iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.43981

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: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Multiple Disease Detection Using Machine Learning: A Survey

Dr. Ankita Karale[1], Uday Talpade[2], Sahil Nikumbh[3], Laxman Wadekar[4], Parikshit Angre[5]
[1]*Professor,* [2,3,4,5] *Student, Sandip Institute of Technology and Research Centre, Nashik, India.*

Abstract— In recent years, some researchers have used various machine learning-based approaches to develop autonomous disease detection systems, and early disease identification may help to reduce the number of people who die. The disease detection models aim to bring the medical and artificial intelligence (AI) fields together so that people can understand how well AI and medicine can work together. To better understand the role of Artificial intelligence in the medical field, we plan to conduct a comprehensive study on AI applications for the healthcare sector. First, we'll go over the highlights and motives for using AI in the healthcare industry. Following that, we go over machine-learning-based algorithms for integrating AI and the healthcare sector in depth. Next, we go over the technical problems of AI in the medical industry first, and then show how machine learning can help. We also look into the impact of machine learning in the medical field. Moreover, we also present several notable initiatives that demonstrate the importance of machine learning in healthcare applications and services. Finally, discuss some existing issues in disease identification and suggest future research and development areas that will lead to the usage of machine learning in the healthcare sector.

Keywords— CNN, Deep Learning, MRI, Artificial Intelligence, Neural Network, Image Processing.

I. INTRODUCTION

Rapid advances in technology and the healthcare industry have contributed to changes in people's lifestyles and socioeconomic situations in recent years, raising the risk of people contracting numerous diseases. Major diseases such as brain tumours, lung cancer, and pneumonia, among others, have a global impact. In 2019, roughly 86000 individuals were diagnosed with a brain tumour, according to the World Health Organization, with a 35 percent average survival rate [1], and lung cancer is a terrible disease that kills one in every five people worldwide, or 1.59 million people, accounting for 19.4 percent of all deaths [2]. With over 37 million verified cases and more than 1 million deaths globally, the coronavirus pandemic has impacted various countries [3], has brought diseases like pneumonia to the forefront. These major disorders increase societal pressure and healthcare costs, thereby impacting the patient's overall health. The primary goal of disease detection is to determine whether or not a person is at risk of contracting one or more serious diseases. This necessitates the consideration of numerous issues, which takes a significant amount of manpower and financial resources.

Medical datasets for health-related data are now easily collected by numerous medical institutes all over the world. Image data, patient reports, and other sorts of medical data are examples. Medical data is highly complicated, irregular, and includes unstructured data, making it more challenging to manage. Manual data entry is impossible, and diagnosis is limited, depending on a variety of criteria such as the patient's medical state, the doctor's level, and the differences in patient reports, among others. These issues are addressed by incorporating a machine learning-based disease detection module to aid in disease prediction and diagnosis.

Deep learning uses algorithms to identify and analyze patterns in medical images. In a variety of medical applications, deep learning has improved to the point where it is currently the top of the line. Deep learning can process large amounts of data and extract multiple data features. Deep learning is utilised in domains such as image recognition, natural language processing, and speech recognition. Deep learning has grown into a profound level, also referred as a deep neural network (DNN), as more people seek models, more data, and more processing capacity. As shown in Fig.1 There are three layers to a DNN: input, hidden, and output. It works with forward and backward propagation concepts.

Deep learning models, like human brains, are fed image data and extract distinguishing features, emulating human brain processes like vision as well as other intelligent behaviours. It imitates medical experts during disease identification and accumulates experiences over time through continuous practise to increase detection accuracy and hence make the model more resilient. Deep learning's application has yielded amazing results [4].

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

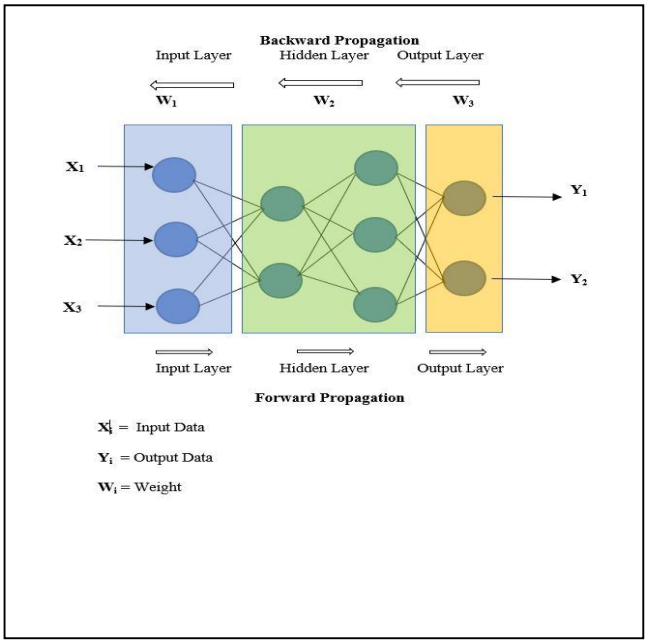


Fig.1 Deep Neural Network.

II. METHODOLOGY

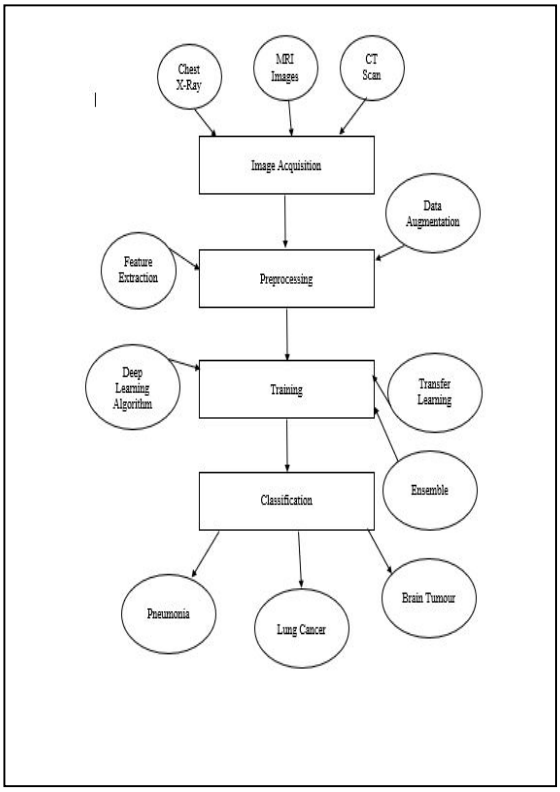


Fig.2 Overview of Methodology of Deep Learning Model.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

A. Image Acquisition Phase:

The acquisition of disease-related images is the initial phase of the disease detection model. Because the CNN technique is used, the model must be trained on a vast number of images. In the scope of this research, images provide critical data for the diagnosis of different diseases. Images such as CT scans and chest X-rays can be used. The first phase's output comprises of images, which are fed into the model for training. Here image dataset is acquired from Kaggle.

B. Data Pre-processing Phase

The image is modified at this stage to increase image quality. Because the images in the dataset are of varying sizes, they are adjusted to have a shape of (224, 224) = (image width, image height) so as to feed it as an input to the neural network, as all images must have the same shape. To expand the quantity of data available, data augmentation is performed on the images. To scale pixel values to the range 0–1, normalization is used. Feature extraction is done so that the DNN model can find relevant features that can be used to classify certain class. The result is a series of images that have been upgraded in quality or have undesired elements removed.

C. Training Phase

The selection of a deep learning algorithm is done in the third phase, training. The previously described CNN is an example of a deep learning algorithm. Algorithms can learn in a variety of ways. Certain algorithms work best with specific types of data. CNN is a adept at using images. The type of data should determine which deep learning method is used. The models created from the data learned are the result of this step.

D. Classification Phase

The last phase is classification, in which the trained model predicts which class an image belongs to. For example, if a model has been trained to distinguish between normal and tumorous brain in MRI images, it should categorise images accordingly. The model assigns a probability score to each image, indicating how probable it is that the image belongs to a given class.

III.DATA AUGMENTATION

Data Augmentation is a technique for increasing training datasets without having to gather new images. Data augmentation alters the original images in some way. This is accomplished by using various processing techniques including as rotations, flips, zooming, and adding noise, among others. Large training datasets are significant in deep learning since they improve the training model's accuracy. It also aids in the avoidance of overfitting. The downsides of data augmentation include increased training time, transformation computation costs, and higher memory expenses.

IV.RELATED WORK

According to [5], a new strategy incorporating random forests (RF) and contour-based models is utilised to extract glioma features from multivariate volumetric MR images. They also use random forests algorithm as feature training kernels to analyse both geographic data, accuracy data from multiple images for tumour diagnosis utilising a feature representational technique for learning.

Discriminatory features and sparseness were included into the PCA model by [6]. Instead of the standard sparse PCA, which enforces sparseness on the loadings, sparse components are created to reflect the data.

According to [7], A suggested strategy based solely on 3D convolutional neural networks (CNN) provides effective performance on the publicly available dataset for lung nodule identification and malignancy classification. While methods for detecting nodules are frequently developed and improved separately, the relationship between component identification and classification is crucial.

According to [8], For lung nodule recognition and categorization, an enhanced multidimensional Region-based Fully Convolutional Network (mRFCN) based automated system was used. The mRFCN is being used to investigate the multi-Layer fusion Region Proposal Network (mLRPN) using position-sensitive score maps (PSSM) as an image classifier for extracting features. Then, using the suggested mLRPN, a median intensity projection was employed to take benefit of 3D information from CT scans, and then a de-convolutional layer was added to the architecture to autonomously choose possible zones of interest.

According to [9], The current surveys, and also latest deep learning-based approaches for brain tumor categorization, were thoroughly examined. The survey covers the basic methods of deep learning-based brain tumor categorization techniques such Data preprocessing, extraction of features, and categorization, as well as their accomplishments and limitations.

Notable changes, according to [10], include the addition of brain invasion as a list of requirements for atypical meningioma, as well as the inclusion of a soft tissue grading system for the newly merged entity of isolated fibrous tumor hemangiopericytoma—a divergence from what other CNS tumors are graded—are among the changes. Overall, the 2016 CNS WHO is intended to fund medical, scientific, and epidemiologic research that will help people with brain tumors live better lives.

Most spouses witnessed months of global dysfunction prior to the symptom that led to physician appointment, according to [11]. The patient factors of "less alien symptoms," "personality change," and "avoidance," as well as the spouse factors of "spouse's passivity" and "spouse's successive adaptation," and the physician factors of "reasonable alternative diagnosis," "physician's inflexibility," and "physician's personal values," were identified as roadblocks on the way to appropriate medical care.

According to [12], The phrase "brain tumors" refers to a broad variety of tumors that grow from intracranial tissue and different tissue layers and can be benign or malignant in nature. Each tumor has its own biology, therapy, and prognosis, and distinct risk factors are likely to cause it. Because of their position in the brain, their ability to penetrate regionally, and their inclination to turn malignant, they are particularly dangerous, even "benign" tumors can be deadly. This complicates the classification of brain tumors and makes it difficult to describe the epidemiology of these diseases.

According to [13], The Brain Cancer Module is a brain tumour scale that can be combined with other questionnaires. In both low-grade and high-grade glioma patients, HRQL measurement and neuropsychiatric evaluation were utilised to assess the impact of radiotherapy and surgery, and also the impact of tumour size, tumour distribution, performance level, and age.

According to [14], Noncognitive computer user interfaces can detect gestures and carry out commands based on them. The concept is implemented on a Linux system, however Python modules can easily be installed on a Windows machine. The platforms used for the identification are OpenCV and KERAS. The vision-based algorithms recognise the gestures depicted on the screen. Lenet architecture trained an assortment of skin colour masks in KERAS for recognition using a background removal technique.

The normal brain tumour and the malignant brain tumour should be recognised, according to [15]. MRI is used to investigate several types of brain tumors, such as metastatic bronchogenic carcinoma tumours, glioblastoma, and sarcoma. In order to identify and categorise MRI brain tumours, several wavelet techniques and SVM algorithms are applied. For medical analysis and evaluation, effective and autonomous identification of MRI images of brain is critical.

Entropy, mean, correlation, contrast, energy, and homogeneity are six aspects that [16] focus on. The accuracy, sensitivity, and specificity performance metrics are calculated to show that the suggested method outperforms existing methods. The suggested technique is used to determine the size and location of a brain tumour by utilising an MRI image and MATLAB.

The goal can be achieved, according to [17], by executing the major measures to follow: Brain image pre-processing, segmentation of disease tissues, extraction of meaningful data from every segmented tissues, and categorization of tumour images using a Neural Network. The Quality Rate with normal and abnormal MRI images is often used to evaluate the experimental outcomes and assessments.

According to [18], the suggested method is highly effective and exact in diagnosing, categorising, and segmenting brain tumours. This necessitates the use of accurate automatic or semi-automatic techniques. The study provides an automatic segmentation method that finds small 3x3 kernels using CNN (Convolution Neural Networks). Segmentation and classification can be accomplished by combining these two techniques. CNN evolved from NN, a machine learning technology that employs layers to identify outcomes (Neural Networks). Data collection, Data pre-processing, filtering, segmentation, extraction of features, Convolutional neural network via classification, and identification are among the phases included in the proposed methodologies. Data mining techniques can be used to extract important patterns and relationships from data.

According to [19], Optimal Feature Level Fusion (OFLF) is used to fuse low and high level features of a brain image, and the images are classified as benign or malignant based on this analysis. The experiment findings include analysing performance indicators and comparing current classifiers based on this medical images. When compared to existing classifier, the suggested MRI image classification technique has an accuracy of 96.23 percent, a sensitivity of 92.3 percent, and a specificity of 94.52 percent. This proposed methodology is implemented using the MATLAB working platform.

V. RELATED WORK

A. Existing Problems

By extracting characteristics and classifying them, the deep learning model achieves great results. However, the current system's model interpretability is inadequate. It is necessary to connect medical domains and deep learning interpretability of models. The interpretability of models refers to the degree to which humans can comprehend decision-making logic. Deep learning's problem is that it is entirely data-driven, with no regard for prior domain expertise or experience, as well as risk considerations. The deep learning model has been fully trained, and new data is being fed into it, with detection results being generated. The model, on the other hand, only presents the classification results depending on the input data and does not indicate how to detect or predict. The model's credibility is determined by its interpretability. As a result, future research should pay more attention to model interpretability.

The majority of disease detection models are still in the theoretical stage and have yet to be put into practice. The following are some of the reasons for this:

1) *Stability:* A high level of stability is required to apply deep learning to healthcare systems. It is possible that the sample training dataset does not match the real sample dataset when a convolutional neural network is utilised in actuality. If the model's stability cannot be assured, it will have a negative effect on performance, efficiency, as well as imprecise disease prediction, which could put the patient's life in danger.
2) *Data Security and Privacy:* Medical records and private details of patients are essential in order to predict diseases, Hence data security and privacy must be considered. To protect privacy, different alternatives such as blockchain based technology, decentralized networks should be examined.

B. Issues:

This section discusses the issues with detection of diseases using deep learning model that have been reported in the literature.
Handling large image sizes,
Limited available datasets.
Data imbalance were recognised as three key issues.

1) *Handling of Large Image Size:* Because it is difficult to train a model using the original image size as it is computationally expensive and time intensive, image sizes are typically reduced during model training.
2) *Limited Available Datasets:* A high number of images are necessary for a more accurate training model, yet training data is less than optimal due to limited datasets.
3) *Data Imbalance:* If one class gets a lot more data than the other when building a classification model, the final model will be biased. If each class has the same number of images, that is ideal.

VI. CHALLENGES IN MODEL TRAINING

The most difficult aspect of model training is still data quality. High-quality medical data is required for Deep learning models to perform well in prediction and diagnosis. Despite the ease with which medical data can be obtained under current circumstances, the data quality is still poor. To offer a proprietary label to much medical facts, medical experts must have a great deal of experience. Image feature analysis is particularly significant [20], because medical data sets are kept in different institutions due to several privacy issues. A large majority of the data sets cannot be used in legitimate research since they are closed rather than open. Many novel models are hindered by the inability to obtain proper training [21].

VII. FUTURE WORK

This section describes future research that should be done to improve disease detection capability using deep learning.
Data from private hospitals was used in several research. In order to obtain larger datasets, efforts such as de-identification of personal patient data may be undertaken. If more data was supplied, the classifiers developed would be more accurate. This is due to the fact that more data means more diversity. As the model is trained on more examples, it becomes more general, the generalization error is reduced. Medical information is difficult to come by. As a result, if the databases were made public, researchers would have access to additional information.
The use of cloud computing for training may be able to solve the difficulty of dealing with large image sizes. On a local mid-range PC, training with large graphics will be slow. Even if a high-end computer could speed up the process, it might not be feasible. However, we can leverage several GPUs at a minimal cost by using cloud computing to train the deep learning model. This enables higher-cost computational training to be performed faster and for less money.

VIII. CONCLUSION

Under the backdrop of artificial intelligence and deep learning techniques, the future of medical healthcare has more modern prospects. Deep learning has emerged as a primary driving force for future progress in the face of medical data instability, thanks to its unique feature processing approach and variable model structure. The deep learning models are linked together and learn from one another, creating a more complex deep learning system network which contributes to the improvement of the medical profession by assisting in the development of medical diagnosis and practical applications.
We have highlighted the most common deep learning approaches in this study. The approach and existing challenges, as well as the limitations of deep learning are also highlighted. We also go over the many security and privacy issues and obstacles that have been encountered. We also highlight several relevant papers, as well as various research concerns that need to be addressed further.

REFERENCES

[1] Rehman A, Khan MA, Saba T, Mehmood Z, Tariq U, Ayesha N. Microscopic brain tumor detection and classification using 3D CNN and feature selection architecture. Microsc Res Tech. 2020;1–17.

[2] Mushtaq, R., & Loan, F. A. (2021). Lung cancer research in India and Iran: a scientometric study. *Library Philosophy and Practice (e-Journal), 4761.*

[3] Ibrahim, A.U., Ozsoz, M., Serte, S. *et al.* Pneumonia Classification Using Deep Learning from Chest X-ray Images During COVID-19. *Cogn Comput* (2021).

[4] Cao, C., Liu, F., Tan, H., Song, D., Shu, W. et al. (2018). Deep learning and its applications in biomedicine.Genomics, Proteomics & Bioinformatics, 16(1), 17–32. DOI 10.1016/j.gpb.2017.07.003.

[5] C. Ma, G. Luo, and K. Wang, "Concatenated and connected random forests with multiscale patch driven active contour model for automated brain tumor segmentation of MR images," IEEE Trans. Med. Imag., vol. 37, no. 8, pp. 1943–1954, Aug. 2018.

[6] C.-M. Feng, Y. Xu, J.-X. Liu, Y.-L. Gao, and C.-H. Zheng, "Supervised discriminative sparse PCA for com-characteristic gene selection and tumor classification on multiview biological data," IEEE Trans. Neural Netw. Learn. Syst., vol. 30, no. 10, pp. 2926–2937, Oct. 2019.

[7] Onur Ozdemir, Rebecca L. Russell "A 3D Probabilistic Deep Learning System for Detection and Diagnosis of Lung Cancer Using Low-Dose CT Scans" IEEE TRANSACTIONS ON MEDICAL IMAGING *last revised 21 Jan 2020*.

[8] Anum Masood, Bin Sheng, Po Yang, Ping Li "Automated Decision Support System for Lung Cancer Detection and Classification via Enhanced RFCN with Multilayer Fusion RPN" IEEE Transactions on Industrial Informatics (Volume: 16, Issue: 12, Dec. 2020).

[9] Khan Muhammad; Salman Khan; Javier Del Ser; Victor Hugo C. de Albuquerque. "Deep Learning for Multigrade Brain Tumor Classification in Smart Healthcare Systems: A Prospective Survey" IEEE Transactions on Neural Networks and Learning Systems (Volume: 32, Issue: 2, Feb. 2021).

[10] David N. Louis, Arie Perry, et al. , "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary" , Acta Neuropathol , Springer may 2016.

[11] Pär Salander, A Tommy Bergenheim, Katarina Hamberg, Roger Henriksson, Pathways from symptoms to medical care: a descriptive study of symptom development and obstacles to early diagnosis in brain tumour patients, Family Practice, Volume 16, Issue 2, April 1999, Pages 143–148,

[12] McKinney PA ,"Brain tumours: incidence, survival, and aetiology",Journal of Neurology, Neurosurgery & Psychiatry 2004;75:ii12-ii17.

[13] Heimans, J., Taphoorn, M. Impact of brain tumour treatment on quality of life. J Neurol 249, 955–960 (2002).

[14] Malavika Suresh, et al. "Real-Time Hand Gesture Recognition Using Deep Learning", International Journal of Innovations and Implementations in Engineering (ISSN 2454- 3489), 2019, vol 1

[15] M. Gurbină, M. Lascu and D. Lascu, "Tumor Detection and Classification of MRI Brain Image using Different Wavelet Transforms and Support Vector Machines", 42nd International Conference on Telecommunications and Signal Processing (TSP), Budapest, Hungary, 2019.

[16] Somasundaram S and Gobinath R, "Early Brain Tumour Prediction using an Enhancement Feature Extraction Technique and Deep Neural Networks", International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278- 3075, Volume-8, Issue-10S, August 2019.

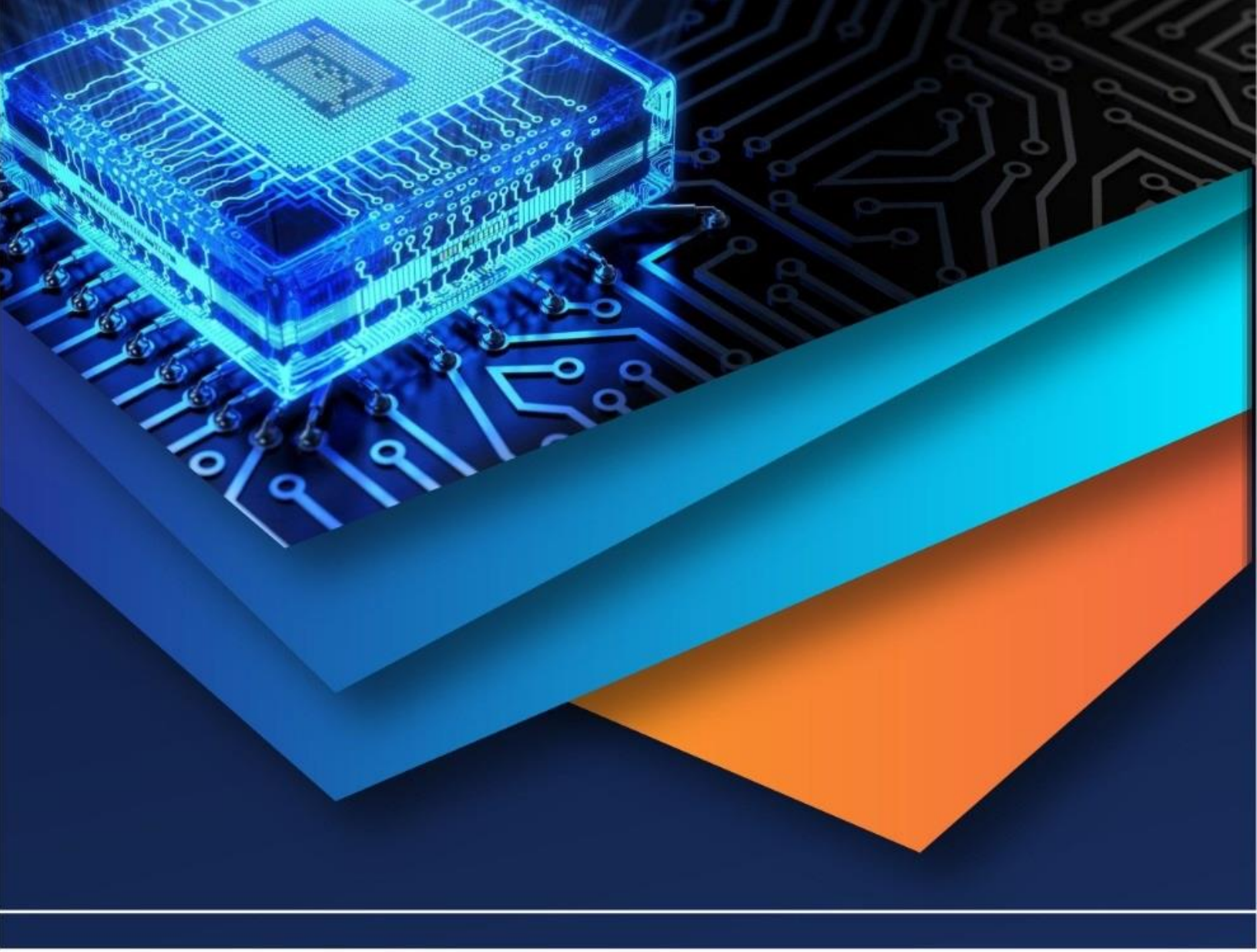
[17] Damodharan S and Raghavan D, "Combining Tissue Segmentation and Neural Network for Brain Tumor Detection", The International Arab Journal of Information Technology, Vol. 12, No.1, January 2015.

[18] G. Hemanth, M. Janardhan and L. Sujihelen, "Design and Implementing Brain Tumor Detection Using Machine Learning Approach", 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019.

[19] Shankar, Dr & Elhoseny, Mohamed & Lakshmanaprabu, S.K. & .M, Ilayaraja & RM, Vidhyavathi & Abu Elsoud, Mohamed. (2018). Optimal feature level fusion based ANFIS classifier for brain MRI image classification. Concurrency and Computation Practice and Experience. 10.1002/cpe.4887.

[20] Yang, X., Zhang, T., Xu, C., Yan, S., Hossain, M. S. et al. (2016). Deep relative attributes. IEEE Transactionson Multimedia, 18(9), 1832–1842. DOI 10.1109/TMM.2016.2582379.

[21] Levine, A. B., Schlosser, C., Grewal, J., Coope, R., Jones, S. J. et al. (2019). Rise of the machines: Advancesin deep learning for cancer diagnosis. Trends in Cancer, 5(3), 157–169. DOI 10.1016/j.trecan.2019.02.002.



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)