iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** III **Month of publication:** March 2019

DOI: http://doi.org/10.22214/ijraset.2019.3096

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: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

Design and Development of Online Portal for Diagnosis of Multiple Diseases using AI

Keval Chavda[1], Chiranjev Koul[2], Pravin Deshatwar[3]
[1, 2, 3]*Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India*

Abstract: ***The number of medical ailments are increasing at a significant rate having high intensity. Occasionally, this makes it difficult for the doctors to identify the diseases and are often not able to offer any assurance for the disease being present. A solution to such type of problems can be the use of machine learning algorithms. The machine learning algorithms can help identify the diseases to a great extent if appropriate type of algorithm is used to detect a particular disease. With the development of research in the biomedical field, the algorithms have proved to be able to detect the diseases with substantial accuracy. Our paper provides a detailed analysis of the key algorithms and their results used primarily with different types of diseases. The misdiagnosis of the diseases by some of the inadequately experienced doctors is often proved fatal to the patients. This project can possibly be used to help the doctors and the patients in the early detection of the diseases which are not easily detectable. The result of these can be used to an early treatment and a more healthy recovery.***

Keywords: Machine Learning; Cloud Computing; Artificial Intelligence; Image Processing; Classification Models; Disease Diagnosis

I. INTRODUCTION

Artificial Intelligence has enabled us to efficiently detect diseases with accuracy which previously required an enormous amount of time and calculations. Machine Learning has branched out over the years with the introduction of several different algorithms having different accuracy with different types of datasets. The computational capabilities of the machine learning algorithms have enabled us to rethink the process of detection and analysis of the diseases of the human body. These algorithms build models based on the training data and then the models are used to make predictions on new data. As the healthcare industry is getting further dependent on the computational capabilities, the machine learning models are required by the doctors in the assistance of detection and curing of ailments at an early stage. Such algorithms provide facilities such as accuracy, efficient speed and reliability. Moreover, shifting such models over a cloud platform provides easy access to the people enabling them to use it by their own convenience.

II. LITERATURE SURVEY

A paper by Rajesh Kumar, Harish Kumar, Seema Verma and Smita Jhajharia in 2016 presented a neural network based breast cancer prognosis model. This model used featured for PCA algorithm. In a pathologist's reports, a patient's biological tissue sample are usually considered as the gold standard of many diseases as a review in diagnosis. Cancerous tumor mass has been found as the major cause of breast cancer. Smita Jhajharia, a pathologist, proposed a neural network based breast cancer prognosis model which has the principal component analysis feature. It is a multivariate statistical approach that has been coupled with an artificial intelligence based learning technique to implement a prediction model. PCA pre-processes the data and extracts features in the most efficient form for the training of the ANN. Artificial neural networks learns the pattern in the data for classification of new instances. The accuracy for this experimental analysis is expected to be 96%.

In 2017, Shubangi Khobragade, Aditya Tiwari and Vikram Narke presented a paper on detection of lung diseases using the chest radiographs. This was worked on a feed forward artificial neural network. Artificial Intelligence is often used for improvisation of the accuracy in the diagnosis of lung diseases. Machine learning utilizes algorithms which can learn and perform predictive data analysis. Shubhangi Khobragade has proposed an algorithm for automatic detection of major lung diseases. The lung image segmentation and it's analysis via AI and Neural network will determine major lung diseases by feature extraction. Image classification will determine geometric feature with algorithm and this output accuracy is found to be 86%.

Researchers such as Otoom A.F., Abdallah E. E., Kefaye A., and Ashour gave a paper on diagnosis of heart diseases in 2015. Machine learning algorithms are usually used to help improve the accuracy in the diagnosis of heart diseases. Otoom proposed an algorithm for detection and analysis of Coronary artery disease. Performance analysis is carried out using three algorithms Bayes Net, Support vector machine, and Functional Trees (FT). Bayes Net attained a correction of 85.5%, whereas SVM provided an accuracy of 85.1% and the Functional Trees classified the data showing 84.5% correctness.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

Moreover, another paper on diagnosis of diabetes using multiple classification mining techniques was presented by Iyer A., Jeyalatha S., and Sumbaly R. Diabetic Macular Edema is one of the main cause of visual loss in diabetes. There are many machine learning algorithm for the diagnosis of diabetes retinopathy. Iyer performed a work to predict diabetes disease by using decision tree and Naive Bayes. This paper presents 76.95% accuracy by using Cross Validation and Percentage Split Respectively. Naive Bayes shows 79.56% correctness by using PS. Algorithms displaying the highest accuracy by utilizing percentage split test.

III.PROPOSED WORK

The project is an online portal that can be used by the doctors and patients to scan their reports and get a particular output and information regarding the detected diseases. The portal is hosted on a cloud server which makes it easier for the users to access it on the go. The portal makes the use of machine learning algorithms such as Naive Bayes, Convolution Neural Networks and Support Vector Machines. The model after it is trained would then be used to detect and identify the diseases more efficiently. For the working of the system, initially, the users would have to sign up and have their account verified. This would enable them to access the portal and thus being able to upload the images of the reports. Once the body part and type of disease is selected in the user interface, the model would begin the processing of the uploaded document and provide the users with the resultant information regarding the detection of the disease. All the reports of the users would be stored in the cloud server. The users can access these reports at any time. The proposed system has an interactive, responsive and minimal user interface making it easier for the people to use. This project is expected to be used in clinics, hospital, healthcare and even homes.

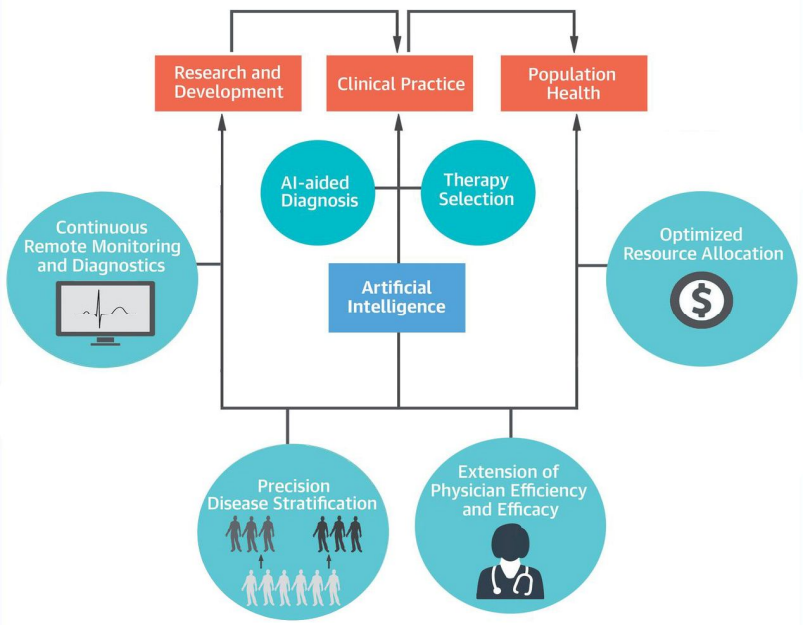


Fig. 1 Working of online portal for diagnosis of disease

The flow of the system is divided into 6 major parts: Image Acquisition, Preprocessing, Segmentation, Feature extraction, Classification and Result. The Image Acquisition enables the user to upload their reports to the system for its processing. The pre-processing helps in improvement of image data that suppresses the unwanted distortions or enhances some image features important for further processing. It can then proceed to the Segmentation which comprises of 2 segments: Adaptive thresholding and Region detection. After the test data is divided into segments. They are used to determine the body part of the ailment and this is then converted into validate data. Machine learning algorithms is then applied on this data to classify the type of disease into the dataset, thus, providing us with the output [1] [5].

A. Malaria Detection

We have utilized Keras deep learning library to automatically analyse medical photographs for malaria testing. Such a deep learning library combined with medicinal imaging framework lessen the significant measure of deaths per year caused by malaria. We trained a deep learning model on medical images to predict if a given patient's blood drops are positive for malaria or not.

Malaria is a contagious disease that is known of causing more than 400,000 deaths per year. It is truly prevalent in some areas of the world, which implies that the disease is consistently found in those regions. In the other regions of the world, malaria is epidemic [8].

Yet in other areas of the world malaria is seldom, if ever, found at all. There are numerous factors that make a region prone to an infectious disease outbreak.

The diagnosis of malaria requires precision and comprises of a numerous steps. First, a blood sample is taken from a patient and after that put on a slide. A contrasting agent is is then used to recolor the example which will help in featuring the malaria parasites in red blood cells [11]. The slide is then inspected by a clinician under a microscope and the number of red blood cells are checked manually. This can be an extensive procedure and are prone to contain unintentional mistakes by the people working on it. We have utilized the malaria dataset provided by the National Institute of Health (NIH). The dataset contains of 27,588 images belonging to two distinct classes:

1) *Parasitized:* which infers that the region contains malaria.
2) *Uninfected:* which infers that there is no disease in the region

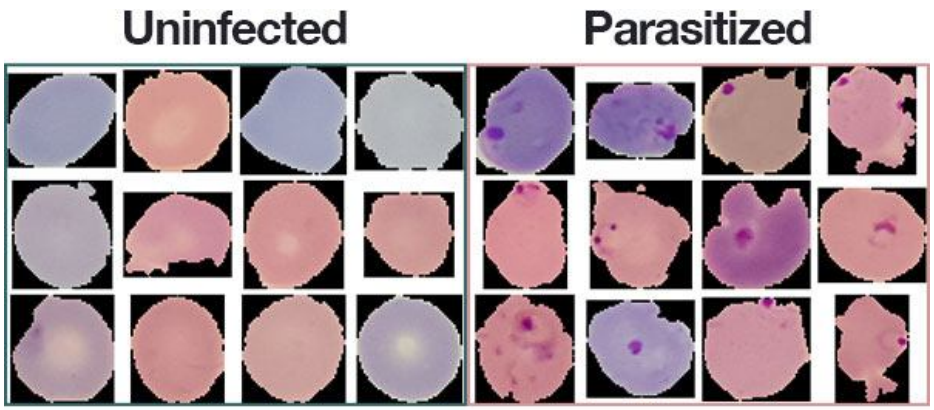


Fig. 2 Comparison of the malaria datasets

The quantity of pictures per class is similarly circulated with 13,794 pictures for each individual class. We used 80% of data for training and 10% of the training data for validation.

Rest of the data was used for testing purposes. We have used the ResNet architecture to train our deep learning model with Keras. After using the model, we were able to notice that a smaller variation of ResNet whose model size is only 17.7MB, we were able to obtain an accuracy of 97%.

B. Breast Cancer Prediction

One of the most common cancer among the women in the world is the breast cancer. We were able to notice that it accounted for about 25% of all cancer cases and affected over 2.1 million people in 2015 alone. It usually starts when cells in the breasts begin to grow out of control.

These cells tend to form tumors which can be seen via X-ray or even be felt as lumps in the breast areas [12].

The chances of survival can significantly increase if it is diagnosed at an early stage. One of the key difficulties against breast cancer's detection is whether how would we characterize tumors into being benign (non-cancerous) or malignant (cancerous). A tumor can be viewed as benign if it does not invade adjacent tissue or spread to the alternate parts of the body in the manner in which cancerous tumors can. A threatening tumor's cells can develop into encompassing tissues or spread to remote areas of the body. Indeed, even benevolent tumors can go about as difficult issues on the off chance that they push on some imperative structures, for example, nerves or veins.

We used the UCI Machine Learning Repository for breast cancer dataset and removed any missing or null data points of the dataset using the Pandas library. Categorical data can be referred to as variables that comprise of label values instead of numerical values. The total amount of probable values is often limited up to a fixed set. For instance, users are typically labelled by gender, age, country, age group, etc.

We have used Label Encoders to label these categorical data. Label Encoder is a part of the scikit-learn library in Python. This Label Encoder is used to form numbers from categorical data or textual data which can be used by our predictive models to better understand. We split the dataset into 80% training data and 20% testing data. We brought all features to the same level of magnitudes using feature scaling. We applied various classification techniques but SVM achieved the highest accuracy of 97%.

C. Lung Disease Prediction

Advancements in the technology and advances computational power has helped with the earlier identification of diseases, particularly lung diseases. We can use these advancements with the earlier and accurate detection, which can save many people as well as reduce the pressure on the system. The health system has not developed in time with the development of the population [14] [15].

There are 15 classes in the dataset, 14 for different lung diseases and 1 for no findings. The dataset contains several images of the chest X-Ray for each of the classes which is used to train the model. These 14 classes include Infiltration, Pleural thickening, Consolidation, Pneumothorax, Edema, Fibrosis, Effusion, Nodule, Atelectasis, Pneumonia, Cardiomegaly, Mass, Hernia and Emphysema. The application is supposed to determine whether or not a person has any of the lung diseases or not.

We have used Tensorflow library to automatically train the model and the users can then predict whether they have any of the 14 specified diseases or no disease. We chose to perform CNN since it gave us the most accurate results and applied techniques such as spacial transformation and VGG fine-tuning on it. Although this does increase the training time, it results in the increase of overall precision and accuracy as well.

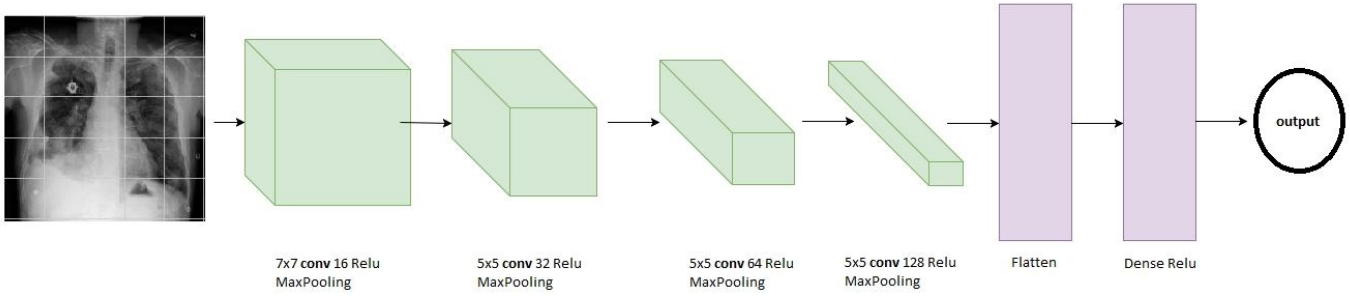


Fig. 3 CNN used on the Lung X-Ray image

D. Diabetic Retinopathy Prediction

The leading cause of blindness among the working age population of the developed world is the diabetic retinopathy. It is estimated to affect over 93 million people. It is estimated by the US Center for Disease Control and Prevention that 29.1 million people in the US have diabetes and estimated by the World Health Organization that 347 million people have diabetic retinopathy worldwide. Diabetic Retinopathy (DR) is a type of eye disease affiliated with well-established diabetes. Between 40% to 50% of the Americans have some stage of the DR. Movement to vision impedance can be moderated or turned away if DR is recognized in time, however this can be very challenging as the disease often shows few symptoms until it is past the point where it is possible to give effective treatment [13].

At present, identifying DR is a tedious and manual process that requires a trained clinician to assess and evaluate digital color fundus images of the retina. The results take time to be obtained by the human results. This can be overdue, often a day or two late. These long-delayed results can lead to lost miscommunication, follow ups, and postponed treatment [16].

The requirement for an automated and complete method of Diabetic Retinopathy screening has for quite some time been perceived as a difficult process. The previous efforts have gained great ground progress using pattern recognition, image classification, and AI. The use of CNN through the Tensorflow library substantiated the results to have an accuracy of above 93%. The user upon uploading the image of the retina would be able to detect if it has presence of the diabetic retinopathy in it or not along with an output from a scale to determine how critical the disease is.

E. The Portal

The front-end has been integrated using the Django framework so that it is easier for the users to access any of the data over the internet. Moreover, to train the models, we have used the Google Colab. Google Colab is a free cloud service and offers its GPUs for free to be used by us. It is a Jupyter based notebook and we can run python scripts on it. We then exported the trained models to the Amazon server. The portal is hosted on an amazon server using an EC2 database to store the user credentials. The AWS s3 is used to store the diagnostic files uploaded by the user which is then tested with the trained model to predict the diagnosis of the patients.

IV. RESULTS & DISCUSSION

Major use of this model would be in medical centre in rural areas. This model can be used to build an Android and iOS App which could be used by consumer to identify the ailments caused in the day to day life. If the usage of this model is very beneficial and helpful to the society then the system can be expanded at global level.

A. Malaria

The model is a smaller variant of the ResNet of only 17.5 MB. Quantizing the weights of the classes in the model would allow us to obtain the model with size less than 10 MB. Using the model, we were able to obtain 97% accuracy in about 1.5 hours. We used 50 epochs for training the model.

B. Breast Cancer

We used several different algorithms to obtain beneficial outputs considering accuracy and performance. SVM seemed to give the highest accuracy (97.13%) with the lowest error rate, outperforming all other algorithms.

C. Lung Diseases

The average accuracy of clinicians' diagnoses was 38%. The clinicians were best at identifying chronic obstructive pulmonary disease (COPD), having accurately diagnosed 74% of the cases of this disease. For other disease groups, the clinician's accuracy rarely exceeded 50%.

The diagnoses made by AI, on the other hand, on average, were 68% accurate. For diagnosing Nodes, the AI achieved a positive predictive value of 83% and a sensitivity of 78%. The positive predictive value and sensitivity of AI for Pneumonia (66% and 82%, respectively) and interstitial lung diseases (52% and 59%) were both significantly greater than those achieved by the clinicians.

D. Diabetic Retinopathy

With the use of a multi-layered perceptron, we were able to achieve a decent accuracy of 78.56%. Later on, using the CNN algorithm, we obtained a maximum sensitivity of 95% and a predictive capacity of 93%.

V. CONCLUSIONS

This paper provides a survey of latest developments in medical field of disease detection and diagnosis using machine learning. The system analyses various machine learning algorithms which can be used to detect diseases like brain tumour, lung cancer, skin cancer etc. that can prevent misdiagnosis by the doctor. It will also help doctors to evaluate their reports and prevent any human error. The various disease diagnosis using machine learning systems can lower the false positive rates and improve the speed of decision making. From the results, it can be clearly noticed that various machine learning algorithms provide enhanced accuracy on detection of different diseases.

VI. FUTURE SCOPE

The project is expected to be deployed on a global scale. Apart from the detection of diseases, the implementation of the suggestions of possible remedies and tips is also planned. Moreover, people would be able to access the system through an android application. We also aim to further expand the entire portal to not only detect human diseases but even plant based diseases which would also be extremely helpful for the farmers. Lastly, we plan to make this system extremely swift, reducing the processing time to a bare minimum.

VII. ACKNOWLEDGMENT

We sincerely are grateful to the Professor and Dean of Academics, Dr. R. R. Sedamkar, of Thakur College of Engineering & Technology, Mumbai University, for providing the facilities to accomplish the present research work. We thank him for his guidance, constant support and teachings to better ourselves and the society.

REFERENCES

[1] Dinu A.J.*, Ganesan R1, Felix Joseph2 and Balaji V3, (2017) "A study on Deep Machine Learning Algorithms for diagnosis of diseases" International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 17 (2017) pp. 6338-6346.
https://www.ripublication.com/ijaer17/ijaerv12n17_03.pdf

[2] Meherwar Fatima1, Maruf Pasha2, 2017, "Survey of Machine Learning Algorithms for Disease Diagnostic" Journal of Intelligent Learning Systems and Applications, 2017, 9, 1-16
https://file.scirp.org/pdf/JILSA_2017012413273284.pdf

[3] Sajda P1., 2006, ‘’Machine learning for detection and diagnosis of disease." Annu Rev Biomed Eng. 2006; 8:537-65.
https://www.ncbi.nlm.nih.gov/pubmed/16834566

[4] Cardoso I1, Almeida E1, Allende-Cid H2, Frery AC1, Rangayyan RM3, Azevedo-Marques PM4, Ramos HS1, 2018, “Analysis of Machine Learning Algorithms for Diagnosis of Diffuse Lung Diseases” 2018 Oct 8. doi: 10.3414/ME17-02-0023.
https://www.ncbi.nlm.nih.gov/pubmed/30296808

[5] Allen Daniel Sunny1, Sajal Kulshreshtha2, Satyam Singh3, Srinabh4, Mr. Mohan Ba5, Dr. Sarojadevi H.6, 2018, “Disease Diagnosis System By Exploring Machine Learning Algorithms”, Volume 10 Issue 2 May 2018 ISSN: 2319-1058
http://ijiet.com/wp-content/uploads/2018/06/3.pdf

[6] Anju Jain, 2015, “MACHINE LEARNING TECHNIQUES FOR MEDICAL DIAGNOSIS: A REVIEW”, ICSTM 2015
http://data.conferenceworld.in/ICSTM2/P2449-2459.pdf

[7] Dinu A.J.*, Ganesan R1, Felix Joseph2 and Balaji V3, 2017, “A study on Deep Machine Learning Algorithms for diagnosis of diseases”, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 17 (2017) pp. 6338-6346
https://pdfs.semanticscholar.org/513e/a1b5f2bbee25522f8c07636d7447fbd9ea5c.pdf?_ga=2.80106391.121878047.1539503517-1373670992.1539503517

[8] Donald M. Steinwachs; Ronda G. Hughes.”Health Services Research: Scope and Significance”,Hughes RG, editor. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008 Apr.
https://www.ncbi.nlm.nih.gov/books/NBK2660/

[9] Stephanie Shimada, Kim Nazi, Gang Luo, and Gillian Strudwick, 2017, “Patient Health Record Systems Scope and Functionalities: Literature Review and Future Directions”, J Med Internet Res. 2017 Nov; 19(11): e388.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5707430/

[10] Adedeji Olugboja; Zenghui Wang, 2017, “Malaria parasite detection using different machine learning classifier”, IEEE Xplore Res. 2017 Nov; 17(16): e388.
https://ieeexplore.ieee.org/document/8107772

[11] Das, Dev Kumar & Ghosh, Madhumala & Pal, Mallika & Maiti, Asok & Chakraborty, Chandan. (2012). Machine learning approach for automated screening of malaria parasite using light microscopic images. Micron (Oxford, England: 1993). 45. 10.1016/j.micron.2012.11.002.
https://www.researchgate.net/publication/233877022_Machine_learning_approach_for_automated_screening_of_malaria_parasite_using_light_microscopic_images

[12] Hiba Asri, Hajar Mousannif, Hassan Al Moatassime, Thomas Noel, 2016, “Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis”
https://www.sciencedirect.com/science/article/pii/S1877050916302575

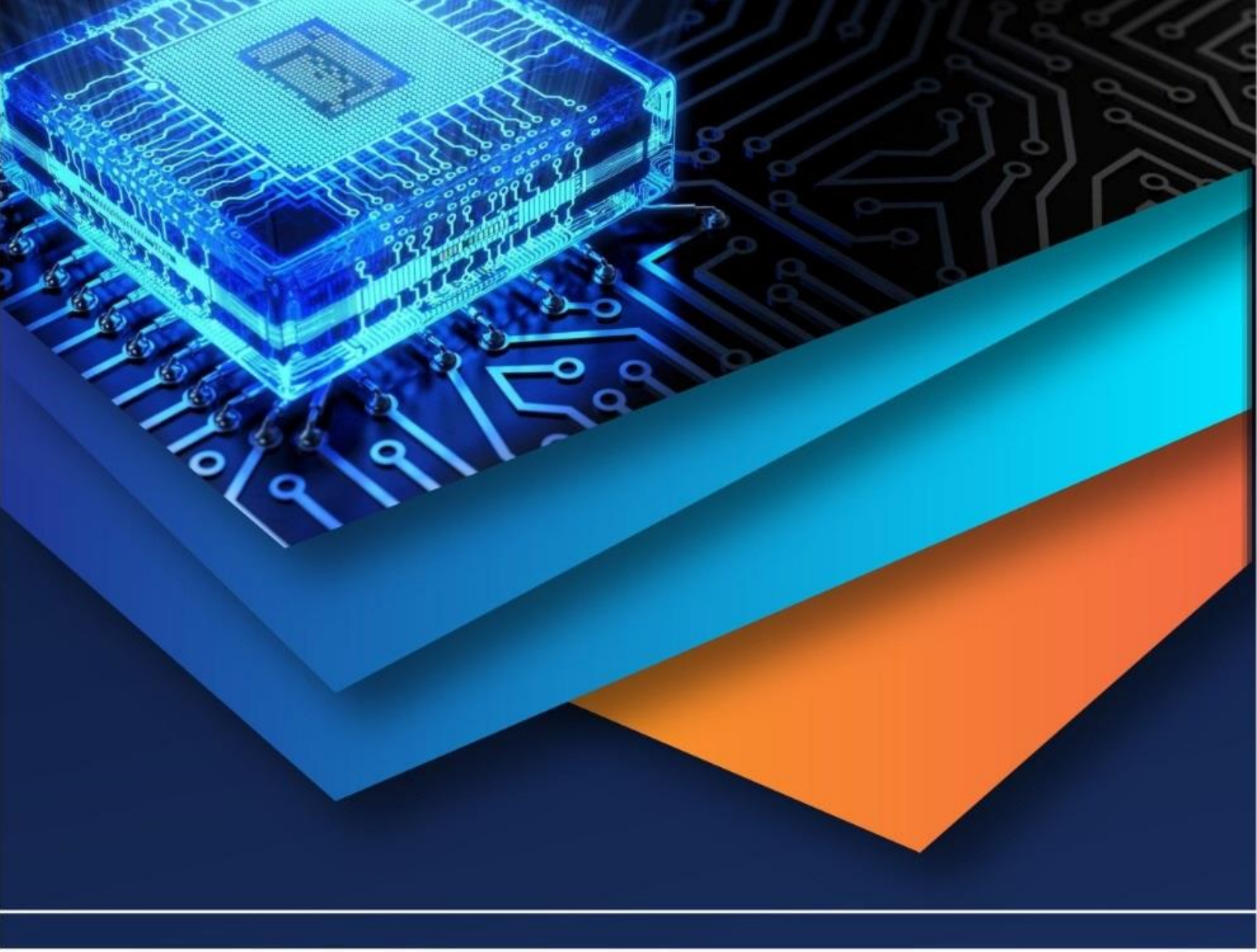
[13] Carrera, Enrique & González, Andrés & Carrera, Ricardo. (2017). Automated detection of diabetic retinopathy using SVM. 10.1109/INTERCON.2017.8079692.
https://www.researchgate.net/publication/317951200_Automated_detection_of_diabetic_retinopathy_using_SVM

[14] N.V. Ramana Murty1 and Prof. M.S. Prasad Babu2, 2017, “A Critical Study of Classification Algorithms for LungCancer Disease Detection and Diagnosis”, International Journal of Computational Intelligence Research, ISSN 0973-1873 Volume 13, Number 5 (2017), pp. 1041-1048.
https://pdfs.semanticscholar.org/8c7d/c550a87c3ac0c705eaadf62dd71f08f31aca.pdf

[15] Ted Bosworth, 2016, “Artificial intelligence beats standard algorithms for diagnosing lung disease”, Conference Coverage.
https://www.mdedge.com/chestphysician/article/113843/pulmonology/artificial-intelligence-beats-standard-algorithms

[16] Carson Lam, MD, Darvin Yi, Margaret Guo, and Tony Lindsey, PhD, “Automated Detection of Diabetic Retinopathy using Deep Learning”,
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5961805/

[17] Shaik Razia, P.Swathi Prathyusha, N. Vamsi Krishna, N. Sathya Sumana, 2017, “A REVIEW ON DISEASE DIAGNOSIS USING MACHINE LEARNING TECHNIQUES”, International Journal of Pure and Applied Mathematics, Volume 117 No. 16 2017, 79-85.
https://acadpubl.eu/jsi/2017-117-16/articles/16/10.pdf



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY