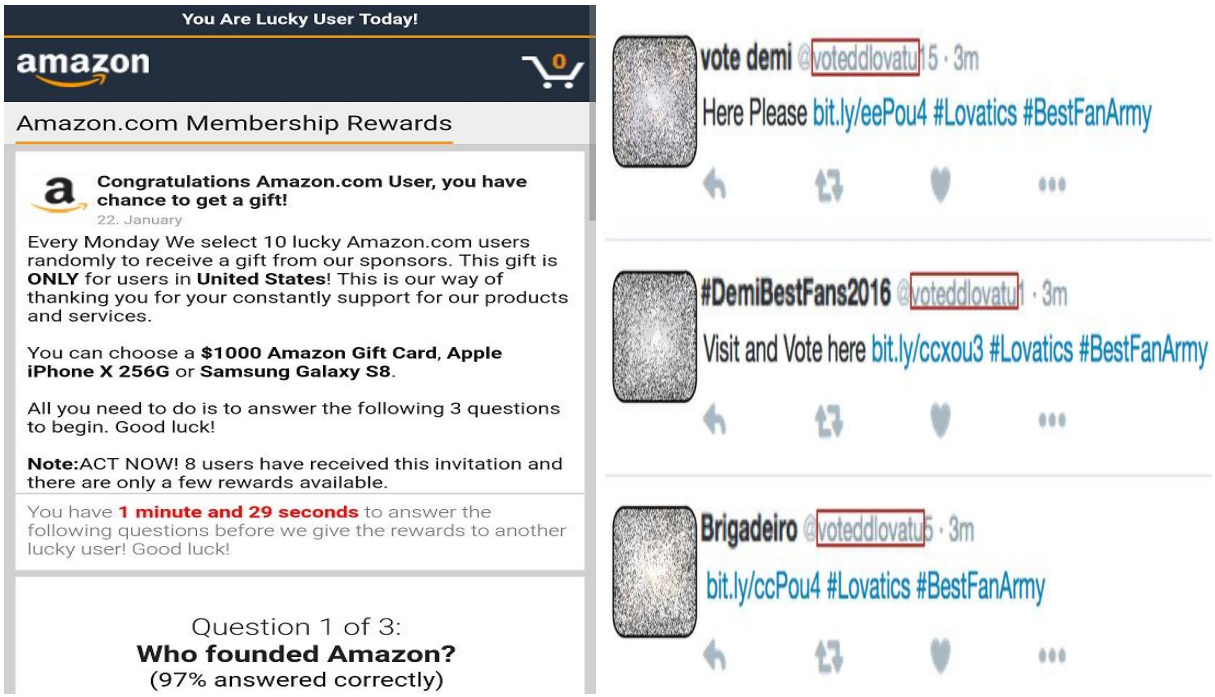

Figure 2. Image Spam: Amazon Gift Card, Spam link images.

Types of Image spam

Several different kinds of image spam

“Images that consist solely of text," "Images that have been divided into slices," and "Images that have been arranged in a random order" are the three distinct types of images that may be found online.

- Text-only images are images that only include text and are referred to by that name. Although they may look like emails composed of words, the messages in question are actually graphics. Optical Character Recognition, more commonly abbreviated to OCR, is a system that can pull text from images In addition to that, it uses the more common text-based filters.
- When arranged in the shape of a jigsaw puzzle, a collection of sliced photos can make it extremely difficult to identify which images belong to the spam category and which do not.
- Images that have been randomized contain pixels that have also been randomized in their placement. Image pixels are manipulated by spammers so that the resulting image is completely random. When dealing with randomized photos, it can be quite difficult to determine whether an image is spam. Figure 3 presents a number of different pictures.

DOI: https://doi.org/10.54216/JCIM.150106
Received: January 29, 2024 Revised: April 28, 2024 Accepted: July 24, 2024

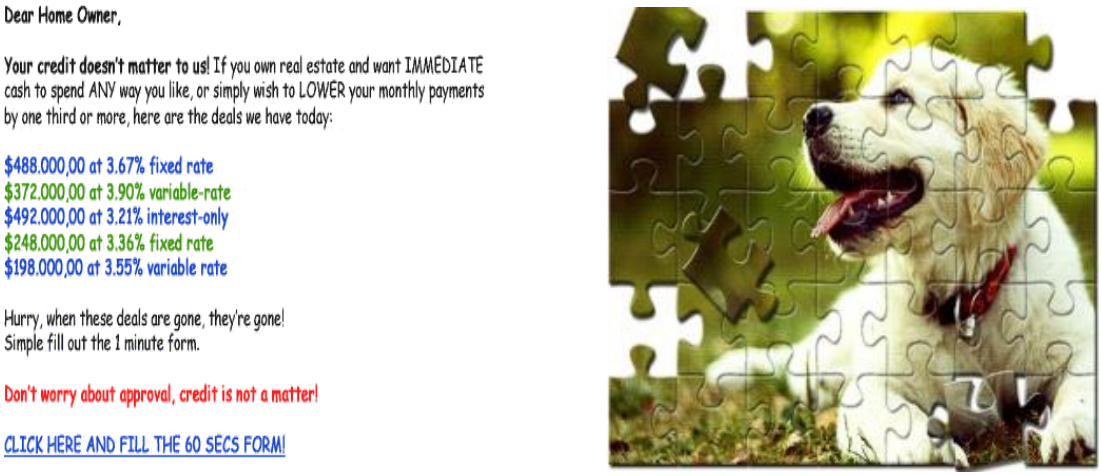

Figure 3. Text image, sliced image, Randomized image

4. Explainable ML

The explainability of machine learning (ML) approaches, particularly in distinguishing between images of spam and legitimate content, is a challenging task. In this paper, we tackle this challenge by comparing various classifiers using two distinct datasets. We employ a convolutional neural network (CNN), represented by the equation.

$$\text{CNN}(x) = softmax(FC(ReLU(Conv(x)))) \quad (1)$$

Where x represents the input image, Conv denotes the convolutional layers; ReLU is the rectified linear activation FC signifies the fully connected layers, and softmax provides the output probabilities. Therefore, we used various classifiers.

4.1. The Support Vector Machine

The Support Vector Machine (SVM) is a supervised machine learning technique used for classification tasks. Nevertheless, it can be classified as a task involving both classification and regression. It is frequently employed for identifying image spam [15]. SVM use a margin hyperplane to classify the dataset, aiming to optimize the distance between different classes. The SVM classifier, utilizing a kernel function, is employed to discern a spam image by analyzing the retrieved characteristics. The SVM algorithm is defined by many essential concepts: Stamp (2017)

- Separating hyperplane — during the training phase, the Support Vector Machine (SVM) aims to partition the labelled input data into two distinct groups. Ideally, the data is linearly separable, meaning that all data points belonging to one class are located on one side of a separating hyperplane, while all data points of the other class are located on the opposite side of the hyperplane.

- Optimize the margin – When creating an ideal hyperplane, only a portion of the training data is truly significant. These specific points are referred to as support vectors. An ideal hyperplane is a hyperplane that maximizes the separation or margin between the support vectors and the hyperplane.

- Engage in tasks involving dimensions beyond the usual three-dimensional space typically, a straight line in the input space cannot separate the training data. By mapping the input data to a feature space with a higher number of dimensions, it is possible to enhance the linear reparability.

- The kernel trick refers to the use of a kernel function that allows us to transform the input space into a higher dimensional feature space without incurring a substantial efficiency cost.

DOI: https://doi.org/10.54216/JCIM.150106
Received: January 29, 2024 Revised: April 28, 2024 Accepted: July 24, 2024

4.2. Feature Selection

In a linear Support Vector Machine (SVM), weights are assigned to each feature in the input space. The weight assigned to a feature indicates the level of importance that the SVM classifier assigns to that feature. Consequently, we may utilize these weights to priorities the features, so decreasing the complexity of the problem without compromising accuracy. Indeed, accuracy can be enhanced by reducing the number of features, as certain features may lack informative value and effectively function as noise. This study focuses on recursive feature elimination (RFE), which involves training an initial linear support vector machine (SVM) utilizing all the available features. Subsequently, we remove the feature with the lowest weight and proceed to train another linear Support Vector Machine (SVM) using the reduced set of features. We persist in diminishing the quantity of characteristics and retraining the Support Vector Machine (SVM) until the target amount of features has been achieved.

4.3. Evaluation Criteria

Accuracy is a statistical measure that quantifies the number of right classifications relative to the total number of classifications. It is calculated by dividing the sum of true positive and true negative classifications by the total number of samples. Accuracy is employed as a quantitative metric in our detection tests to assess the effectiveness of our proposed strategies. The ROC curve is created by graphing the true positive rate against the false positive rate, using different threshold values, based on the outcomes of a binary classification experiment. The range of the area under the receiver operating characteristic (ROC) curve (AUC) is between 0 and 1. A value of 1.0 for the AUC implies complete distinction, meaning that there is a specific threshold at which there are no incorrect positive or negative results. Conversely, an AUC value of 0.5 suggests that the binary classifier performs no better than random chance, like flipping a fair coin. The AUC provides the likelihood that a randomly chosen match case will have a better score than a randomly chosen non-match case (Bradley, 1997; Stamp, 2017). SVM experiments were conducted with a value of 3.4.

In order to conduct our SVM experiments, it is necessary to initially create feature vectors for training the models. The datasets contain photos of varying dimensions. Consequently, we initially adjust the size of all photographs to 32 × 32. Subsequently, we employ the Canny edge detection technique to transform an unprocessed image into a Canny image. In order to construct the feature matrix, we create byte data for every individual pixel in the Canny image. Every pixel is composed of three bytes, which encode the red, green, and blue (RGB) color information, ranging from 0 to 255. To facilitate computation, every integer is normalized to a range between 0 and 1. In addition, we generate a comparable feature vector using the unprocessed byte values, which are similarly standardized. It should be noted that every feature vector has a length of 1024.

We create distinct Support Vector Machine (SVM) models for each of the two datasets in our trials. For every dataset, we do a random shuffling and allocate 70% of the image samples for training purposes, while the remaining 30% are used for testing. Our SVM experiments involve testing both linear and RBF kernels, as well as other features.

Table 1 displays the SVM's accuracy when trained and tested on the ISH dataset. The SVM is trained and evaluated using raw images that have been shrunk to 32×32, and the results are compared to those obtained using images resized to 16×16. By utilizing the RBF kernel, we attain a superior accuracy of 0.9652, surpassing the maximum accuracy of 0.9156 achieved with the linear kernel. The raw picture feature outperforms the Canny image feature in both scenarios. The disparity in feature sizes for the RBF kernel is insignificant.

Table 1: SVM feature size and type comparison (ISH dataset)

Kernel	32 × 32 features Raw	32 × 32 features Canny	16 × 16 features Raw	16 × 16 features Canny
RBF	0.9748	0.9010	0.9752	0.9048
Linear	0.9156	0.8492	0.8838	0.7861

Figure 4 displays the Receiver Operating Characteristic (ROC) curves for the Support Vector Machine (SVM) binary classification outcomes. The curves are generated using the ISH dataset and employ both the RBF and linear kernels. The AUC values for the RBF kernel and the linear kernel are 0.97 and 0.73, respectively. The results demonstrate that employing an SVM with RBF kernel yields a high level of accuracy in distinguishing between ham and spam images, while maintaining a low proportion of false positives.

DOI: https://doi.org/10.54216/JCIM.150106
Received: January 29, 2024 Revised: April 28, 2024 Accepted: July 24, 2024

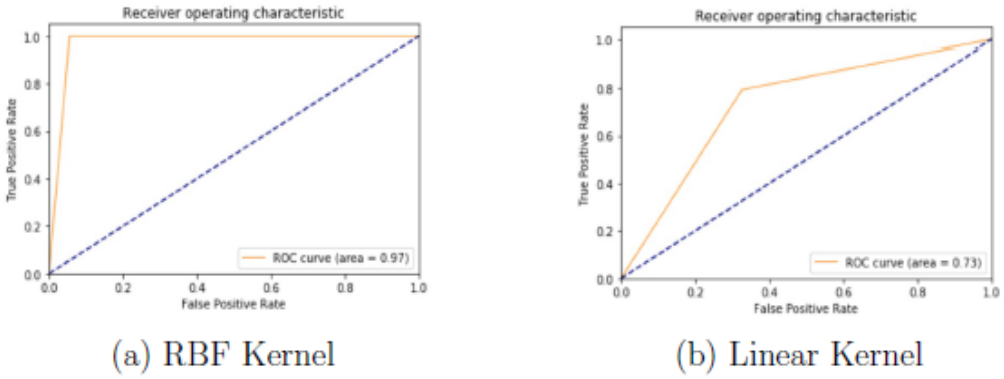

Figure 4. ROC Curves for ISH dataset

Table 2 presents a comparison of the feature types, specifically raw images and Canny images, and the SVM kernel options, namely linear and RBF, across the three datasets being analyzed. It is worth mentioning that the RBF kernel yields the highest performance in all scenarios when applied to raw photos. Therefore, we utilize unprocessed photos that have been adjusted to a size of 16 × 16 for all other studies described in this research work.

Table 2: SVM feature and kernel comparison

Dataset	RBF kernel (Raw)	RBF kernel (Canny)	Linear kernel Raw	Linear kernel Canny
ISH	0.9752	0.9048	0.8838	0.8261
Twitter dataset	0.9685	0.8953	0.9233	0.9265

5. Deep Learning Techniques and Methodology

A CNN, or convolutional neural network, is a specific sort of ANN, or artificial neural network that is utilized for image processing and recognition. It operates by converting images into a pixel format. The neural network is comprised of interconnected nodes known as neurons. Convolutional neural networks offer significant benefits in terms of both efficiency and accuracy when it comes to image analysis. Figure 5 illustrates the organization of neurons into distinct levels, namely the input layer, hidden layer, and output layer

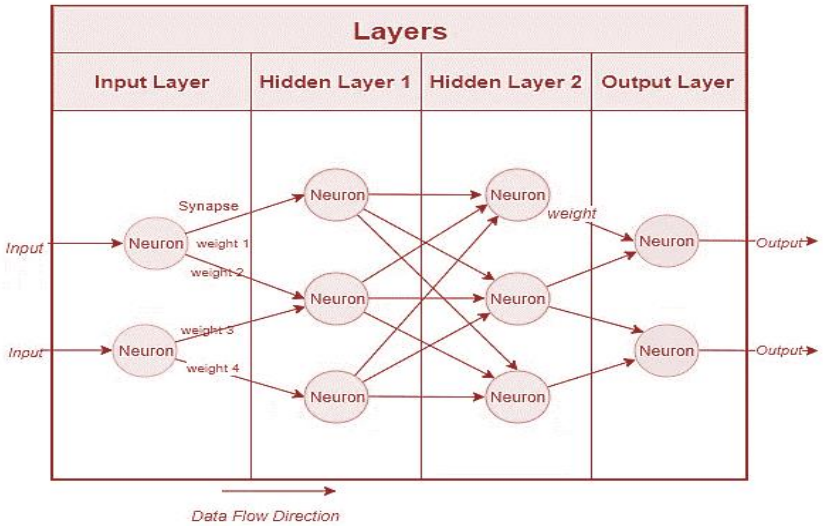

Figure 5. CNN of Input layer, Output layer and Hidden layer

DOI: https://doi.org/10.54216/JCIM.150106
Received: January 29, 2024 Revised: April 28, 2024 Accepted: July 24, 2024

The data is first sent to the hidden layer through the input layers, where it is processed before being passed on to the hidden levels for further processing. During the last stage, the data is transmitted to the relevant layers that handle the final output.

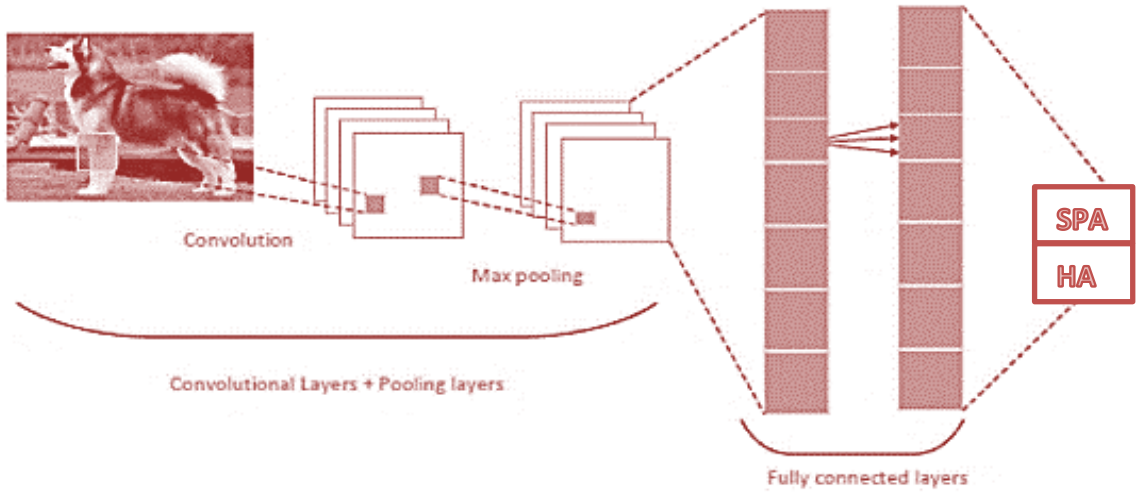

Figure 6. CNN architecture

For example, a picture needs to be identified accurately at this point. For the input layer, the initial step involves arranging the pixels in an array. This marks the initial stage.mThe hidden layer is responsible for extracting features through a variety of operations and calculations. Numerous hidden layers play a crucial role in the extraction processes. Convolution, ReLU, and pooling are three of the layers commonly used in this context. Finally, a layer encompasses all the links and is responsible for accurately determining the photographs. This is the sixth figure.

5.1. Dataset and Preprocessing

We used two different datasets, one of which was the image hunter dataset. You can find it at https://users.cs.northwestern.edu/yga751/ML/ISH.htm.[26][27]. The collection contains a total of 810 regular photographs and 920 spam images. Furthermore, we utilized Twitter's application programming interface. The dataset collected by Image hunter contains a considerable amount of duplicate and damaged files. As part of the initial process, we eliminated the corrupted files and subsequently employed the hashing method to identify any duplicates. If the hash values can be compared, the photos will not be displayed.

Figure 7. Sample of images collected from image spam hunter dataset.

DOI: https://doi.org/10.54216/JCIM.150106
Received: January 29, 2024 Revised: April 28, 2024 Accepted: July 24, 2024

Figure 8. Sample of images extracted from Twitter

5.2. Deep Convolutional Neural Networks (DCNN) Models

An approach known as convolutional neural network is utilized to identify image spam. [19]. the advancements in image processing brought about by CNN and Deep learning architectures have revolutionized the field. These technologies enable the automatic extraction of features, scalability, adaptability, and improved accuracy. Machines are able to understand images in a way that closely resembles human visual perception. Convolutional neural networks are specifically engineered to effectively analyses and interpret spatial data within images. The CNN operations analyses images by taking into account the relationships between neighboring pixels, capturing local patterns, edges, and features. This enables CNN to understand the visual content of images in a way that may present difficulties for traditional machine learning approaches. One of the key advantages of deep learning and specifically CNN's is their remarkable ability to independently extract important features from raw data [20-21]. Eliminating the need for manual feature extraction has made a significant difference in the field of image processing, saving researchers valuable time and effort. A Convolutional Neural Network (CNN) can learn and adapt to the most distinctive features in the data, reducing the need for domain experts.

5.3. Engaged in CNN

In this work, a framework utilizing dual Deep Convolutional Neural networks (DCNN) has been employed. The initial model, known as CNN1, is comprised of three convolutional layers with filter dimensions of 32, 64, and 128, respectively. Each convolutional layer is paired with Rectified Linear Unit (ReLU) activation layers and is subsequently followed by max-pooling layers with a window size of 2. Following the convolutional layers, we have incorporated dropout regularization techniques to enhance the model's capacity for generalization. The model's output is flattened and then passed through a sequence of Dense layers that consist of 128 neurons each [24] [25]. The last layer of the dense architecture employs a single layer with the sigmoid activation function.

In addition, in the later stages of the network, we have included several Machine Learning (ML) classifiers. This study incorporates various ML classifiers, such as Linear Regression, Random Forest, k-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), Linear Support Vector Machine (LSVM), and Radial Support Vector Machine (RSVM) [22][23].

5.4. Environmental Setup

We used Python Jupiter notebook to conduct all of our experiments. We use OpenCV for task processing and utilize a Python library to implement classifiers on Jupiter Notebook. We perform experiments using SVM, CNN, and a range of classifiers [26] [27].

Exploring the CNN Experiment of the Image Spam Hunter dataset.
No, of EPOCHS, we have use 100 and batch size is 32.

DOI: https://doi.org/10.54216/JCIM.150106
Received: January 29, 2024 Revised: April 28, 2024 Accepted: July 24, 2024

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 156, 156, 32)      896

 max_pooling2d (MaxPooling2D  (None, 78, 78, 32)       0
 )

 conv2d_1 (Conv2D)           (None, 78, 78, 64)        18496

 max_pooling2d_1 (MaxPooling  (None, 39, 39, 64)       0
 2D)

 conv2d_2 (Conv2D)           (None, 39, 39, 128)       73856

 dropout (Dropout)           (None, 39, 39, 128)       0

 flatten (Flatten)           (None, 194688)            0

 dense (Dense)               (None, 128)               24920192

 dropout_1 (Dropout)         (None, 128)               0

 dense_1 (Dense)             (None, 1)                 129

=================================================================
Total params: 25,013,569
Trainable params: 25,013,569
Non-trainable params: 0
```

Figure 9. CNN architecture description

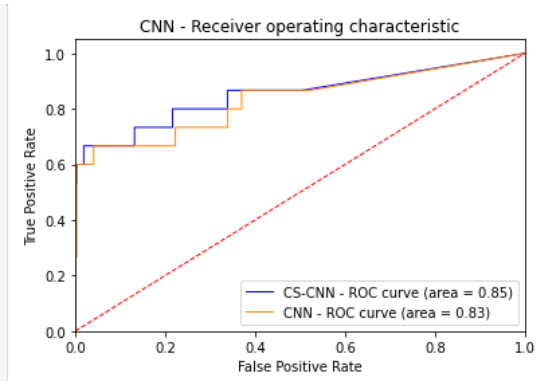

Figure 10. CNN Receiver operating characteristic

6. Result and Discussion

In Figure 11, it is evident that the Logistic Regression of the CNN algorithm has demonstrated impressive performance on the image spam dataset, achieving an accuracy of 97.80%. Logistic Regression achieves an impressive accuracy of 97% for the twitter dataset. LR emerges as the undeniable champion in the Image spam hunter database, boasting an impressive accuracy rate of 97.80%. In figure 12, we achieved the most favorable outcome when compared to other state-of-the-art findings.

DOI: https://doi.org/10.54216/JCIM.150106
Received: January 29, 2024 Revised: April 28, 2024 Accepted: July 24, 2024

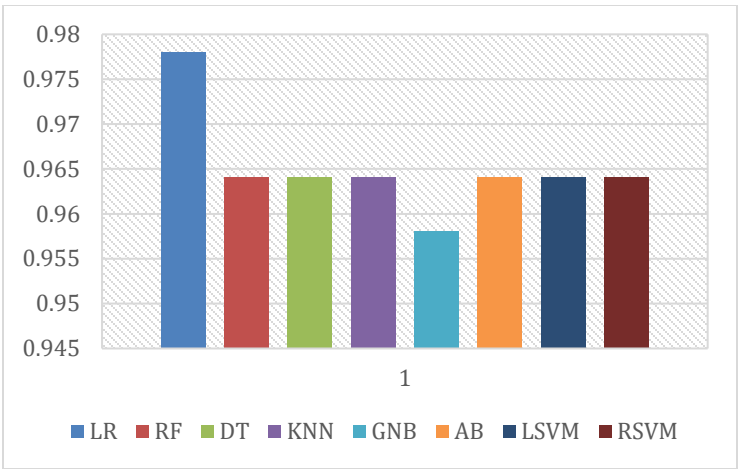

Figure 11. Comparison of various classifiers on image spam hunter dataset

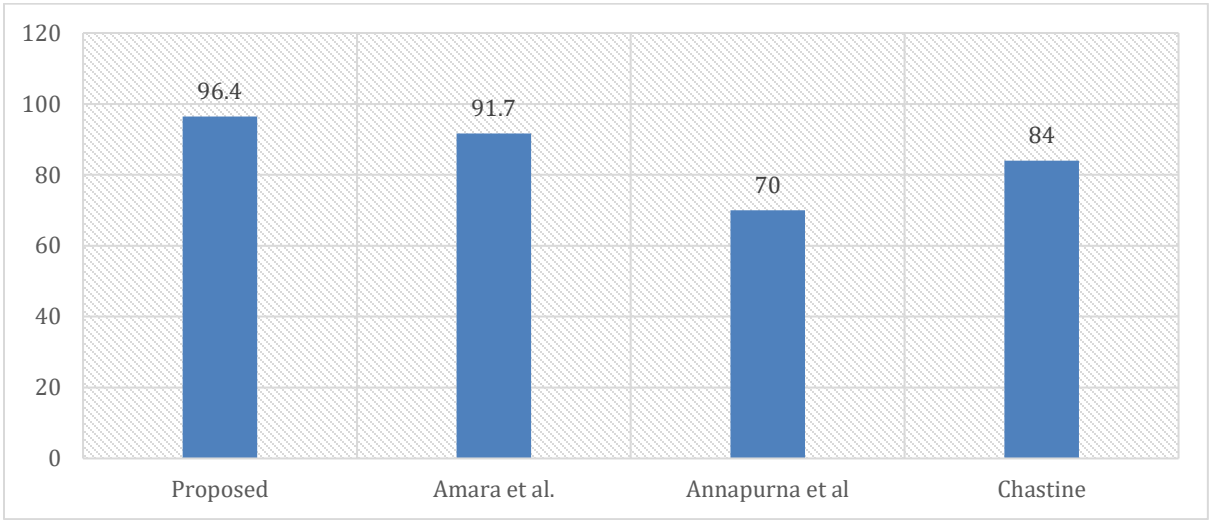

Figure 12. Comparison to previous work Extensive experiment on Twitter dataset based on different classification. Twitter api data

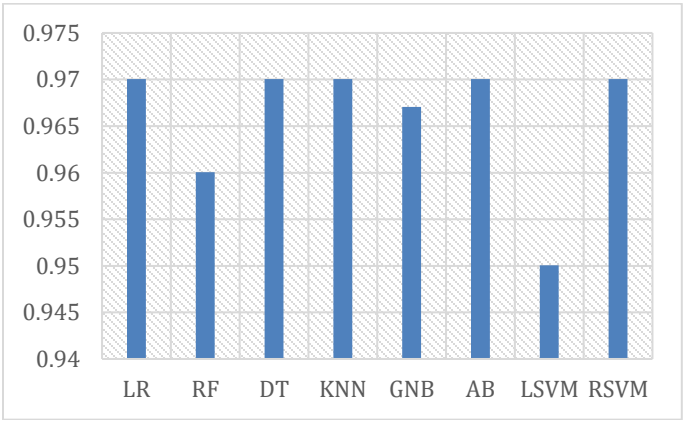

Figure 13. Comparison of various classifiers on Twitter dataset

DOI: https://doi.org/10.54216/JCIM.150106
Received: January 29, 2024 Revised: April 28, 2024 Accepted: July 24, 2024

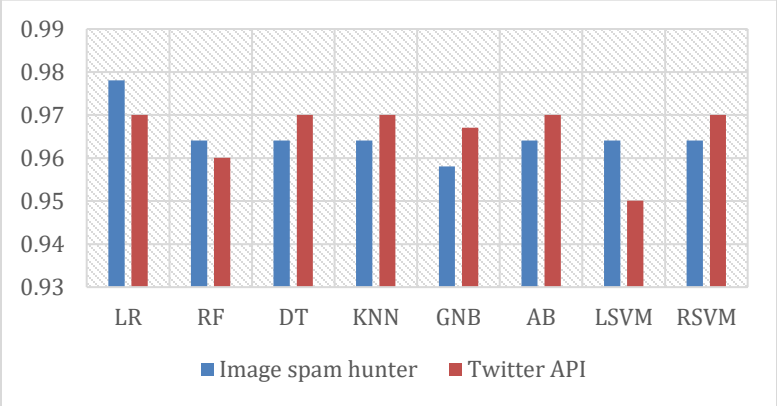

Figure 14. Comparison of image spam hunter database vs twitter dataset.

Table 3 shows the Comparison of images spam hunter dataset vs twitter dataset

Table 3: Comparison of images spam hunter dataset vs twitter dataset

classification	Image spam hunter	Twitter API
LR	0.978071856	0.97005988
RF	0.964071856	0.96005988
DT	0.964071856	0.97005988
KNN	0.964071856	0.97005988
GNB	0.958083832	0.96706587
AB	0.964071856	0.97005988
LSVM	0.964071856	0.95005988
RSVM	0.964071856	0.97005988

Finally, we have successfully implemented CNN models. The architecture of CNN typically includes three types of layers: convolutional layer, subsampling layer, and fully connected layer. Two different architectures of CNN are utilized to evaluate and compare their performance: a 3-layer architecture and a 5-layer architecture. In this study, various parameters of CNN were utilized, including the number of epochs, pool size, dropout rate, and optimizer algorithm. The number of epochs is set to 100, and the optimizer algorithm used is Adam. Additionally, a dropout rate of 0.5 is applied. In the CNN layers, a 2 × 2 pool size is utilized in each layer of the CNN. The initial and subsequent testing scenarios aim to evaluate the efficiency of a 3-layer architecture by analyzing various kernel dimensions. The kernel dimension remains consistent across all layers. The outcomes of the initial and subsequent situations are displayed in Tables 4 and 5. The accuracy of image spam detection is highest at 0.90 for the 3-layer architecture with a 5x5-kernel dimension, and 0.83 for the 5-layer architecture with a kernel dimension of 5x5. When considering the 3-layer and 5-layer architecture, the kernel dimension with the highest accuracy value is chosen. In addition, the CNN method has the advantage of not requiring the feature extraction process. In the traditional classification approach, it is important to choose the right feature extraction method in order to achieve accurate classification outcomes.

DOI: https://doi.org/10.54216/JCIM.150106
Received: January 29, 2024 Revised: April 28, 2024 Accepted: July 24, 2024

Table 4: The results on 3-layer architecture with varying kernel dimension

Kernel dimension	Accuracy	Precision	Recall	Running time (s)
2*2	0.79	0.820	0.65	9610
3*3	0.80	0.815	0.79	10170
4*4	0.82	0.80	0.71	110934
5*5	0.90	0.81	0.62	9432

Table 5: The results on 5-layer architecture with varying kernel dimension

Kernel dimension	Accuracy	Precision	Recall	Running time (s)
2*2	0.80	0.75	0.718	9645
3*3	0.81	0.775	0.79	8976
4*4	0.83	0.80	0.71	9546
5*5	0.83	0.785	0.63	9400

Table 6: Comparing results on 3-layer, 5-layer architectures

CNN architectures	No. of Layer	No. of Neurons	Accuracy	Precision	Recall	Running time (s)
3 layers	8	193	0.95	0.780	0.69	9400
5 layers	12	450	0.96	0.786	0.79	9700

The results show that the highest accuracy achieves 0.96 by using a 5 layers architecture.

7. Conclusion

Images of spam and spam may be difficult to differentiate from one another. This is an extremely challenging issue to find a solution for. In this study, we compare a large number of different classifiers by making use of two distinct data sets. As part of this research, we make use of a convolutional neural network, often known as a CNN, to identify and label spam photos. In addition to the data from over 10,000 Twitter users, we used data from the photo spam hunter dataset. Experiments are carried out on a variety of media, including raw and Canny images, amongst others. The result of the experiment was a significant success when measured against the results of preceding experiments. We tested it on the same dataset, which was the picture spam hunter dataset, and got an accuracy rate that was extremely respectable at 96.0%. We intend to employ different datasets to conduct analyses of picture spam and study additional social media sites like Facebook and WhatsApp so that we can increase the accuracy of our results.

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] N. M. AlShariah and A. K. Jilani Saudagar, "Detecting fake images on social media using machine learning," Int. J. Adv. Comput. Sci. Appl., 2019, doi: 10.14569/ijacsa.2019.0101224.

[2] V. Sharma, I. You, K. Andersson, F. Palmieri, M. H. Rehmani, and J. Lim, "Security, privacy and trust for smart mobile-Internet of Things (M-IoT): A survey," IEEE Access, 2020, doi: 10.1109/ACCESS.2020.3022661.

[3] S. Aphiwongsophon and P. Chongstitvatana, "Detecting fake news with machine learning method," in ECTI-CON 2018 - 15th International Conference on Electrical Engineering/Electronics, Computer,

DOI: https://doi.org/10.54216/JCIM.150106
Received: January 29, 2024 Revised: April 28, 2024 Accepted: July 24, 2024

Telecommunications and Information Technology, 2019. doi: 10.1109/ECTICon.2018.8620051.

[4] F. Aiwan and Y. Zhaofeng, "Image spam filtering using convolutional neural networks," Pers. Ubiquitous Comput., 2018, doi: 10.1007/s00779-018-1168-8

[5] A. M, R. P, S. R, and K. S, "A secure model on Advanced Fake Image-Feature Network (AFIFN) based on deep learning for image forgery detection," Pattern Recognit. Lett., 2021, doi: 10.1016/j.patrec.2021.10.011

[6] S. T. Suganthi et al., "Deep learning model for deep fake face recognition and detection," PeerJ Comput. Sci., 2022, doi: 10.7717/PEERJ-CS.881.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM, 2017, doi: 10.1145/3065386.

[8] G. Choudhary, V. Sharma, I. You, K. Yim, I. R. Chen, and J. H. Cho, "Intrusion Detection Systems for Networked Unmanned Aerial Vehicles: A Survey," in 2018 14th International Wireless Communications and Mobile Computing Conference, IWCMC 2018, 2018. doi: 10.1109/IWCMC.2018.8450305.

[9] T. Sharmin, F. Di Troia, K. Potika, and M. Stamp, "Convolutional neural networks for image spam detection," Inf. Secur. J., 2020, doi: 10.1080/19393555.2020.1722867.

[10] N. Imam and V. G. Vassilakis, "Detecting spam images with embedded Arabic text in twitter," in 2019 International Conference on Document Analysis and Recognition Workshops, ICDARW 2019, 2019. doi: 10.1109/ICDARW.2019.50107.

[11] M. Mateen, M. A. Iqbal, M. Aleem, and M. A. Islam, "A hybrid approach for spam detection for Twitter," in Proceedings of 2017 14th International Bhurban Conference on Applied Sciences and Technology, IBCAST 2017, 2017. doi: 10.1109/IBCAST.2017.7868095.

[12] A. Annadatha and M. Stamp, "Image spam analysis and detection," J. Comput. Virol. Hacking Tech., 2018, doi: 10.1007/s11416-016-0287-x.

[13] C. Fatichah, W. F. Lazuardi, D. A. Navastara, N. Suciati, and A. Munif, "Image spam detection on instagram using convolutional neural network," in Lecture Notes in Networks and Systems, 2019. doi: 10.1007/978-981-13-6031-2_19.

[14] Bhatnagar, Amrita. , Giri, Arun. , Sharma, Aditi. A Hybrid Intrusion Detection Approach for Cyber Attacks. Journal of Journal of Cybersecurity and Information Management, vol. 13, no. 2, 2024, pp. 08-18. DOI: https://doi.org/10.54216/JCIM.130201

[15] B. Kim, S. Abuadbba, and H. Kim, "DeepCapture: Image Spam Detection Using Deep Learning and Data Augmentation," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2020. doi: 10.1007/978-3-030-55304-3_24.

[16] A. Amir, B. Srinivasan, and A. I. Khan, "Distributed classification for image spam detection," Multimed. Tools Appl., 2018, doi: 10.1007/s11042-017-4944-y.

[17] T. C. Lu, "CNN Convolutional layer optimisation based on quantum evolutionary algorithm," Conn. Sci., 2021, doi: 10.1080/09540091.2020.1841111.

[18] D. H. Kim and H. Y. Lee, "Image manipulation detection using convolutional neural network," Int. J. Appl. Eng. Res., 2017.

[19] A. Neisari, L. Rueda, and S. Saad, "Spam review detection using self-organizing maps and convolutional neural networks," Comput. Secur., 2021, doi: 10.1016/j.cose.2021.102274.

[20] A. Choudhary, A. Tripathi, A. Sharma and R. Singh, "Evolution and comparative analysis of different Cloud Access Security Brokers in current era," 2022 International Conference on Fourth Industrial Revolution Based Technology and Practices (ICFIRTP), Uttarakhand, India, 2022, pp. 36-43, doi: 10.1109/ICFIRTP56122.2022.10059416.

[21] V. Sreenivasulu and M. A. Wajeed, "Image based classification of rumor information from the social network platform," Trait. du Signal, 2021, doi: 10.18280/ts.380516.

[22] Ashish Dixit, R. P. Aggarwal, B. K. Sharma, Aditi Sharma. (2023). Safeguarding Digital Essence: A Subband DCT Neural Watermarking Paradigm Leveraging GRNN and CNN for Unyielding Image Protection and Identification. Journal of Journal of Intelligent Systems and Internet of Things, 10 (1), 33-47 (Doi : https://doi.org/10.54216/JISIoT.100103)

[23] V. Gupta, N. Kumar, A. Sharma and A. Abraham, "Sensor Routing Protocol with Optimized Delay and Overheads in Mobile based WSN", Journal of Information Assurance & Security, vol. 16, no. 4, 2021.

[24] K. V. Samarthrao and V. M. Rohokale, "Enhancement of email spam detection using improved deep learning algorithms for cyber security," J. Comput. Secur., 2022, doi: 10.3233/JCS-200111.

[25] G. Stringhini et al., "Follow the green: Growth and dynamics in Twitter follower markets," in Proceedings of the ACM SIGCOMM Internet Measurement Conference, IMC, 2013. doi: 10.1145/2504730.2504731.

DOI: https://doi.org/10.54216/JCIM.150106
Received: January 29, 2024 Revised: April 28, 2024 Accepted: July 24, 2024

[26] R. Asif and M. A. Islam, "Finding most collaborating mathematicians a co-Author network analysis of mathematics domain," in 2016 International Conference on Computing, Electronic and Electrical Engineering, ICE Cube 2016 - Proceedings, 2016. doi: 10.1109/ICECUBE.2016.7495240.

[27] J. Hussain and M. A. Islam, "Evaluation of graph centrality measures for tweet classification," in 2016 International Conference on Computing, Electronic and Electrical Engineering, ICE Cube 2016 - Proceedings, 2016. doi: 10.1109/ICECUBE.2016.7495209.

[28] Q. Li, Z. Qin, L. Chai, H. Zhang, J. Guo, and B. Bhanu, "Representative reference-set and betweenness centrality for scene image categorization," in 2013 IEEE International Conference on Image Processing, ICIP 2013 - Proceedings, 2013. doi: 10.1109/ICIP.2013.6738670.

[29] Jiang Li, William Rich, Donald Buhl-Brown,"Texture Analysis of Remote Sensing Imagery with Clustering and Bayesian Inference", IJIGSP, vol.7, no.9, pp.1-10, 2015.DOI: 10.5815/ijigsp.2015.09.01

[30] Cariow, A.; Papli´nski, J.P.;Makowska, M. VLSI-FriendlyFiltering Algorithms for Deep NeuralNetworks. Appl. Sci. 2023, 13, 9004.https://doi.org/10.3390/app13159004

DOI: https://doi.org/10.54216/JCIM.150106
Received: January 29, 2024 Revised: April 28, 2024 Accepted: July 24, 2024