iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35528

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Disease Classification with E-Report Generation, Authentication and Encryption

Ashish Kurane[1], Ameena Firdose[2], Janardhan U[3], Lohitha H. D.[4], Vinod Kumar H[5]
[1, 2, 3, 4]*Students of Telecommunication Department, Dayananda Sagar College of Engineering, Bangalore, India*
[5]*Assistant professor of Telecommunication Department, Dayananda Sagar College of Engineering, Bangalore, India*

Abstract: ***The assortment of data analysis on the origin of diseases and consequences of mortality is essential to keep track of death rates caused due to diseases. Thus, the classification of diseases is very crucial. Cancer is one of the huge and major diseases of concern in the world. Machine learning is extensively implemented in the medical field in the anticipation of medical errors and early revelation of diseases. Along with the implementation of technology in medical field there is need for authentication to safeguard the privacy rights of patient's health information. Thus, in this paper, revelation of disease using CNN (Convolutional Neural Networks) algorithm is achieved along with authentication and automatic generation of e-medical report which is further encrypted using RSA (Rivest, Shamir, Adleman) algorithm to overcome the breach of information while being shared from one hospital to another.***

Keywords: Convolutional Neural Network (CNN), Python, Graphical User Interface (GUI), RSA (Rivest, Shamir, Adleman), Magnetic Resonance Imaging (MRI), E-report, Authentication, Encryption

I. INTRODUCTION

Cancer is one of the most death-dealing and wide-reaching health issues in the world. Cancer is a disorderly and unusual growth of cells in the region of the body. Surveys performed estimate about 1.76 million people are diagnosed with cancer in the USA alone in 2019. The deaths caused due to cancer are most common for the people above 50 years old. The origin of cancer in the specific region of the body gives rise to the name of cancer. The varieties of cancers are over 200 in number. The most common being diagnosed as lung cancer, stomach cancer, brain cancer, liver cancer and, so on. After heart attack, the second largest death threat to the world is by lung cancer. In comparison with other types of cancers, the largest number of deaths are caused by lung cancer. Lung cancer is the abnormal growth of cells in our spongy like organs in our chest i.e., lungs. According to the reports, it is observed that globally 19% of the deaths are major because of alcohol and tobacco consumption. The survival Period of the patient is approximately 5 years. the main reason for the high death rates due to lung cancer is the procrastinated detection in successive stages which in turn leads in hampering the required treatment. The possibility of survival can be hiked by 50 - 70% if the revelation is done in rudimentary stages. Lung cancer is basically classified into non-small cell and small cell lung cancer. Among the two, the non-small cell lung cancer makes about 85 -90% of the total lung cancers and small cell lung cancer contributes to the rest of 10 - 15% of the cases. The lung cancer is divided into four stages depending upon in its expanse and severity. Stages-1 occurs in lungs, stage-2 to and stage 3 are found in the chest, while stage-4 occurs when the lung cancer has affected other organs and infected them too. After lung cancer, stomach cancer falls under the list of top five common cancers occurring globally. Every year the number of people prone to stomach cancer is increasing gradually. Recently according to the study, the counts on the stomach cancer was estimated to be 30000 per year. Also, on a global scale,26% of the males and 11% of females are affected with stomach cancer. The chief reason for this is said to be the consumption of smoking and salted food and smoking. The stomach cancer commonly occurs in the mucosa (the innermost lining of the stomach). To stop the rate of increase in this type of cancer early detection becomes a necessary step. Moving towards Brain Tumor, which stands as 10th prime cause of death globally and whose survival rate is 5 years, is said to be 36% for women and 34% for men. It is stated by the World Health Organization (WHO) that 400,00 people are suffering from brain tumor worldwide. Benign (non-cancerous) and malignant (cancerous) are the two classifications under brain tumor. The latter spreads faster to the other tissues and worsens the patient's condition over time. As a part of the cell cycle, the older or the damaged cells have to be annihilated and replaced by new cells, but sometimes the older or the damaged cells do not get annihilated which results in tumor by the formation of additional cells. The size, shape, location and type of tumor make it difficult for the detection of a brain tumor in rudimentary stages. However, the chances of patient's survival can be made high by detecting and diagnosing the tumor early.

©IJRASET: All Rights are Reserved

For the diagnosing of above-mentioned diseases, the most suitable method is Magnetic Resonance Imaging (MRI) in the field of Medical Detection System (MDS). In our proposed idea, we implement the MRI images as input for CNN classifier which detects identifies and classifies the respective disease. Having known the type of disease, the information is sent to the respective specialist and an E-report is generated. The report includes patients' personal details health details and the precautions. The report generated is saved in an encrypted format and can be accessed only by the doctor in order to ensure privacy for the patients details and avoid the misuse of patient's personal details while being communicated from one hospital to another. Also, authentication is provided by the doctor for double verification. Consequently, detection of diseases is achieved with report generation along with authentication and encryption.

II. LITERATURE REVIEW

Fascination of machine learning and curiosity of automatic computer virus diseases analysis has made advancements in Medical Science due to which highly efficient algos (algorithm) were developed in past few years. These algorithms help to recognize new and Peculiar disease traits that are crucial in analysis and understanding of the tests conducted. A particularly convolutional neural network is seen to be one of the best algorithms in identifying Peculiar characteristics of diseases and present them helping early perception of diseases. For instance,[1] the authors here speak on a proposal to identify CTC circulating tumor cells using convolutional neural network on images of unstained blood. [2] here author gives a study on Deep learning implemented on analyzing medical images.[3] Brain Tumor is discussed by authors mentioning how dangerous and critical disease it is and which needs more accuracy based on early perception. Most detection and analysis method depend on the decision of doctors and specialists. [4] as Seen here described by the author we understand that half of the population who suffer from stomach cancer are late in spotting of the disease. [5] author share gave insights on understanding one of the diagnosis methods by using neural networks such as machine learning algorithm support vector machine (SVM).[6] here we get brief about the phases of lung cancer and degree of its spread in the lungs. [7] here the authors discuss on CAD-based Framework image of lungs are divided by using region growing technique and smoothing morphologically. [8] and here author presents efficient RSA algorithm key generation exponents are increased then the original is done.

III. METHODOLOGY

In deep learning, CNN is a type of model used for studying and processing that data which exists in a matrix format, such as images. CNN is developed based on operation of the animal visual cortex. It is made up of three fundamental blocks (layers): convolution layer, pooling layer and, fully connected layer. The first two blocks i.e., convolution and pooling layer extract the features of image, while the third Block i.e., fully connected layer performs the classification of the input image. The convolution layer is comprised of linear operation called 'convolution' which plays a vital role in CNN. Kernels make the CNN highly efficient in feature extraction. The kernel is a small matrix comprising of parameters. In CNN as the output of one layer enters the next layer, the intricacy of the model is peaked. In order to abate the differences between the output obtained and actual ground truth table, training is introduced. The following is the overall architecture of our proposed model and detailed architecture of CNN:

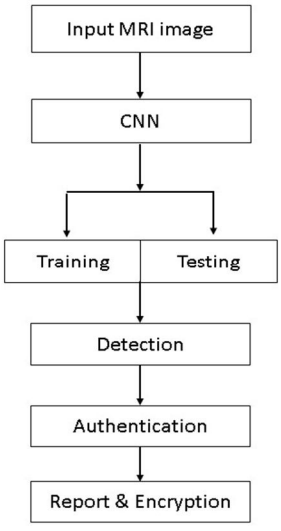


Figure1. overall architecture of proposed model

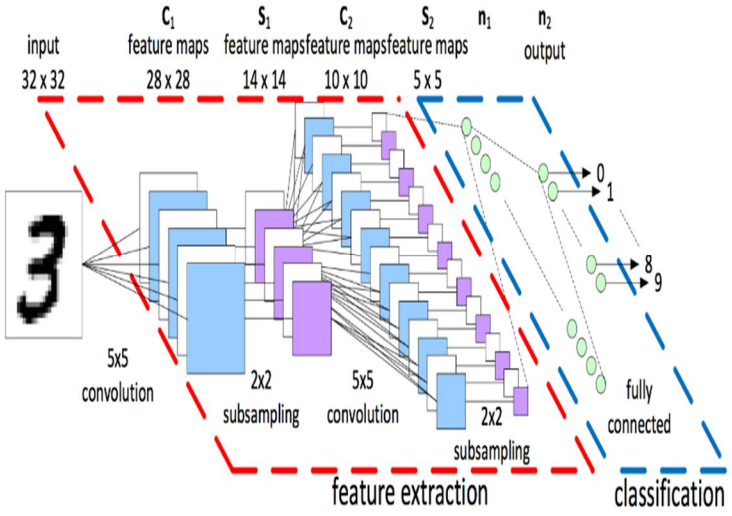


Figure 2. Architecture of CNN

©IJRASET: All Rights are Reserved

A. Input Image

The first layer of CNN is where the input MRI image is fed to the model. The size and dimensions of input image can vary from image to image. Thus, the dimensions and size of input image are the deciding factors of the input data number. The input data number is given by:

Input data number = X*Y*Z

Where X, Y, Z are the dimensions of the image, for instance input data number is 268203 if the dimensions are 299*299*3.

B. Convolution Layers

The next layer is 2D convolutional layer of CNN. It comprises of the 'convolution' function- a type of linear Operation. The output of convolution layer is termed as feature map or activation map, which is obtained by convoluting the input image with filter matrix. the feature map is given by

$$(I * F)(x, y) = \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} I(u, v) * F(-u + x, -v + y)$$

In the above equation I(x, y) is the image, F(x, y) is the filter, inversion and F(-u+x, -v+y) is the shifting of filter performed. In the output matrix, the number of convolution layers is equivalent to the number of filters taken for convolution.

C. Kernels and Shape

The kernel is a matrix which includes parameters like edge, shape, curve, etc. It is also called a filter. The feature map depends on the filter i.e., different feature maps are produced based on different filters implemented. Shooting -up the number of filters will result in a huge number of feature maps. The depth of the kernel relies on the depth of matrix which is convoluted with the kernel. The shape of the kernel can be either square or rectangle.

D. Padding

When the filter is applied to the image, the main drawback can be losing information on the corners of the image. To overcome this, padding is implemented. Thus, padding is the process of adding zeros to the borders of an image. By doing this, the edges of the image are covered and more pixels are achieved. This increases the overall efficiency of CNN. The padding is given by:

P=((S-1)*W-S+F)/2

Where F -filter size, W- input size, S- stride (number of steps the filter moves after performing convolution).

E. Rectification Linear Unit (ReLU)

There is no significance for the convoluted output if it is a negative value. Therefore, ReLU layer modifies all existing negative values to zeros by the implementation of sigmoid and tanh function. The output of this layer is non-linear output. ReLU is a type of activation function given by:

Y= max(0,x)

Where x is the value of the point (Negative value)

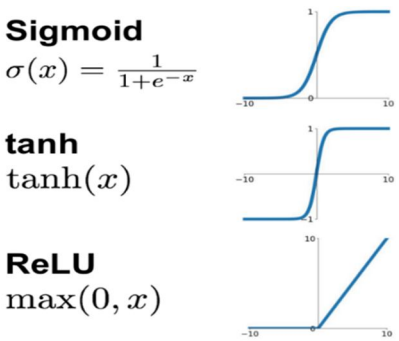


Figure 3. Activation functions

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

F. Pooling Layer

It is the next building block after convolution layer. It is used to shrink the spatial dimensions of the feature map independently. By doing so, it lessens the computational complexity in a network, as the number of parameters gets reduced. The pooling can be performed as: max pooling, average pooling or global pooling. In our model, max pooling is devised. Max pooling selects the most prominent features of the previous feature maps and gives the output. The pooling layer lays the foundation of robustness to CNN. Max pooling is given by:

$$a_j = \max_{(p,q)\in R_{ij}} a_{kpq}$$

G. Fully Connected Layers

It is in this layer; the image finally gets classified into respective diseases. The input fed to this layer is the image which will be converted into vector. These vectors are fused to produce a model. Then finally an activation function classifies the image. After this, the CNN model is trained with weights and parameters using back propagation.

IV. AUTHENTICATION AND REPORT GENERATION

In order to ensure authenticity of the user, as well as Doctor, GUI is created using Python, which can be accessed and processed only by the doctor. The interactive GUI provides login authentication by the doctor. The doctor has the user- ID and password and only he can have access to the report. All the credentials entered by the doctor are verified too. This ensures double verification by the doctor. Furthermore, a medical report is generated automatically using the concept of file handling in python. Python file handling plays an essential role while creating a file or reading data from a file. File handling enables vast number of operations to be performed on a file. File handling in other programming languages is very sophisticated. Thus, python is used.

V. ENCRYPTION

The Automatically generated report consists of the patient's personal information and the disease detected. Encryption becomes a necessity to secure the patient's health information. Failing to ensure security not only poses technical issues but also ethical issues like denial of health insurance or misuse of health insurance by an intruder. Thus, there is a need for encryption. In our proposed model, RSA encryption method is utilized. RSA stands for Rivest, Shamir, Adleman based on the names of its developers - Ron Rivest, Adi Shamir and Leonard Adleman, who developed RSA algorithm in 1977. It falls under public-key encryption (asymmetric encryption). Due to distinguishable functions of mathematics like one- way trapdoor function, Euler's theorem and modular arithmetics, being implemented in RSA, it generates two keys public-key and private- key. The working of RSA algorithm includes four modules: key generation, key distribution, encryption and decryption. The following is an algorithm for RSA:

Step 1: Select two distinctive prime numbers, represented as p and q.

Step 2: Determine n=pq, where 'n' is expressed in bits, represents key length for both keys.

Step 3: Determine Carmichael totient function for the product (pq). $\lambda(n)= LCM\{(p-1),(q-1)\}$

Step 4: Select a number between $1<e<\lambda(n)$ such that it is coprime to $\lambda(n)$.

Step 5: Determine d, the modular multiplicative inverse, given by: $d=e(\text{mod } \lambda(n))$

Step 6: For public- key(n,e) with a plaintext message m, the encryption function is given by: $c(m)=m^e \text{ mod } n$

Step 7: With private-key(n,d) and ciphertext c, the decryption function is given by: $m(c)= c^d \text{ mod } n$

VI. SIMULATION RESULTS

In our proposed methodology of 'Disease Classification With E-Report Generation, Authentication and Encryption' the chief libraries included are TensorFlow, NumPy and PyQt5. These python libraries enhance the implementation of neural networks with higher data rate flow, allows performance of complex mathematics and create interactive GUIs. The presented 5-layer CNN model can allow a maximum of 2^{27}- 1 input image scans per class. A data set of 500 MRI scans of diseases like brain tumor, stomach cancer and lung cancer have been taken for training the CNN algorithm. To intensify the accuracy, 5000 training steps are applied with training data ratio of 80 and 20. Consequently, the algorithm gives an accuracy of 99.97% on the prediction and classification of diseases. Additionally, an e- report is generated automatically with double verification by doctor and encryption. The report is encrypted using RSA algorithm of key size 4 digits and input message bit size of 16 bits. This makes the report secure and prevents from mishaps with patient's health information.

©IJRASET: All Rights are Reserved

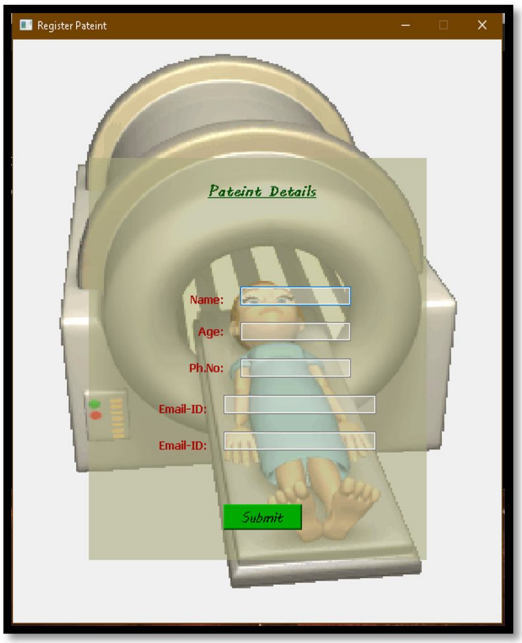


Figure 4. Registration of patient's personal details



Figure 5. Authentication and Decryption of e-report

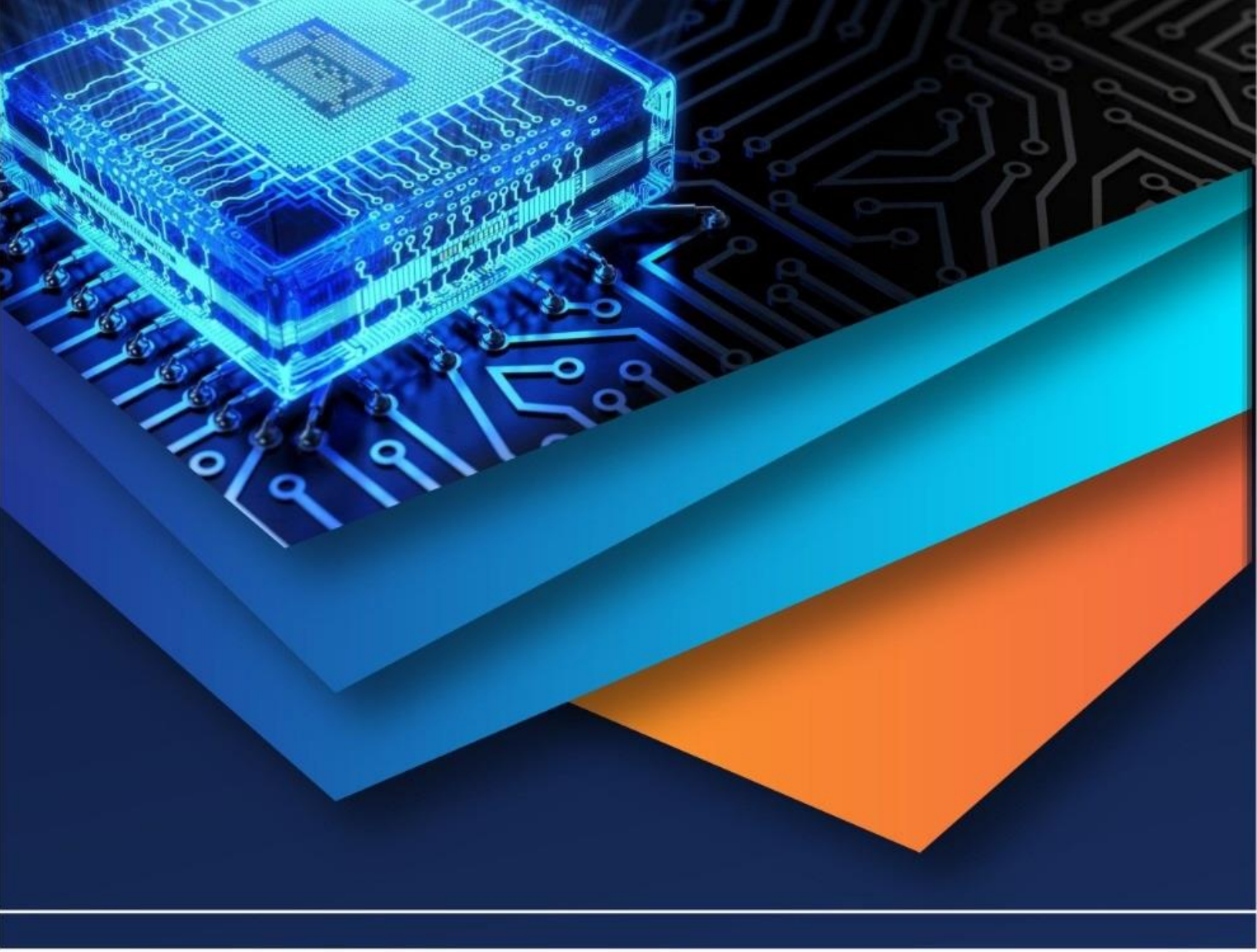
VII. CONCLUSION

The model proposed in this paper results in an efficient disease classification of various diseases like brain tumor, stomach cancer and lung cancer using CNN algorithm. It gives an efficiency of 99.97% which beats the drawback of time and energy consumption in disease revelation. It is highly authenticated and secured. The GUI implemented is interactive. The report generated automatically is encrypted using RSA. Thus, serving the need for security to patient's personal information. The proposed model makes use of highly understanding and robust programming language- Python. In future, we intend to work on the classification of more diseases that are laborious for revelation and time-consuming, which can increase the mortality rate. Also, a stronger encryption algorithm can be implemented for security issues.

REFERENCES

[1] Automatic Detection of Tumor Cells in Microscopic Images of Unstained Blood using Convolutional Neural Networks Ioana Mocan, Razvan Itu, Anca Ciurte, Radu Danescu Computer Science Department Technical University of Cluj-Napoca Cluj-Napoca, Romania radu.danescu@cs.utcluj.ro Rares Buiga Anatomic Pathology Department The Oncology Institute "Prof. Dr. Ion Chiricuta" Cluj-Napoca, Romaniararesbuiga@yahoo.fr
[2] G. Litjens, et al, "A Survey on Deep Learning in Medical Image Analysis", 2017, available at arXiv:1702.05747.
[3] Ehab F. Badran, Esraa Galal Mahmoud, Nadder Hamdy, "An Algorithm for Detecting Brain Tumors in MRI Images", 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, 2017.
[4] Lambert, R., Guilloux, A., Oshima, A., Pompe-Kirn, V., Bray, F., Parkin, M., Ajiki,W., Tsukuma, H. (2002). Incidence and mortality from stomach cancer in Japan,Slovenia and the USA. Internationa Journal Of Cancer,97, 811–818.
[5] Ahmadzadeh, D., Fiuzy, M., Haddadnia, J. (2013). Stomach Cancer Diagnosis by Using a Combination of Image Processing Algorithms, Local Binary Pattern Algorithm and Support Vector Machine.Journal of Basic and Applied Scientific Research, 3(2), 243-251.
[6] V. Krishnaiah, G. Narsimha, C. Subhash.(2013, December) "Diagnosis of lung cancer prediction system using data mining classification techniques,"in International Journal of Computer Science and Information Technologies, Vol. 4, issue 1, pp. 39-45.
[7] K. Murphy, B. Ginneken, A. M. R. Schilham, B. J. Hoop, H. A. Gietema, and M. Prokop, "A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification," Medical Image Analysis, vol. 13, issue 5, pp. 757–770, 2009.
[8] 2014 International Conference on Communication and Network Technologies (ICCNT) 145 Variants of RSA and their Cryptanalysis Kannan Balasubramanian Professor, Dept. of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, India. kannanbala@mepcoeng.ac.in

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)