IMAGE SPAM FILTERING USING MACHINE LEARNING TECHNIQUES

Abhishek Rungta, Bhawna Arya, G. Usha

Abstract**:** ***Unsolicited visual data is undesirable in any form. The art of hiding malicious content in images and adding them as attachments to electronic mails has become a popular nuisance. In recent years, attackers have developed various new techniques to evade traditional spam classification systems. Text-based spam classification has been in focus for a long time and, researchers have successfully created a prodigal system for identifying spam text in electronic mails using Optical Character Recognition technology. In the last decade, extensive work has been performed to tackle image spam but with unsatisfactory results. Various algorithms and data augmentation techniques are used today to develop an optimal model for image spam recognition. Many of these proposed systems come close to the ideal system but do not provide 100 percent accuracy. This paper highlights the role of three popular techniques in image spam filtering. We discuss the importance and application of Optical Character Recognition, Support Vector Machines and, Artificial Neural Networks in unsolicited visual data filtering. This paper sheds light on the algorithms of these techniques. We provide a comparison of their accuracy, which helps us draw useful insights for developing a robust unsolicited visual data classification system. This paper aims to bring clarity regarding the feasibility of using these techniques to develop an unsolicited visual data filtering system. This paper records that the most favourable results are obtained using Artificial Neural Networks.***

Index Terms**:** ***Artificial Neural Networks, Data Augmentation, Optical Character Recognition, Support Vector Machines***

I. INTRODUCTION

Unsolicited visual data refers to unwanted images and videos content shared as electronic mail attachments. These files generally contain spam content used to influence, sell or promote an idea. At times, these emails contain malware attachments that can compromise the security of an individual.Wasted bandwidth, choked memory space, and utter annoyance are some of the reasons which increase the irritation associated with image spam. Typical spam classification systems fail to detect such spam content in image form. Spammers have developed a new craft of sending broken images as email attachments. Adding subconscious messages in the frames of an animated Graphics Interchange Format (GIF) is another popular practice. Novice systems designed as anti-spam undoubtedly fail to classify these images.

Spam attackers actively strive to develop new methods to sabotage the efficiency of anti-spam systems. They deploy many disguises [1] such as multi-frame animated GIFs, hand-written images, geometric variance, and noisy images. The latest trends in visual data spamming are forge-header information, cartoon recolouring, forge-sender information, template-driven, patchy font and randomization. The major obstacle faced by amateur classification systems is the lack of diverse datasets which can help train models to effectively classify spam images and videos.

Our paper is a comprehensive study on unsolicited visual data filtering and its techniques. Firstly, we focus on Optical Character Recognition (OCR), Artificial Neural Network (ANN) and Support Vector Machines (SVM) techniques used to classify unsolicited visual data by providing a brief about their working and algorithms. We tabulate the results by highlighting the distinction between approaches and their accuracy. We also use various visual stimuli to represent the working of each technique adequately.

The paper is structured as follows: Section II introduces a diagrammatic representation of the system architecture of a generic anti-spam filter model and Section III presents an outline of the three techniques popularly used in anti-spam systems- Optical Character Recognition, fabricated Artificial Neural Network and Support Vector Machines. Lastly, we focus on results and discussions.

II. SYSTEM ARCHITECTURE

As shown in figure 1, we represent the concept of operations involved in an anti-spam system. We aim to visualize the working of various physical and logical components involved by providing a high-level view of the system. There exists a database that stores multiple files which act as input to the filter model.

The input is transformed to meet the requirements of the classification system. Most of these systems run on a cloud and provide an output- the nature of the image.

Revised Manuscript Received on July 05, 2019.

Abhishek Rungta, Software Engineering, SRM Institute of Science and Technology, Chennai, India. abhishekrungta_ar@srmuniv.edu.in

Bhawna Arya, Software Engineering, SRM Institute of Science and Technology, Chennai, India.

Dr. G. Usha, Associate Professor, Software Engineering, SRM Institute of Science and Technology, Chennai, India.

Retrieval Number: B10350782S419/2019©BEIESP
DOI: 10.35940/ijrte.B1035.0782S419

Published By:
Blue Eyes Intelligence Engineering
Sciences Publication

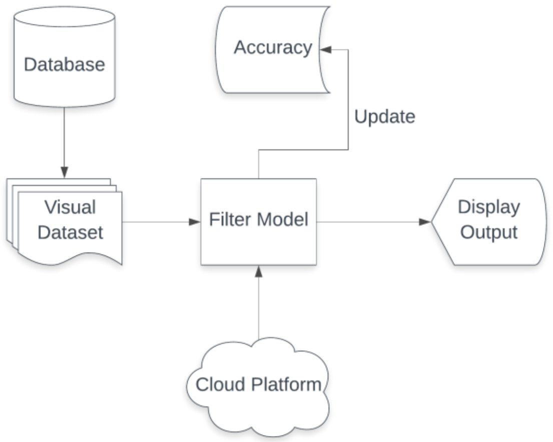

Fig 1. System Architecture of an anti-spam system

In the next section, we give a brief description of the various categories of spam. Following that, we elaborate and give an overview of OCR, ANN and SVM techniques.

III. CLASSIFICATION TECHNIQUES

According to [2], spam accounts for 60% of all inbound electronic mail exchanges. Spam attackers deploy new approaches to transfer malicious content over the internet. Some of the classes of spam [3] are randomization, colouring, wild backgrounds, text only, and broken images [4]. Spam advertisements are a famous means of hype for sales marketers. The aim of the spam classification is to distinguish between spam image and useful image. Spam attachments spread information through advertisements of products. They [5], often include messages of financial, adult, political or spiritual nature. The spam text embedded in the images cannot be detected using simple OCR classifiers. Hence, a need arises to develop a robust system which can handle all spam types efficiently.

In this section, we discuss in detail about unsolicited visual data classification techniques. We begin by describing the Optical Character Recognition Technique [6] used to detect the optical patterns in visual data. Next, we give information on neural networks and highlight the key features of an artificial neural network. In the end, we summarize the SVM classification technique and its algorithm.

A. Optical Character Recognition

Optical Character Recognition [7] is used to detect optical patterns obtainable from a digital image. It is the machine-driven or auto electronic conversion of scanned images of in writing, typewritten or printed text into machine text. It is a three-stage process. Firstly, the image undergoes segmentation. Then features are extracted from an image. At last, image classification occurs. This technique allows us to transform text caught in a digital image into editable document format. An OCR engine observes all characters present in an image from the available classes. It gives near accurate results for text-based spam classification. The accuracy and efficiency of optical character recognition are not reliable as it cannot identify images concealed by CAPTCHA [8]. Moreover, the technique is expensive to compute. It does not handle a large database well.

Optical Character Recognition [9] is useful for unsolicited visual data classification of text-only images. It can identify characters written in a digital image. In [10], the author uses pattern recognition in combination with Optical Character Recognition technology to detect spam images sent over the email. The model developed is an improved approach to traditional Optical Character Recognition technology. Many authors [11] have explored the domain of detecting image spam using Optical Character Recognition technology by creating an engine that recognizes a certain set of words frequently occurring in the image. The accuracy of such models is respectable, but these systems fail to observe noisy or broken images efficiently. Furthermore, there are not enough datasets available to train a model to its full potential.

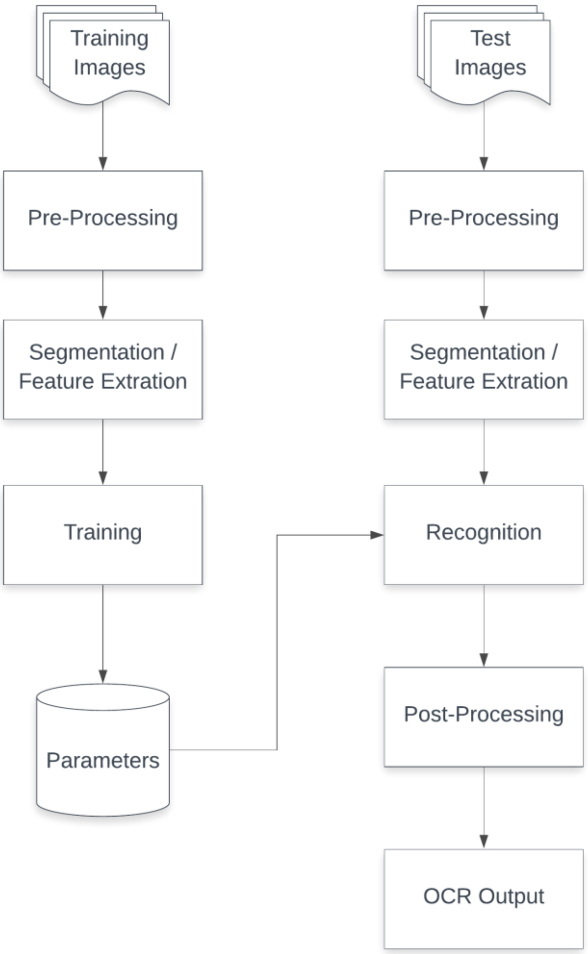

Fig 2. Optical Character Recognition Sequence

As shown in figure 2, we elaborate on the working sequence of an OCR system. We depict all the stages, starting with an initial input image and ending with the final output text. Next, we discuss the use of Artificial Neural Networks in unsolicited visual data classification. We provide the

DOI: 10.35940/ijrte.B1035.0782S419

Published By:
Blue Eyes Intelligence Enginee
& Sciences Publication

algorithmic framework for using such networks for model training.

B. Artificial Neural Network

An artificial neural network [12] is a shred of the vast artificial intelligence system. It consists of fundamental processing units called neurons. This network replicates the working of the human nervous system. Artificial neural network imitates the activity of neurons in the human brain. Every neuron is capable of training and testing, that is each neuron can be taught to deliver an output. Multiple neurons commit and form a single neural network layer. Their sequence forms an artificial neural network.

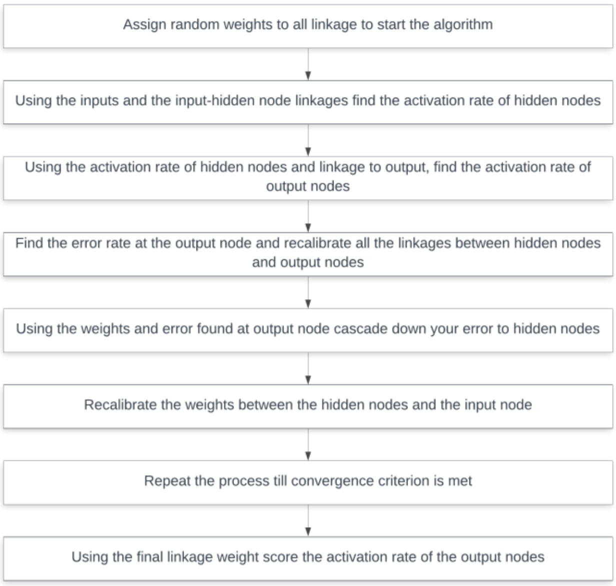
Assign random weights to all linkage to start the algorithm
Using the inputs and the input-hidden node linkages find the activation rate of hidden nodes
Using the activation rate of hidden nodes and linkage to output, find the activation rate of output nodes
Find the error rate at the output node and recalibrate all the linkages between hidden nodes and output nodes
Using the weights and error found at output node cascade down your error to hidden nodes
Recalibrate the weights between the hidden nodes and the input node
Repeat the process till convergence criterion is met
Using the final linkage weight score the activation rate of the output nodes

Fig 3. Artificial Neural Network Framework

In [13], the author uses an optimized approach for image spam filtering. They use data augmentation to improve the quality and quantity of the data sample. They discuss techniques using Adam as a network optimizer and weighted discriminative spatial pyramid for feature extraction. They make use of the Convolutional Neural Network (CNN) architecture of Artificial Neural Network. Their CNN model provides genuinely accurate results on the improved dataset. Artificial Neural Network is a non-parametric classifier with high computing efficiency, but the network training time and overfitting risks are high. According to [14], there is only one critical difference between a traditional Artificial Neural Network and Convolutional Neural Network. A Convolutional Neural Network has a fully connected last layer, on the other hand, an Artificial Neural Network has each neuron connected to every other neuron.

As shown in figure 3, we explain the framework on which Artificial Neural Network works. We describe the sequential process with a sufficient description of every process step. In the next section, we discuss the benefit of using Support Vector Machines for unsolicited visual data classification.

C. Support Vector Machine

Support Vector Machines use hyperplanes or a set of hyperplanes to classify data when image data is in more quantity or infinite dimensions. It is a binary classifier and, separates two classes with the help of a linear boundary. A good classification observes that the nearest training point of any class holds the farthest reach from a particular hyperplane. Support vector machines [15-16] impartially optimize the use of training data.

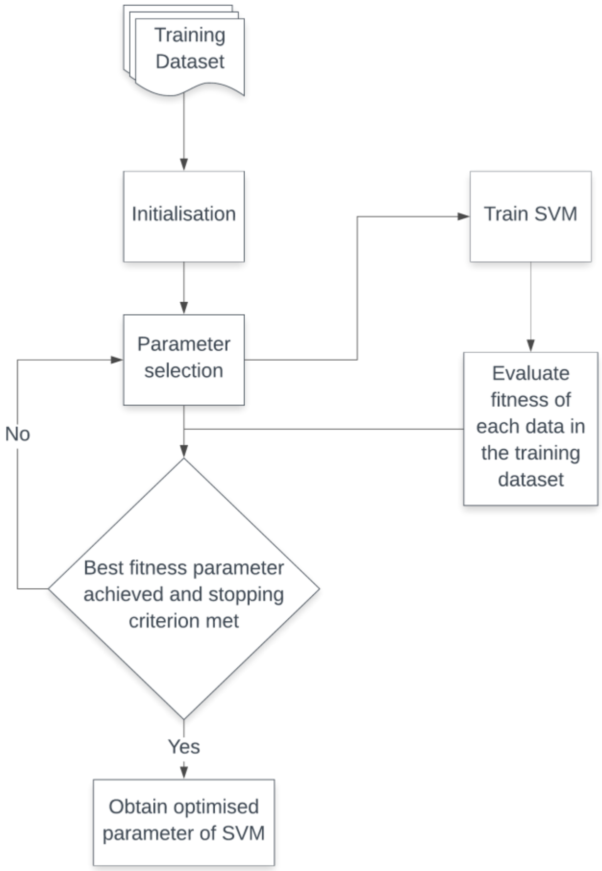
Training Dataset
Initialisation
Train SVM
Parameter selection
Evaluate fitness of each data in the training dataset
No
Best fitness parameter achieved and stopping criterion met
Yes
Obtain optimised parameter of SVM

Fig 4. Algorithm of Support Vector Machines

As shown in figure 4, we illustrate the various steps involved in an algorithm of SVM classification process. The performance and accuracy of the support vector machines are affected by kernel parameter and hyperplane selection.

The performance and accuracy do not get influenced by prior assumptions made on the unsupervised data. Support Vector Machines are used for non-parametric classification as they can manage large input data efficiently. They can efficiently achieve generalization. Overfitting is rectified. It is difficult to accurately define the optimal parameters required for support vector machines [19,20] owing to the high complexity of the algorithm.

Moreover, training takes time and, the output transparency remains low. In the next section, we highlight the results and discuss the feasibility of each technique.

Retrieval Number: B10350782S419/2019©BEIESP
DOI: 10.35940/ijrte.B1035.0782S419

Published By:
Blue Eyes Intelligence Engineering
Sciences Publication

IV. RESULTS AND DISCUSSIONS

The papers [19-23] present the various approaches which are applied to classify an image as useful or useless. In [24], the author uses low-level features and the OCR technique using SVM classifier to obtain 95 per cent accuracy. In [25], the author provides a literature review of OCR techniques and their limitations. In [26], the author highlights the potential of Artificial Neural Networks in recognising image spam. In [27], nearly 98 per cent accuracy is obtained using Backpropagation Neural Networks.

Next, we discuss the performance of different filter models built on distinct techniques. We tabulate the accuracy obtained for these systems and represent the performance of various sets of classifiers.

Table 1: Accuracy of Techniques

Reference Number	**Classifier**	**Performance**	**Dataset**
28	SVM	84%	Spam Archive and Ling Spam Images
29	SVM	96%	Spam Archive Princeton, dataset by Dredze et al, personal ham dataset
30	OCR	91%	Personal Dataset
31	OCR	95%	Personal Dataset
32	ANN	97.89%	Personal Corpus
33	ANN	94.30%	Personal Corpus

As shown in Table 1, we present a summarization of classifier-type, performance and, dataset Using [28-33] papers. We observe that the performance of ANN is better than others. We understand that improving the quality of datasets provides better accuracy.

V. CONCLUSION

In conclusion, an Artificial Neural Network helps to build a robust system for unsolicited visual data detection. This paper discusses three popular machine learning techniques - Optical Character Recognition, Support Vector Machines and Artificial Neural Network. We examine the use of these techniques in classifying unsolicited visual data and discover that the accuracy of ANN is the highest at nearly 98 per cent. We have observed that ANN is a non-parametric classifier depending on the training data for learning. Also, its accuracy varies depending upon the input and network structure. It is self-adaptive and capable of efficiently dealing with noisy inputs. Hence, this paper observes the benefits of ANN, as it yields high computing efficiency and presents many key characteristics.

REFERENCES

[1]. Attar, Abdolrahman & Moradi Rad, Reza & Ebrahimi Atani, Reza. (2011). A survey of image spamming and filtering techniques. Artificial Intelligence Review - AIR. 40. 1-35. 10.1007/s10462-011-9280-4.

[2]. Annadatha, A. & Stamp, M. J Comput Virol Hack Tech (2018) 14: 39. https://doi.org/10.1007/s11416-016-0287-x

[3]. Sneha Nikam, Rujata Chaudhari, "A review paper on image spam filtering",https://www.ijltet.org/journal_details.php?id=911&j_id=3690, Volume 8 Issue 2 - March 2017, 307-313, #ijltetorg

[4]. Khawandi, Shadi & Abdallah, Firas & Ismail, Anis. (2019). A Survey On Image Spam Detection Techniques. 13-27. 10.5121/csit.2019.90102.

[5]. Liu Q, Qin Z, Cheng H, Wan M (2010) Efficient modelling of spam images. In: 3rd International Symposium 3rd intelligent information technology and security informatics, IEEE, China

[6]. Technology AI, Rao NV, Pradesh A, Pradesh A, Pradesh A. OPTICAL CHARACTER RECOGNITION TECHNIQUE. 2016;83(2).

[7]. Nagabhushan, P. and Shivananda Nirmala. “Text Extraction in Complex Color Document Images for Enhanced Readability.” Intelligent Information Management 2 (2010): 120-133.

[8]. K. Kawattikul and P. Chomphuwiset, "A simple text detection in document images using classification-based techniques," 2017 IEEE 4th International Conference on Soft Computing & Machine Intelligence (ISCMI), Mauritius, 2017, pp. 119-122.doi: 10.1109/ISCMI.2017.8279610

[9]. E.M. Bahgat, S. Rady, W. GadAn E-mail filtering approach using classification techniques The 1st international conference on advanced intelligent system and informatics, Beni Suef, Egypt, Springer International Publishing (2016), pp. 321-331

[10]. Fadiora, B & Wada, F & Longe, Olumide. (2019). Combining Optical Character Recognition (OCR) and Edge Detection Techniques to Filter Image-Based Spam.

[11]. Yamakawa, Daisuke & Yoshiura, Noriaki. (2012). Applying Tesseract-OCR to detection of image spam mails. 1-4. 10.1109/APNOMS.2012.6356068.

[12]. Maind, S.B. & Wankar, P. (2014). Research paper on basic of Artificial Neural Network. International Journal on Recent and Innovation Trends in Computing and Communication. 2. 96-100.

[13]. Aiwan, F. & Zhaofeng, Y. Pers Ubiquit Comput (2018) 22: 1029. https://doi.org/10.1007/s00779-018-1168-8

[14]. Gogul, I & Kumar, Sathiesh. (2017). Flower species recognition system using convolution neural networks and transfer learning. 1-6. 10.1109/ICSCN.2017.8085675.

[15]. H. Drucker, Donghui Wu, and V. N. Vapnik. 1999. Support vector machines for spam categorization. Trans. Neur. Netw. 10, 5 (September1999),1048-105,DOI: https://doi.org/10.1109/72.788645

[16]. Malon C, Uchida S, Suzuki M. Mathematical symbol recognition with support vector machines. Pattern Recognit Lett [Internet]. 2008;29(9):1326–32.

[17]. H Bhavsar, MH Panchal. A Review on Support Vector Machine for Data Classification, 1 (10) (2012), pp. 185-189

[18]. Reena Sharma, Gurjot Kaur E-mail spam detection using SVM and RBF Int J Modern Education Comput Sci (IJMECS), 8 (4) (2016), p. 57

[19]. Long, Xianzhong & Lu, Hongtao & Peng, Yong & Wang, Xianzhong & Feng, Shaokun. (2015). Image classification based on improved VLAD. Multimedia Tools and Applications. 10.1007/s11042-015-2524-6.

[20]. Ketari & Mohammed; Chandra, Lamia & Khanum, Munesh & Akheela, Mohammadi. (2012). A Study of Image Spam Filtering Techniques. Proceedings - 4th International Conference on Computational Intelligence and Communication Networks, CICN 2012. 10.1109/CICN.2012.34.

[21]. Al-Duwairi, Basheer & Khater, Ismail & Al-Jarrah, Omar. (2013). Detecting Image Spam Using Image Texture Features. International

Journal for Information Security Research. 3. 10.20533/ijisr.2042.4639.2013.0040.

[22]. S. Dhanaraj and V. Karthikeyani, "A study on e-mail image spam filtering techniques," 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, Salem, 2013, pp. 49-55.doi: 10.1109/ICPRIME.2013.6496446

[23]. Amiza Amir, Bala Srinivasan, and Asad I. Khan. 2018. Distributed classification for image spam detection. Multimedia Tools Appl. 77, 11 (June 2018), 13249-13278. DOI: https://doi.org/10.1007/s11042-017-4944-y

[24]. Amit Kumar Sharma, Renuka Yadav Spam mails filtering using different classifiers with feature selection and reduction technique 2015 Fifth international conference on communication systems and network technologies (CSNT), IEEE (2015)

[25]. Mehta B, Nangia S, Gupta M, Nejdl W (2008) Detecting image spam using visual features and near-duplicate detection. security and privacy. ACM, Beijing

[26]. Bowling, J.R., Hope, P. and Liszka, K.J., “Spam image identification using an artificial neural network,” MIT SPAM Conference, pp.1-11, 2008.

[27]. Chowdhury M, Gao J, Chowdhury M (2015) Image spam classification using neural network. In: Thuraisingham B, Wang X, Yegneswaran V (eds) Security and privacy in communication networks: 11th international conference, SecureComm 2015, Dallas, TX, USA, October 26-29, 2015, Revised Selected Papers, Springer International Publishing, Cham, ISBN 978-3-319-28865-9, pp 622–632. doi:10.1007/978-3-319-28865-9_41

[28]. Khawandi, Shadi & Abdallah, Firas & Ismail, Anis. (2019). A Survey On Image Spam Detection Techniques. 13-27. 10.5121/csit.2019.90102.

[29]. K. Kawattikul and P. Chomphuwiset, "A simple text detection in document images using classification-based techniques," 2017 IEEE 4th International Conference on Soft Computing & Machine Intelligence (ISCMI), Mauritius, 2017, pp. 119-122.doi: 10.1109/ISCMI.2017.8279610

[30]. Wanli Ma, Dat Tran and Dharmendra Sharma, "Detecting Image-Based Spam Email," ICHIT Springer, pp. 168-177, 2007.

[31]. B. Fadiora, "Combining Optical Character Recognition (OCR) and Edge Detection Techniques to Filter Image-Based Spam", *African J. Comput. ICT January*, vol. 5, no. 1, pp. 59-68, 2012.

[32]. Chowdhury M., Gao J., Chowdhury M. (2015) Image Spam Classification Using Neural Network. In: Thuraisingham B., Wang X., Yegneswaran V. (eds) Security and Privacy in Communication Networks. SecureComm 2015. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 164. Springer.

[33]. Idris, Ismaila. (2011). E-mail Spam Classification With Artificial Neural Network and Negative Selection Algorithm. International Journal of Computer Science and Communication Networks. 01.

AUTHORS PROFILE

Abhishek Rungta has completed his B.Tech in Software Engineering from SRM Institute of Science and Technology, Chennai, India in 2019. His research interests include Machine Learning, Computer Vision and Interaction Design.

Bhawna Arya has completed her B.Tech in Software Engineering from SRM Institute of Science and Technology, Chennai, India in 2019. Her research interests include Computer Vision, Software Development Models and Critical Systems Group (Crysis).

Dr.Usha is currently working as an associate professor at the software engineering department in SRMIST. She has 12 years of teaching experience.While working in Anna University chennai she worked in research projects for Smart and Secure techniques Research Lab. Her research interests include network security, machine learning, Bioinformatics. Dr. Usha published nearly 40 research articles in peer reviewed journals and international conferences. She is GATE scorer and awarded as college first rank holder in UG. She is an editorial board member for the journal Progress of Electrical and Electronic Engineering. She was awarded as Outstanding Reviewer within Top 10 percentile of reviewers in Elsevier-Pattern Recognition Letters in 2017. She is reviewer of Elsevier Journal - Computer and Electrical Engineering, Elsevier Journal- Pattern Recognition Letters, Springer- Multimedia tools and Applications, IEEE Access. She has coordinated IET sponsored Workshop on Cyber Security, National Workshop on Internet of Things, National workshop on VANET and its security IET sponsored National Conference on Big data, cloud and Security. She is an active member of IET, ISTE, Indian Science Congress. Currently she is guiding 6 Phd Students.