ISSN: 2278-4632
Vol-14, Issue-4, April: 2024

LUNG CANCER DETECTION USING DEEP LEARNING

B.Swarajya Lakshmi1[1], G.Bramarambika[2], N.MeghanaReddy[3], U.GuruKavitha[4], T.SudheshnaReddy[5]

[1]Assistant Professor in Department of Computer Science and Engineering, Santhiram Engineering College, Nandyal, Kurnool, Andhra Pradesh,India.

[2,3,4,5] Student, Department of Computer Science and Engineering, Santhiram Engineering College, Nandyal, Kurnool, Andhra Pradesh, India.

email: swarajyalakshmi.cse@srecnandyal.edu.in , 20x51a0595@srecnandyal.edu.in

Abstract: Early detection of lung cancer is crucial for effective treatment. Traditional methods like X-ray and CT scans are widely used but not always accessible globally. This study proposes a comprehensive approach for lung cancer detection using medical image mining techniques. The process involves lung field segmentation, feature extraction, and classification utilizing neural networks and support vector machines (SVM). Two datasets were utilized, and various experiments were conducted, including feature selection and training Capsule Networks (CapsNet) with different parameters. Results indicate promising accuracy rates: Convolutional Neural Network (CNN) achieved 21%, K-Nearest Neighbors (KNN) achieved 43%, SVM achieved 90%, and Random Forest achieved 60%. Notably, CapsNet outperformed all other methods with an accuracy of 96%. This suggests the potential of CapsNet in enhancing lung cancer detection. The findings underscore the importance of advancing automated techniques for early-stage lung cancer diagnosis, especially in regions with limited access to advanced screening technologies.

Keywords : *Lung Cancer Detection, CNN, SVM, KNN, Random Forest, CapsNet, Deep Learning, Machine Learning.*

1. INTRODUCTION

Lung cancer stands as one of the most formidable adversaries in the realm of oncology, characterized by unbridled cellular proliferation within the lung tissues. Early detection is paramount to curative interventions, as treatment efficacy dramatically declines in advanced stages. While modalities such as X-ray, CT scans, and MRI are pivotal in diagnosis, their widespread availability remains elusive in many regions, underscoring the significance of simpler yet effective screening methods like chest radiography.

The primary objective of this paper is to automate the classification process for early-stage lung cancer prediction, thereby facilitating timely interventions. Leveraging advanced technologies like Neural Networks and Machine Learning, this research endeavors to streamline the classification process, enabling accurate and swift identification of abnormal lung patterns indicative of carcinoma. The evaluation of proposed methodologies hinges on the accurate classification of sample data, serving as a litmus test for the efficacy of the developed algorithms.

Central to this endeavor are fundamental techniques in medical image mining, encompassing lung field segmentation, image processing, feature extraction, and classification. Through the meticulous application of these techniques, digital X-ray chest films are categorized into two distinct classes: normal and abnormal. The classification process is underpinned by the deployment of sophisticated algorithms, including Neural Networks and Support Vector Machines (SVMs), which are trained and fine-tuned using diverse datasets.

The experimental framework of this study encompasses a multifaceted approach. Two distinct datasets are curated, each subjected to various learning experiments. Feature selection techniques are employed to optimize model performance, while SVMs are trained with a spectrum of parameters to ascertain the most effective configurations. The ensuing results are meticulously scrutinized, compared, and reported, shedding light on the relative efficacy of different methodologies in automating lung cancer classification.

2. LITERATURE SURVEY

[1] Zakaria Suliman Zubi and Rema Asheibani Saad's research focused on utilizing data mining techniques for the early diagnosis of lung cancer. They explored the application of various algorithms to aid in the timely detection of this disease, aiming to improve patient outcomes through early intervention.

[2] In a study by Paola Campadelli, Elena Casiraghi, and Diana Artioli, a fully automated method for lung nodule detection from postero-anterior chest radiographs was proposed. Their research aimed to develop a robust algorithm capable of accurately identifying nodules in chest X-rays, facilitating early detection and treatment planning.

[3] Rohit Chitale et al. investigated the utilization of Convolutional Neural Networks (CNN) and ensemble methods to support medical practitioners in categorizing lung cancer. Their study delved into the potential of advanced machine learning techniques to assist healthcare professionals in accurately classifying lung cancer cases.

[4] V. Krishnaiah, Dr. G. Narsimha, and Dr. N. Subhash Chandra's work focused on the development of a lung cancer prediction system using data mining classification techniques. Their research aimed to create a predictive model capable of identifying patterns indicative of lung cancer, aiding in early diagnosis and treatment.

[5] Astha Pathak et al. conducted a survey-based study on machine learning algorithms for lung cancer prediction. Their research aimed to provide insights into the various machine learning techniques employed in predicting lung cancer, highlighting their strengths and limitations in clinical applications.

[6] Kambam Shreya et al. proposed a machine learning approach for lung cancer analysis. Their study aimed to develop a robust model capable of analyzing data related to lung cancer and providing valuable insights for clinical decision-making.

[7] K R Lathakumari, A C Ramachandra, U C Avanthi, C Basil Ronald, and T Bhavatharani explored the classification of non-small cell lung cancer using deep learning techniques. Their research aimed to leverage the power of deep learning algorithms to accurately classify different subtypes of lung cancer, thereby aiding in personalized treatment strategies.

[8] Asha V and Bhavanishankar K conducted a review on lung cancer detection using CT scans and image processing through deep learning techniques. Their study aimed to provide an overview of the current state-of-the-art methods in lung cancer detection, focusing on the role of deep learning algorithms in analyzing CT scan images.

[9] P. Chitra et al. investigated lung cancer detection using classification algorithms. Their research aimed to develop and evaluate various classification algorithms for accurately detecting lung cancer, providing insights into

the performance of different machine learning techniques in this domain.

[10] Nitha V. R and Vinod Chandra S. S. explored lung cancer malignancy detection using a voting ensemble classifier. Their research aimed to develop an ensemble learning approach capable of combining multiple classifiers to improve the accuracy of lung cancer malignancy detection.

These studies collectively highlight the growing interest in leveraging advanced technologies, such as machine learning and deep learning, for the early detection and diagnosis of lung cancer. By harnessing the power of data-driven approaches, researchers aim to enhance clinical decision-making and improve patient outcomes in the battle against this deadly disease.

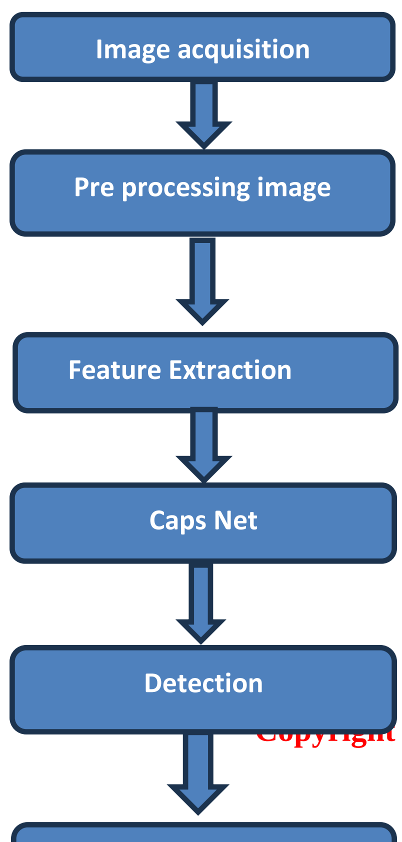
3. METHODOLOGY

i) Proposed Work:

The proposed system employs medical image mining techniques for comprehensive lung cancer detection, focusing on early-stage diagnosis. Initially, the system conducts lung field segmentation, followed by image processing and feature extraction. Utilizing neural networks and support vector machines (SVM), the extracted features are then classified into two categories: normal and abnormal. Training of the system involves diverse datasets, including computed tomography scan images with abnormal physiology, enhancing its ability to identify subtle abnormalities indicative of lung cancer. Additionally, the system incorporates Capsule Networks (CapsNet) for improved classification accuracy. Experimental evaluations demonstrate promising results, with CapsNet exhibiting a remarkable accuracy of 96%, surpassing traditional methods such as CNN, KNN, SVM, and Random Forest. Overall, the proposed system showcases the potential of advanced machine learning techniques in enhancing the early detection of lung cancer, particularly in regions with limited access to advanced screening technologies.

ii) System Architecture:

The first image is transformed to gray scale image. After that, removal of the noises and contrast enhancement is finished for obtaining the improved images. After image acquisition the system perform preprocessing on image understand affected regions and their characteristics in style of data. This data is classed using RF,KNN,SVM,CNN & CapsNet. CapsNet classify it as normal or diseases lung and identify lung diseases.



ISSN: 2278-4632
Vol-14, Issue-4, April: 2024

Fig3. 1 Proposed Architecture

iii) Dataset Collection:

The dataset utilized in this study comprises a collection of digital X-ray and computed tomography (CT) scan images of the chest region. These images encompass a diverse range of cases, including both normal and abnormal physiological conditions associated with lung cancer. The dataset is curated to represent various stages and manifestations of the disease, facilitating robust training and evaluation of the proposed system. Additionally, specialized subsets within the dataset focus on specific characteristics relevant to lung cancer diagnosis, ensuring comprehensive coverage of relevant image features for accurate classification.

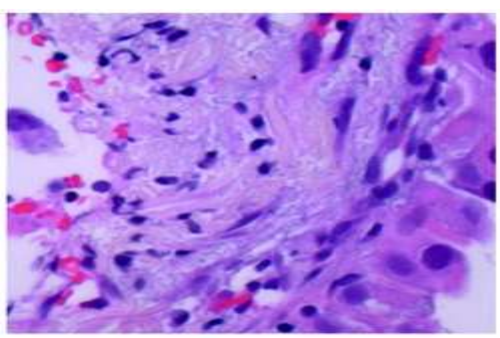


Fig 3.2 Sample Dataset Image

iv) Image Processing:

Image processing plays a crucial role in preparing and enhancing images for various applications, including detection tasks. One approach involves using the ImageDataGenerator module, which provides several augmentation techniques to enrich the dataset. These techniques include re-scaling the image to normalize pixel values, shear transformation to introduce controlled distortions, zooming to simulate varying perspectives, and horizontal flipping to increase dataset diversity. Additionally, reshaping the image ensures uniform dimensions for compatibility with neural network architectures. Another method utilizes Torchvision, offering pre-defined functions tailored for object detection tasks. This framework enables operations such as resizing, normalization, and transformation of images to optimize their suitability for detection algorithms. By incorporating these image processing techniques, the quality and diversity of the dataset are enhanced, leading to improved model performance and robustness in detection tasks.

v) Training & Testing:

The dataset is partitioned into training and testing sets using an 80:20 ratio, respectively. This division ensures that 80% of the data is allocated for training the model, allowing it to learn patterns and features from a diverse range of examples. The remaining 20% of the dataset is reserved for testing the trained model, enabling evaluation of its performance on unseen data. This split ensures that the model's generalization ability is effectively assessed, providing insights into its effectiveness in accurately classifying lung cancer images while mitigating the risk of overfitting to the training data.

vi) Algorithms:

CNN (Convolutional Neural Network): A deep learning algorithm specifically designed for processing and classifying visual data, such as images. CNNs consist of multiple layers, including convolutional layers for feature extraction and pooling layers for dimensionality reduction, followed by fully connected layers for classification.

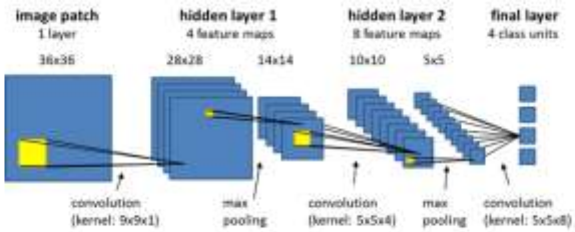


Fig 3.5 CNN

CapsNet (Capsule Network): A type of neural network architecture introduced to overcome the limitations of traditional CNNs in handling hierarchical relationships and spatial hierarchies within data. CapsNet incorporates capsules, which are groups of neurons that encode the instantiation parameters of specific entities in the input data, enabling better generalization and robustness.

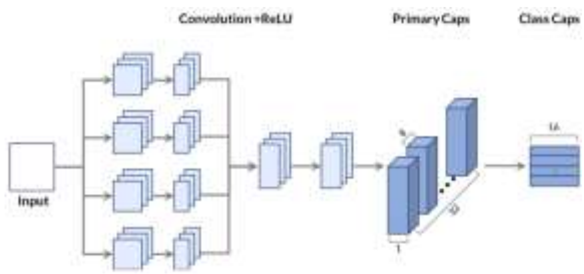


Fig 3.3 CapsNet

KNN (K-Nearest Neighbors): A simple yet effective supervised learning algorithm used for classification and regression tasks. KNN works by assigning a class label to an input sample based on the majority class label among its k nearest neighbors in the feature space, where distance metrics such as Euclidean distance are commonly used to measure similarity.

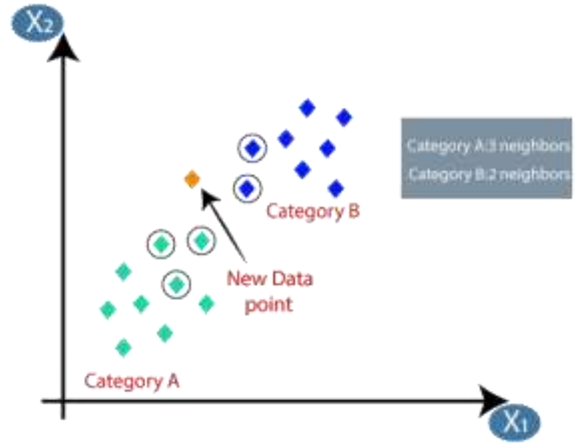


Fig 3.6 KNN

SVM (Support Vector Machine): A powerful supervised learning algorithm used for classification and regression tasks. SVM works by finding the hyperplane that best separates different classes in the feature space, maximizing the margin between classes while minimizing classification errors. SVM can handle linear and non-linear data separation using different kernel functions.

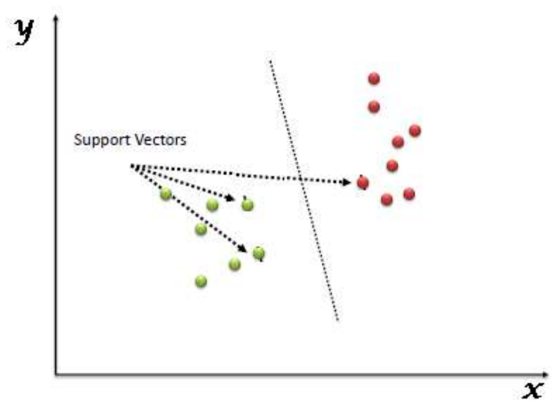


Fig 3.7SVM

Random Forest: An ensemble learning algorithm that operates by constructing a multitude of decision trees during training and outputting the mode of the classes (classification) or the mean prediction (regression) of the individual trees. Random Forest mitigates overfitting and

ISSN: 2278-4632
Vol-14, Issue-4, April: 2024

improves robustness by combining predictions from multiple decision trees trained on different subsets of the data.

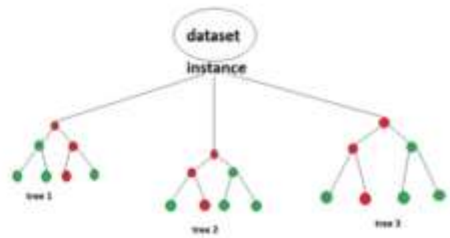


Fig 3.8 Random Forest

4. EXPERIMENTAL RESULTS

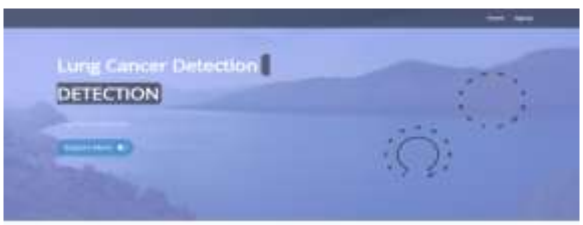


Fig 4.1 Home Page

Fig 4.2Registration Page

Fig 4.3Login Page

Fig 4.4Uploaded Input Image

Fig 4.5Predict Result for Given Input Image as Patient has Lung Aca

5. CONCLUSION

Eventually, this project involved different image processing procedures of lung modules that were intended for the most appropriate lung cancer detection. Among the variety of image processing methods, fuzzy filter, undoubtedly, showed the best performance in terms of noise suppression and contrast improvement. Segmentation by watershed marker-based

ISSN: 2278-4632
Vol-14, Issue-4, April: 2024

algorithm provided the means for water areas division, thus facilitating feature extraction. Through this method, not only accuracy and reliability were enhanced, but also faster result generation was made possible. The final outcome of these efforts was the creation of the CapsNet classifier which turned out to be very accurate in differentiating lung nodules between benign or malignant ones. The CapsNet classifier exhibit an impressive accuracy rate of 96%, which point to its high effectiveness in accurately detecting cancerous tumors in lung images. Such high accuracy, in fact, opens new opportunities to improve early-stage lung cancer diagnosis and treatment planning. The results of this project demonstrate the significance of using the most advanced image processing and machine learning systems in the medical imaging field for improving disease detection and enhancing patient care.

6. FUTURE SCOPE

Future endeavors in this domain could explore integrating additional advanced image processing techniques and machine learning algorithms to further enhance the accuracy and efficiency of lung cancer detection. Research could focus on optimizing computational efficiency and scalability to enable real-time application in clinical settings. Additionally, the development of robust automated systems capable of processing large-scale datasets could pave the way for personalized medicine approaches and improved patient outcomes in the field of oncology.

REFERENCES

[1] ZakariaSulimanZubi and RemaAsheibaniSaad, “Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer,” Recent Researches in Artificial Intelligence, Knowledge Engineering and Data Bases, Libya, 2007.

[2] Paola Campadelli, Elena Casiraghi, and Diana Artioli, “A Fully Automated Method for Lung Nodule Detection From Postero-Anterior Chest Radiographs,” In Proc. of IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 25, NO. 12, DECEMBER 2006.

[3] V.Krishnaiah, Dr.G.Narsimha, Dr.N.Subhash Chandra. 2013, “Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques,” International Journal of Computer Science and Information Technologies, Vol. 4 (1), 2013, 39 – 45.

[4] Astha Pathak, Sunil Kumar Dewangan, Mahendra Kumar Sahu, M Gayatri, Gitanjali Sahu, Prakriti Verma, "A Survey Based on Machine Learning Algorithm for Lungs Cancer Prediction", 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), vol.1, pp.1-6, 2023.

[5] Nitha V. R, Vinod Chandra S. S., "Lung Cancer Malignancy detection Using Voting Ensemble Classifier", 2023 2nd International Conference on Computational Systems and Communication (ICCSC), pp.1-5, 2023.

[6] Mahammad, F. S., & Viswanatham, V. M. (2020). Performance analysis of data compression algorithms for heterogeneous architecture through parallel approach. The Journal of Supercomputing, 76(4), 2275-2288.

[7] Karukula, N. R., & Farooq, S. M. (2013). A route map for detecting Sybil attacks in urban vehicular networks. Journal of Information, Knowledge, and Research in Computer Engineering, 2(2), 540-544.

[8] Farook, S. M., & NageswaraReddy, K. (2015). Implementation of Intrusion Detection Systems for High Performance Computing Environment Applications. Inter national journal of Scientific Engineering and Technology Research, 4(0), 41.

[8] Sunar, M. F., & Viswanatham, V. M. (2018). A fast approach to encrypt and decrypt of video streams for secure channel transmission. *World Review of Science, Technology and Sustainable Development*, *14*(1), 11-28.

[9] Mahammad, F. S., & Viswanatham, V. M. (2017). A study on h. 26x family of video streaming compression techniques. *International Journal of Pure and Applied Mathematics*, *117*(10), 63-66.

[10] Devi,S M. S., Mahammad, F. S., Bhavana, D., Sukanya, D., Thanusha, T. S., Chandrakala, M., & [11] [17] Swathi, P. V. (2022).” Machine Learning Based Classification and Clustering Analysis of Efficiency of Exercise Against Covid-19 Infection.” Journal of Algebraic Statistics, 13(3), 112-117.

[11] Devi, M. M. S., & Gangadhar, M. Y. (2012)." A comparative Study of Classification Algorithm for Printed Telugu Character Recognition." *International Journal of Electronics Communication and Computer Engineering*, *3*(3), 633-641.

[12] Devi, M. S., Meghana, A. I., Susmitha, M., Mounika, G., Vineela, G., & Padmavathi, M. MISSING CHILD IDENTIFICATION SYSTEM USING DEEP LEARNING.

[13] V. Lakshmi chaitanya. "Machine Learning Based Predictive Model for Data Fusion Based Intruder Alert System." journal of algebraic statistics 13, no. 2 (2022): 2477-2483.

[14] Chaitanya, V. L., & Bhaskar, G. V. (2014). Apriori vs Genetic algorithms for Identifying Frequent Item Sets. International journal of Innovative Research &Development, 3(6), 249-254.

[15] Chaitanya, V. L., Sutraye, N., Praveeena, A. S., Niharika, U. N., Ulfath, P., & Rani, D. P. (2023). Experimental Investigation of Machine Learning Techniques for Predicting Software Quality.

[16] Lakshmi, B. S., Pranavi, S., Jayalakshmi, C., Gayatri, K., Sireesha, M., & Akhila, A. Detecting Android Malware with an Enhanced Genetic Algorithm for Feature Selection and Machine Learning.

[17] Lakshmi, B. S., & Kumar, A. S. (2018). Identity-Based Proxy-Oriented Data Uploading and Remote [18] Data Integrity checking in Public Cloud. International Journal of Research, 5(22), 744-757.

[19] Lakshmi, B. S. (2021). Fire detection using Image processing. Asian Journal of Computer Science and Technology, 10(2), 14-19.

[20] Devi, M. S., Poojitha, M., Sucharitha, R., Keerthi, K., Manideepika, P., & Vasudha, C. Extracting and Analyzing Features in Natural Language Processing for Deep Learning with English Language.

[21]Kumar JDS, Subramanyam MV, Kumar APS. Hybrid Chameleon Search and Remora Optimization Algorithm-based Dynamic Heterogeneous load balancing clustering protocol for extending the lifetime of wireless sensor networks. Int J Commun Syst. 2023; 36(17):e5609. doi:10.1002/dac.5609

[22] David Sukeerthi Kumar, J., Subramanyam, M.V., Siva Kumar, A.P. (2023). A Hybrid Spotted Hyena and Whale Optimization Algorithm-Based Load-Balanced Clustering Technique in WSNs. In: Mahapatra, R.P., Peddoju, S.K., Roy, S., Parwekar, P. (eds) Proceedings of International Conference on Recent Trends in Computing. Lecture Notes in Networks and Systems, vol 600. Springer, Singapore. https://doi.org/10.1007/978-981-19-8825-7_68

[23] Murali Kanthi, J. David Sukeerthi Kumar, K. Venkateshwara Rao, Mohmad Ahmed Ali, Sudha Pavani K, Nuthanakanti Bhaskar, T. Hitendra Sarma, "A FUSED 3D-2D CONVOLUTION NEURAL NETWORK FOR SPATIAL-SPECTRAL FEATURE LEARNING AND HYPERSPECTRAL IMAGE CLASSIFICATION," J Theor Appl Inf Technol, vol. 15, no. 5, 2024, Accessed: Apr. 03, 2024. [Online]. Available: www.jatit.org

[24] Prediction Of Covid-19 Infection Based on Lifestyle Habits Employing Random Forest Algorithm FS Mahammad, P Bhaskar, A Prudvi, NY Reddy, PJ Reddy journal of algebraic statistics 13 (3), 40-45

[25] Machine Learning Based Predictive Model for Closed Loop Air Filtering System P Bhaskar, FS Mahammad, AH Kumar, DR Kumar, SMA Khadar, ...Journal of Algebraic Statistics 13 (3), 609-616

[26] Kumar, M. A., Mahammad, F. S., Dhanush, M. N., Rahul, D. P., Sreedhara, K. L., Rabi, B. A., & Reddy, A. K. (2022). Traffic Length Data Based Signal Timing Calculation for Road Traffic Signals Employing Proportionality Machine Learning. Journal of Algebraic Statistics, 13(3), 25-32.

[27] Kumar, M. A., Pullama, K. B., & Reddy, B. S. V. M. (2013). Energy Efficient Routing In Wireless Sensor Networks. International Journal of Emerging Technology and Advanced Engineering, 9(9), 172-176.

[28] Kumar, M. M. A., Sivaraman, G., Charan Sai, P., Dinesh, T., Vivekananda, S. S., Rakesh, G., & Peer, S. D. BUILDING SEARCH ENGINE USING MACHINE LEARNING TECHNIQUES.

[29] "Providing Security in IOT using Watermarking and Partial Encryption. ISSN No:

2250-1797 Issue 1, Volume 2 (December 2011)

[30] The Dissemination Architecture of Streaming Media Information on Integrated CDN and P2P, ISSN 2249-6149 Issue 2, Vol.2 (March-2012)

[31] Provably Secure and Blind sort of Biometric Authentication Protocol using Kerberos, ISSN: 2249-9954, Issue 2, Vol 2 (APRIL 2012)

[32]R sumalatha, dr.m.subramanyam, “image denoising using spatial adaptive mask filter”, ieee international conference on electrical, electronics, signals, communication & optimization (eesco-2015), organized byvignans institute of information technology, vishakapatnam, 24 th to 26th january 2015. (scopus indexed)

[33] P.balamurali krishna, dr.m.v.subramanyam, dr.k.satya prasad, “hybrid genetic optimization to mitigate starvation in wireless mesh networks”, indian journal of science and technology,vol.8,no.23,2015. (scopus indexed)

[34] Y.murali mohan babu, dr.m.v.subramanyam,m.n. giri prasad,” fusion and texure based classification of indian microwave data – a comparative study”, international journal of applied engineering research, vol.10 no.1, pp. 1003-1009, 2015. (scopus indexed)