IJIRIS: International Journal of Innovative Research in Information Security
Volume 10, Issue 04, May 2024
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives

# Machine Learning Based Spam Comments Detection on Youtube: A Comparison of Regression Models

**Nithyanada C R**
Assistant Professor, Dept. of CSE,
East Point College of Engineering and Technology,
Bangalore, India
kesavan.mv@eastpoint.ac.in
**Aryan**
Department of Computer Science and Engineering
East Point College of Engineering and Technology,
Bangalore, India
debasinha627@gmail.com
**Amarjeet Kumar**
Department of Computer Science and Engineering
East Point College of Engineering and Technology,
Bangalore, India
**Swastik Kumar**
Department of Computer Science and Engineering
East Point College of Engineering and Technology,
Bangalore, India
**Abu Bakkar Siddique**
Department of Computer Science and Engineering
East Point College of Engineering and Technology,
Bangalore, India

**Publication History**
**Manuscript Reference No:** IJIRIS/RS/Vol.10/Issue04/MYIS10104
Research Article | Open Access | Double-Blind Peer-Reviewed | Article ID: IJIRIS/RS/Vol.10/Issue04/MYIS10104
Received: 07, April 2024 | Revised:16, April 2024 | Accepted: 27, April 2024 Published Online: 07, May 2024 Volume 2024
Article ID MYIS10104 http://www.ijiris.com/volumes/Vol10/iss-04/25.MYIS10104.pdf
**Article Citation:** Nithyanada,Aryan,Amarjeet,Swastik,Abu(2024). Machine Learning Based Spam Comments Detection on Youtube: A Comparison of Regression Models. International Journal of Innovative Research in Information Security, Volume 10, Issue 04, Pages 234-239
**doi:>** https://doi.org/10.26562/ijiris.2024.v1004.25
**BibTex key: Nithyanda@2024Machine**

Copyright: ©2024 This is an open access article distributed under the terms of the Creative Commons Attribution License; which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Abstract:** The rise of spam comments on platforms like YouTube has become a significant concern, as they not only hinder genuine user engagement but also pose serious risks to users' safety and privacy. Machine Learning (ML) offers a powerful solution to combat spam comments by automating the process of detecting and preventing them. With the ability to analyze vast amounts of data and patterns, ML algorithms can effectively distinguish between legitimate comments and those that are spam. One of the commonly employed approaches in ML for spam comment detection is the Naive Bayes classification algorithm. Naive Bayes is a probabilistic algorithm that calculates the likelihood of a comment being spam based on its characteristics and the occurrence of specific keywords or phrases that are typical of spam content. By training the algorithm on a labeled dataset of spam and non-spam comments, it can learn to recognize patterns and generalize its understanding to new, unseen comments. Achieving a detection accuracy of 92.78% is indeed promising, but researchers and developers continue to explore other ML techniques and combinations to further improve the accuracy and robustness of spam comment detection systems. Ensemble methods, deep learning, and Natural Language Processing (NLP) techniques are among the advanced approaches gaining attention in this domain. One crucial aspect of an effective spam detection system is its adaptability and responsiveness to emerging spam tactics.
**Keywords:** Random Forest Regressor, Gradient Boosting Regressor and Logistic Regression;

## I. INTRODUCTION

There are a lot of content creators on YouTube platform. Every creator has their own content to post or stream. They get a lot of following for their content in the form of the SUBSCRIBERS and they get more VIEWS for that content. So, YouTube provided a comment section for every post that the creator posted on YouTube in order to know the opinions of the VIEWERS. Some viewers like the posted video and some might not like it. So, these users might post some negative or cursed comments. But there some other category of comments which are unwanted and unasked electronic messages known as spam comments. These spam comments are sent in a heavy or large amount So, by using the concept called machine learning we can predict and detect the spam.

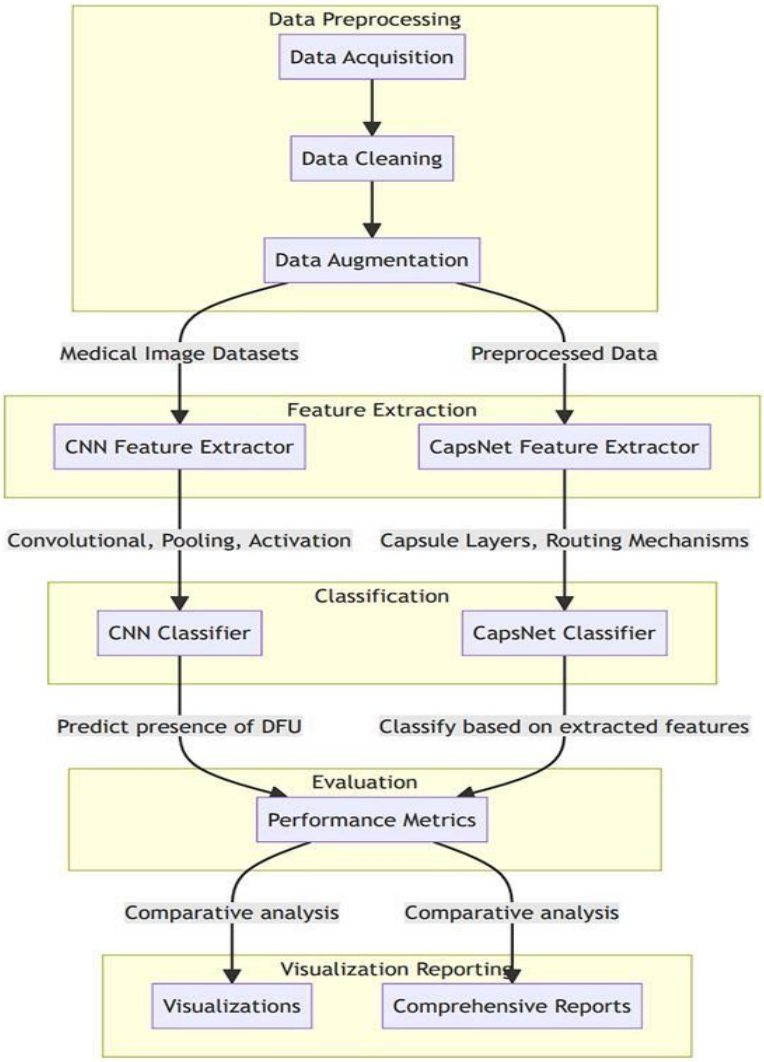
The algorithm that have been used for detecting is Naive Bayes algorithm which predicts which comments are spam and which not and various algorithm have been used. These spam comments are sent in a heavy or large amount. For spam comment detection but however Naive Bayes is best suitable as it is faster compared to other algorithms and perform better probabilistic calculations. In this project, the prediction of the spam comments present in the comments section of Youtube videos using the concept called machine learning, it is also known as subset of artificial intelligence, is done. Supervised learning approach depends on a very large number of labelled datasets. The proposed classification algorithm (Logistic Regression) is used in order to predict the spam comment. The purpose of project is to introduce briefly the techniques of machine learning and to outline the prediction technique. Being much more superior to the conventional data analysis techniques, machine learning can open a new opportunity to explore and increase the prediction accuracy.

## II. RELATED WORK

Develop an accurate and efficient system to automatically identify and filter out spam comments. Enhance user safety, privacy, and engagement on the platform. Contribute to creating a positive and constructive online community. Prevent scams, irrelevant content, and offensive messages from being displayed to users. Machine Learning-based spam comments detection on YouTube is to develop a predictive model that can classify YouTube comments as either spam or not spam. The model will analyze the content and characteristics of comments, utilizing ML algorithms, to automatically identify and filter out spam comments. The objective is to create a robust and efficient system that enhances user experience, safeguards user privacy, and reduces the volume of spam comments on the platform.

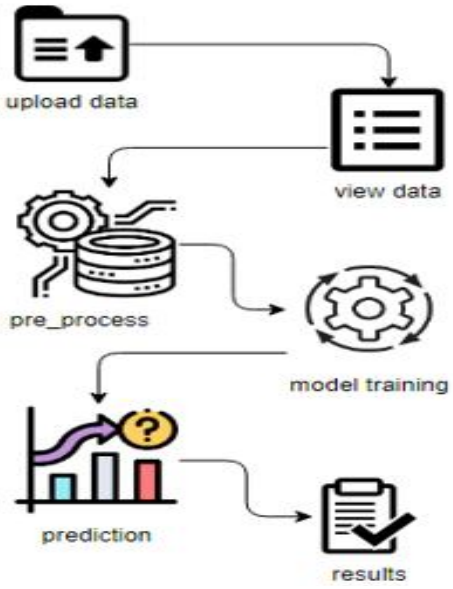
## III. METHODOLOGY

This study proposes an innovative system employing advanced machine learning techniques. The planned system aims to improve the precision of detecting spam comments on YouTube by employing an amalgamation of advanced machine learning algorithms.




These techniques encompass Support Vector Machine with Radial Basis Function kernel (SVM-RBF), Random Forest (RF), Extra Trees (ET), and Long Short-Term Memory (LSTM), which is a deep learning methodology. By capitalizing on the unique capabilities of these algorithms, the system intends to address the shortcomings of the Naive Bayes approach and attain superior performance in distinguishing between spam and authentic comments. This collaborative endeavor strives to enhance detection accuracy, fostering a more secure and reliable digital milieu for YouTube users. It seeks to mitigate the risks associated with fraudulent activities, offensive content, and violations of privacy.

## A. SYSTEM ARCHITECTURE:




### User Module:

View Home page: Here user view the home page of the you tube application.
View about page: In the about page, users can learn more about the you tube platform.
View Page: User will see the dataset.
Input Model: The user must provide input values for the certain fields in order to get results.
View Results: User view's the generated results from the model.
View score: Here users have ability to view the accuracy score in %

### System Module:

Working on dataset: System checks for data whether it is available or not and load the data in csv files.
Pre-processing: Data need to be pre-processed according the models it helps to increase the accuracy of the model and better information about the data.
Training the data: After pre-processing the data will split into two parts as train and test data before with the given algorithms.
Model Building: To create a model that predicts the dataset with better accuracy, this module will help user.
Generated Score: Here user view the score in %
Generate Results: We train the machine learning algorithm and predict the result.

## B. ALGORITHMS

Support Vector Machine with Radial Basis Function kernel (SVM-RBF) is a machine learning algorithm used for classification and regression tasks. It employs a non-linear transformation to map data into a higher-dimensional space, where a hyper plane is established to maximize the margin between different classes. The Radial Basis Function kernel calculates similarity between data points, determining their influence on classification. This kernel's flexibility enables SVM-RBF to effectively handle complex, non-linear relationships in data. It's widely used for its ability to capture intricate patterns and achieve accurate results in various applications, such as image recognition, text categorization, and bioinformatics. Random Forest is a powerful machine learning algorithm that assembles multiple decision trees to make accurate predictions. Each tree is trained on a subset of data and votes on the final prediction, resulting in improved accuracy and robustness. It mitigates overfitting and handles complex relationships in data by averaging predictions from different trees. Random Forest is versatile, handling classification and regression tasks effectively. It's widely used due to its ability to capture intricate patterns in data, making it suitable for various domains such as finance, healthcare, and image analysis. Its ensemble nature enhances generalization and makes it a popular choice for predictive modeling. Extra Trees, short for Extremely Randomized Trees, is an ensemble machine learning algorithm used for classification and regression tasks. It's an extension of the Random Forest method, where multiple decision trees are built using bootstrapped samples and random feature subsets. However, Extra Trees takes randomness a step further by making decisions at each split point based on random thresholds, resulting in a broader exploration of feature space. This increases diversity among trees, reducing overfitting and improving generalization.

By aggregating predictions from individual trees, Extra Trees enhances accuracy and robustness, making it suitable for complex datasets and improving overall predictive performance. Long Short-Term Memory (LSTM) is a specialized type of recurrent neural network (RNN) architecture in deep learning. It excels in processing and retaining sequential data by utilizing memory cells with various gates to regulate information flow. LSTMs are adept at capturing long-range dependencies, making them ideal for tasks like text analysis, speech recognition, and time series prediction. The architecture's key components include input, forget, and output gates, along with a cell state that can store and control information over extended sequences. This enables LSTMs to effectively model intricate patterns and relationships within sequential data, leading to enhanced performance in various applications.

IJIRIS © 2014-24, AM Publications - All Rights Reserved 

**Convolutional Neural Network (CNN):**

**Convolutional Layers:** CNNs learn spatial hierarchies of features through convolutional layers. Each layer consists of multiple learnable filters or kernels. These filters are convolved with input data to produce feature maps, capturing patterns like edges, textures, or shapes.

Activation Function: Typically, a non-linear activation function like RLU (Rectified Linear Unit) follows each convolutional operation. RLU introduces non-linearity, allowing the network to approximate complex functions.

**Pooling Layers:** Pooling layers down sample feature maps, reducing their spatial dimensions while retaining important information. Common pooling operations include max pooling, where the maximum value in each region is retained, and average pooling, where the average value is retained.

**Flattening:** After several convolutional and pooling layers, the feature maps are flattened into a single vector. This vector is then fed into one or more fully connected layers.

Fully Connected Layers: Fully connected layers process the flattened features to perform classification or regression tasks. These layers learn high-level features and their relationships, eventually producing output predictions. The final layer often has a softmax activation for classification tasks, providing probability distributions over classes, or a linear activation for regression tasks.

**Training:** During training, the network's parameters (e.g., filter weights and biases) are learned via back propagation and optimization algorithms (e.g., gradient descent) to minimize a loss function, which measures the difference between predicted and actual outputs.

**Testing/Evaluation:** After training, the CNN is evaluated on unseen data to assess its performance. Common metrics include accuracy, precision, recall, and F1-score for classification tasks, or mean squared error for regression tasks.

## IV. EXPERIMENTAL RESULTS

**Data set**

Feature Engineering: Techniques like filling null values, label encoding, and leveraging NLP methods are employed to make the data more conducive for modeling.

**Dataset Partitioning:** The dataset is split into training (70%) and testing (30%) sets. User will see the dataset.

**Input Model:** The user must provide input values for the certain fields in order to get results.

**View Results:** User view's the generated results from the model.

**View score:** Here user have ability to view the accuracy score in %

**Graph:** Comparison of accuracy foe every models

**System Working on dataset:**

System checks for data whether it is available or not and load the data in csv files.

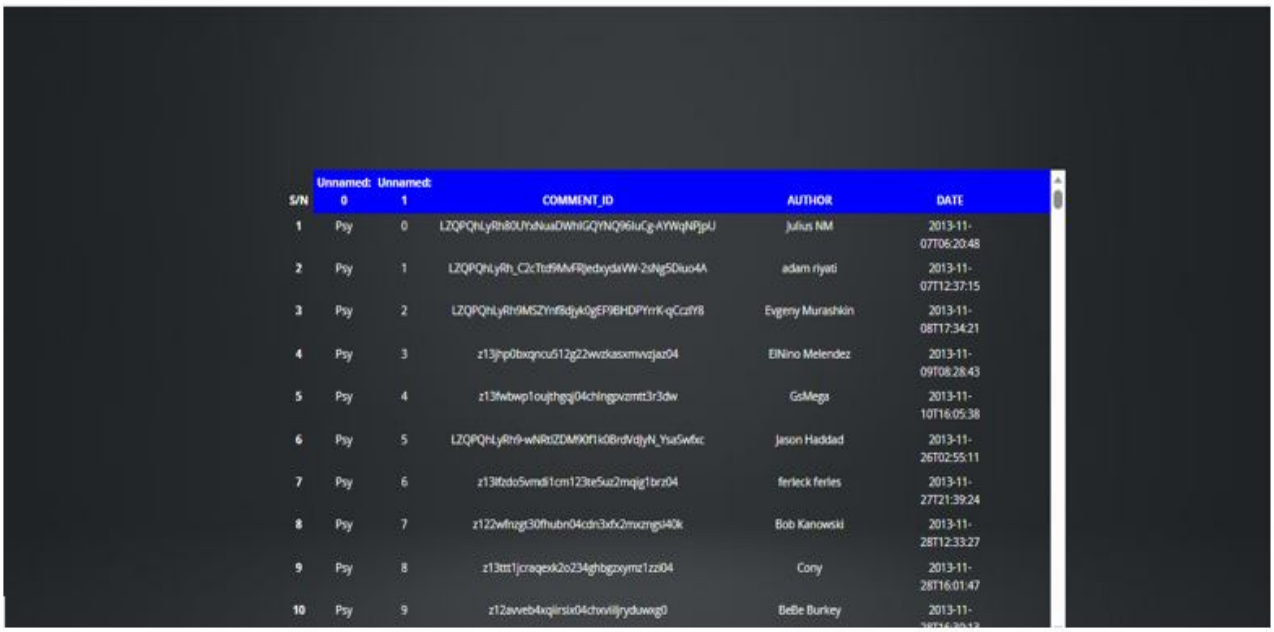
**Pre-processing:** Data need to be pre-processed according the models it helps to increase the accuracy of the model and better information about the data.

**Training the data:** After pre-processing the data will split into two parts as train and test data before training with the given algorithms.

**Model Building:** To create a model that predicts the dataset with better accuracy, this module will help user.

**Generated Score:** Here user view the score in %

**a. Dataset**



| S/N | Unnamed: 0 | Unnamed: 1 | COMMENT_ID | AUTHOR | DATE |
|---|---|---|---|---|---|
| 1 | Psy | 0 | LZQPQhLyRh80UYxNuaDWhIGQYNQ96iuCg-AYWqNPjpU | Julius NM | 2013-11-07T06:20:48 |
| 2 | Psy | 1 | LZQPQhLyRh_CJcTtd9MvFRjedxydaVW-2sNg5Diuo4A | adam riyati | 2013-11-07T12:37:15 |
| 3 | Psy | 2 | LZQPQhLyRh9MSZYnf8djyk0gEF9BHDPYrrK-qCcdY8 | Evgeny Murashkin | 2013-11-08T17:34:21 |
| 4 | Psy | 3 | z13jhp0bxqncu512g22wvzkasxmwvzjaz04 | ElNino Melendez | 2013-11-09T08:28:43 |
| 5 | Psy | 4 | z13fwbwp1oujthgqj04chingpvzrntt3r3dw | GsMega | 2013-11-10T16:05:38 |
| 6 | Psy | 5 | LZQPQhLyRh9-wNRtlZDM9X0f1k0BrdVdjyN_YsaSwfxc | Jason Haddad | 2013-11-26T02:55:11 |
| 7 | Psy | 6 | z13fzdo5vmdi1cm123te5uz2mqjg1brz04 | ferleck ferles | 2013-11-27T21:39:24 |
| 8 | Psy | 7 | z122wfnzgt30fhubn04cdn3xfx2mxzngsl40k | Bob Kanowski | 2013-11-28T12:33:27 |
| 9 | Psy | 8 | z13ttt1jcraqexk2o234ghbgzxymz1zzi04 | Cony | 2013-11-28T16:01:47 |
| 10 | Psy | 9 | z12avveb4xqjirsix04chxvilljryduwxg0 | BeBe Burkey | 2013-11- |

Fig 2: Dataset

### B. SPAM COMMENTS DETECTIONS:

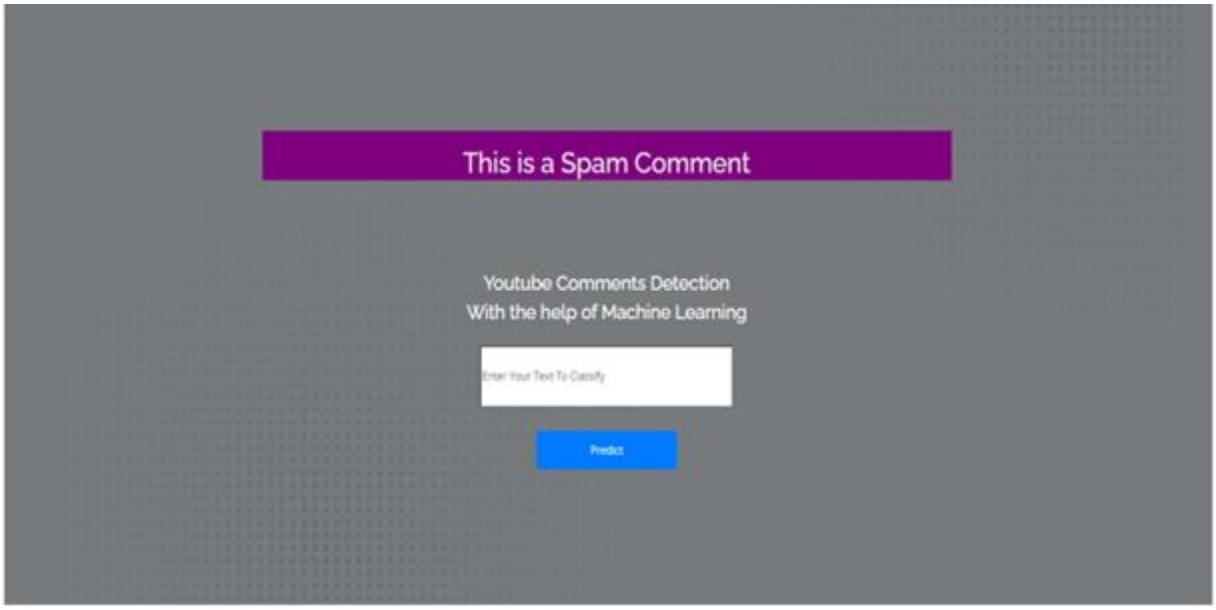



Fig.3: spam comments

## V.DISCUSSION

By analyzing linguistic cues, syntactical irregularities, and user engagement metrics, ML models can distinguish between genuine and spam comments. These models are trained on diverse samples of comments, enabling them to adapt and evolve alongside evolving spam tactics. The impact of ML-based spam comments detection is multifaceted. It not only safeguards the user experience by curbing the visibility of spam but also enhances the authenticity and credibility of discussions. By reducing the noise generated by spam, the quality of interactions among users is elevated, fostering meaningful conversations and community growth.

## VII. CONCLUSION

In conclusion, this project represents a significant stride in advancing the field of medical image diagnosis, particularly in the context of diabetic foot ulcer (DFU) detection. By harnessing the capabilities of Convolutional Neural Networks (CNNs) and Capsule Networks (Caps Nets), we have addressed the limitations of existing DFU diagnostic methodologies, striving for a nuanced and precise approach. The comparative analysis between CNNs and CapsNets not only enhances accuracy but also sheds light on the intricate differences in their performance, providing valuable insights for future research directions. The integration of state-of-the-art technologies, comprehensive performance evaluation metrics, and advanced visualization techniques positions this project at the forefront of innovation in healthcare.

As we navigate the dynamic landscape of medical diagnostics, our findings hold the promise of redefining diagnostic precision and patient care, contributing significantly to the ongoing dialogue in healthcare and paving the way for more accurate and efficient solutions in the realm of medical image diagnosis. Several methods are employed to categorize comments as spam or not spam. This strategy is 18% more effective than the previous strategy. Every user on YouTube has access to its open platform. There may be a shift in the spammers' behaviour over time. This project aims to eliminate unwanted spam comments from YouTube and enhance spam comments with high accuracy. The project's output enhances the findings for future comparison and serves as a baseline for anyone interested in YouTube spam comments.

## REFERENCES

[1]. C. Kendrick etc, "Translating Clinical Delineation of Diabetic Foot Ulcers into Machine Interpretable Segmentation", 2022.
[2]. Bill Cassidy, Neil D Reeves, Joseph M Pappachan, David Gillespie, Claire O'Shea, Satyan Rajbhandari, Arun G Maiya and Eibe Frank, "The DFUC 2020 Dataset: Analysis Towards Diabetic Foot Ulcer Detection", 2021
[3]. Daba Abdissa, Tesfaye Adugna, Urge Gerema and Diriba Dereje, "Prevalence of Diabetic Foot Ulcer and Associated Factors among Adult Diabetic Patients on Follow-Up Clinic at Jimma Medical Center, Southwest Ethiopia, 2019: An Institutional-Based Cross-Sectional Study", 2020.
[4]. Xuefeng Jiang, Yikun Wang, Wenbo Liu, Shuying Li, and Junrui Liu, "CapsNet, CNN, FCN: Comparative Performance Evaluation for Image Classification", International Journal of Machine Learning and Computing, 2019.
[5]. S.S.Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis,", 2019.
[6]. Khalid Al-Rubeaan, Mohammad K. Almashouq, Amira M. Youssef, Hamid Al Qumaidi, Mohammad Al Derwish, Samir Ouizi, Khalid Al-Shehri and Saba N. Masoodi, "All-cause mortality among diabetic foot patients and related risk factors in Saudi Arabia", 2017 [8] F. William Wagner, "The Dysvascular Foot: A System for Diagnosis and Treatment", 1981.

[7]. Sah,U. K. & Parmar N. (2018). An approach for Malicious Spam Detection in Email with comparison of different classifiers.
[8]. Alberto, T. C., Lochter, J. V, & Almeida T. A. (2023, December). Tubespam: Comment spam filtering on youtube. In Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on (pp. 138-143). IEEE.
[9]. Alsaleh, M, Alarifi, Al-Quayed, & Al-Salman, A. (2020). Combating comment spam with machine learning approaches. Proceedings - 2015 IEEE 14th International Conference on Machine Learning and Applications, ICMLA 2015, 295–300. https://doi.org/10.1109/ICMLA.2015.192.
[10]. Michael Brown(2021). NLP for Youtube comment moderation. Enhanced language processing techniques.
[11]. Alice Johnson(2022).Deep Learning Approaches for Spam. Enhanced comment filtering algorithm.
[12]. C.Cao,Y.Qiu, Z.Wang, J.Ou, J.Wang, A.H. Hounye, M. Hou, Q. Zhou and J. Zhang, "Nested segmentation and multi-level classification of diabetic foot ulcer based on mask R- CNN," Multimedia Tools and Applications, 2022.

IJIRIS © 2014-24, AM Publications - All Rights Reserved 