iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** III **Month of publication:** March 2023

DOI: https://doi.org/10.22214/ijraset.2023.49482

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 11 Issue III Mar 2023- Available at www.ijraset.com

Prediction of Chronic Diseases from X-Rays Using Distributed Machine Learning

Uddhav Kolhe[1], Omkar Janaskar[2], Diksha Sonavne[3], Yash Dubale[4], Dr. Kavita Wagh[5]
Department of Information Technology, Vasantdada Patil Pratishthan's College of Engineering, Sion, Mumbai, Maharashtra 400022

Abstract: ***There are various types of disease prediction models available in the market which predicts major chronic diseases but drawback is that each disease has a separate model which predicts and provides output to the user, there is no such model which will include multiple models into a single platform, for example heart disease has a separate model for prediction, covid-19 has a separate model which will provide result using machine learning algorithms, This article will briefly describe those drawbacks and will provide a proposed system for those issues.***
Keywords: Feature Extraction, Machine learning Models, diseases (lungs, Heart, diabetes) Clustering, Supervised Learning, Deep Learning, Diabetes, Heart, Random Forest, Boost.

I. INTRODUCTION

Nowadays, with the drastic change in technology, Healthcare Systems are majorly adopting those changes and becoming technologically sound and advanced with respect to data interpretation and data acquisition. From handling Patients data, to handling hospitals whole simulation AI and Machine Learning are inferencing in hospital management to make system more reliable and more efficient, Key benefits of using AI, Machine learning and Data Mining in Hospital management are as follows:

A. Electronic Medical Records

The goals of healthcare management and administration include fostering relationships between patients and clinicians, increasing transparency, lowering risk, and innovating healthcare. Electronic medical records, which address all of these issues as well as healthcare cost, value, and access, have been put into place thanks to technology. Clinical staff can make better diagnoses, data collection is more effective, and real-time data sharing increases patient satisfaction. Complete medical records are now readily available electronically. In the end, electronic medical records assist medical facilities in saving money by decreasing the time required for diagnosis and the need for additional testing that isn't necessary.

B. Customer Relationship Management Technology

The development of new customer relationship management (CRM) technology, for example, has significantly changed healthcare management. This new business technology was created specifically for medical use. By coordinating data, systems, and people across medical institutions, these healthcare software systems help to automate crucial business processes, cut down on waste, improve workflow, and enhance patient outcomes. These software programmes combine solutions for patient retention, compliance, customer service, claims processing, marketing, and more.

C. Proposing IT in Healthcare

Healthcare management is also undergoing constant change as a result of information technology integration. Consumer health IT applications, the use of electronic prescriptions, the implementation of computerised disease registries, support for clinical decisions, computerised provider order entries, and telehealth are a few examples of health information technology. For both specific patients and patient populations, new tools provide creative, exciting ways to manage information about healthcare and patients' health. As information technology is used more frequently in healthcare management, there are fewer medical errors, lower healthcare costs, and the ability to maintain more complete, accurate patient information, increasing the overall value of care.
Healthcare executives must learn how to use this technology effectively to make sure healthcare facilities are providing effective, high-quality healthcare delivery, as new, emerging technology is constantly reshaping the future of healthcare. Administrators must be ready to make the most of new medical technology while overcoming these new challenges because new technology, of course, also brings new obstacles to be overcome.

There are a few things that need to be taken into consideration when switching from a conventional procedure to a modern approach: challenges are nothing more than the concern of obstacles or issues that we faced during configurations with new technology or methodologies,

D. Serious Digital Risk

The current system for delivering healthcare is extremely complicated. It includes multiple layers of processes, a network of partners and patients, delivery models that are reimbursed by insurance, and regulatory requirements. Cyberattacks or data breaches could compromise both the security and privacy of the entire hospital.

E. Poor Training and Onboarding

There are numerous applications that are used in healthcare. Every day, a variety of applications are used by everyone from hospitals to manufacturers of medical equipment to ensure that operations run smoothly. Ineffective training or bad advice will reduce the effectiveness and accuracy of a system or model.

F. Strict Compliance Regulations and Data Integrity

Any organization involved in the healthcare sector is required to abide by all applicable laws and regulations. Administrative errors account for 86% of errors made in the healthcare industry and are the third leading cause of death (behind heart disease and cancer). Employees must enter data into applications accurately and follow procedures precisely to prevent errors.

II. METHODOLOGY

It is a framework of procedures described in Block Diagram which proposed system follows to find accurate results with definite values.

Fig: System Flow Diagram

1) *Data Collection:* According to the model Data is collected from various sources like kaggle, Internet to apply algorithms in dataset.
2) *Training Data and Test Data:* The collected data is distributed between training data and testing data.
3) *Training Data:* Training data is use to train the model for prediction and accuracy, the data assign to training is 80% of total dataset.
4) *Testing Data:* After Training the model test data is used to test the model for accuracy and prediction.
5) *Data preprocessing:* In this method, data is cleaned in different methods like Binning, Regression as well as Clustering, this processes are used to smoothen the dataset for analysis and prediction purpose.

III. RELATED WORK & ALGORITHM ANALYSIS

A. Diabetes

In this project we will determining Diabetes disease base on some input values which will take as an input to the prediction model and it will provide result based on its analysis, We used Random Forest Algorithm for prediction where we used following factors which is used to train the model with the help of algorithm.

- *Gender*: Describe Gender of the patient.
- *Glucose* concentration: Defines amount of glucose person.
- *Blood pressure*: This is the conditional input and based on this result may vary, According to the studies, Blood pressure should be below 140/80mmHg for people with diabetes or below 130/80mmHg if you have kidney or eye disease or any condition that affects blood vessels and blood supply to the brain. But it is important to speak to your healthcare team about your target.
- *Insulin levels*: this one is too a conditional input value whereas study defines a blood sugar level of less than 140 mg/dL (7.8 mmol/L) is normal., above this specified value can become critical BMI, diabetes pedigree: This values should be between 25 to 30, increased BMI value can become cause of it.
- *Age*: Defines how old you are which is also matters during the time of model prediction Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

	A	B	C	D	E	F	G	H	I
1	Pregnancie	Glucose	BloodPres	SkinThickn	Insulin	BMI	DiabetesPe	Age	Outcome
2	6	148	72	35	0	33.6	0.627	50	1
3	1	85	66	29	0	26.6	0.351	31	0
4	8	183	64	0	0	23.3	0.672	32	1
5	1	89	66	23	94	28.1	0.167	21	0
6	0	137	40	35	168	43.1	2.288	33	1
7	5	116	74	0	0	25.6	0.201	30	0
8	3	78	50	32	88	31	0.248	26	1
9	10	115	0	0	0	35.3	0.134	29	0
10	2	197	70	45	543	30.5	0.158	53	1
11	8	125	96	0	0	0	0.232	54	1
12	4	110	92	0	0	37.6	0.191	30	0
13	10	168	74	0	0	38	0.537	34	1
14	10	139	80	0	0	27.1	1.441	57	0
15	1	189	60	23	846	30.1	0.398	59	1
16	5	166	72	19	175	25.8	0.587	51	1
17	7	100	0	0	0	30	0.484	32	1
18	0	118	84	47	230	45.8	0.551	31	1
19	7	107	74	0	0	29.6	0.254	31	1
20	1	103	30	38	83	43.3	0.183	33	0
21	1	115	70	30	96	34.6	0.529	32	1
22	3	126	88	41	235	39.3	0.704	27	0
23	8	99	84	0	0	35.4	0.388	50	0
24	7	196	90	0	0	39.8	0.451	41	1
25	9	119	80	35	0	29	0.263	29	1
26	11	143	94	33	146	36.6	0.254	51	1

1) Random Forest Algorithm

The Random Forest Algorithm's ability to handle data sets with both continuous variables, as in regression, and categorical variables, as in classification, is one of its most major elements. In terms of classification problems, it delivers better results. It is based on the idea of ensemble learning, which is a method of combining various classifiers to address complex issues and enhance model performance. Random Forest, as the name implies, is a classifier that uses a number of decision trees on different subsets of the given dataset and averages them to increase the dataset's predictive accuracy.

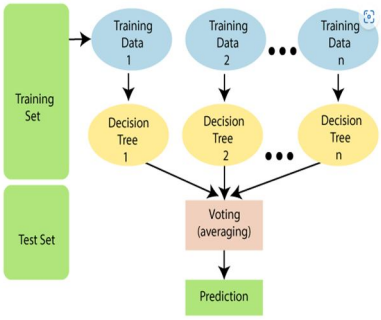


Fig: Train and Test Data Flow

The Working process can be explained in the below steps and diagram:

- Step-1: Select random K data points from the training set.
- Step-2: Build the decision trees associated with the selected data points (Subsets).
- Step-3: Choose the number N for decision trees that you want to build.
- Step-4: Repeat Step 1 & 2.
- Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

B. Heart Disease

Heart Disease or Cardiovascular Disease is one of the severe disease, so as the prediction should be accurate as well as distinct because diseases severity and affect, Here we use XGBOOST algorithm due to its accuracy and prediction rate Following attributes needs to considered for model prediction.

- Old peak
- Max heart rate achieved if exercise induces angina
- Number of blood vessels
- Type of chest pain
- Age

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
2	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
3	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
4	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
5	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
6	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
7	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
8	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
9	44	1	1	120	263	0	1	173	0	0	2	0	3	1
10	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
11	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
12	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
13	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
14	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
15	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1

1) XGBOOST Algorithm

Large dataset performance, usability, and speed are all priorities in the design of XGBoost. It does not require parameter optimization or tuning, so it can be used right away after installation with no additional configuration. The gradient boosted trees algorithm is implemented using the open-source software known as XGBoost, which stands for extreme gradient boosting. Decision trees are used as the algorithm's "weak" predictors in a gradient boosting scheme. In addition, its implementation was specially designed for the best speed and performance.

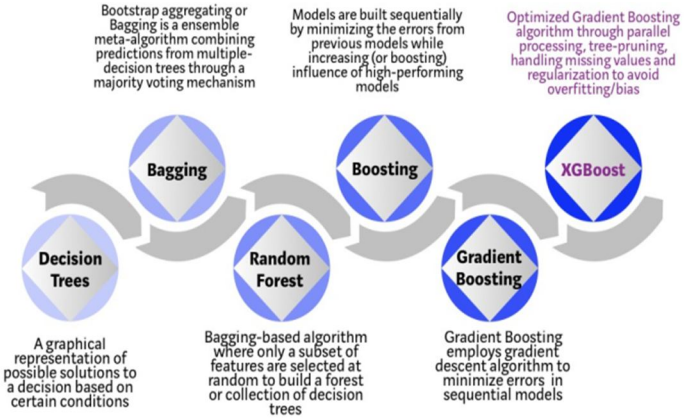
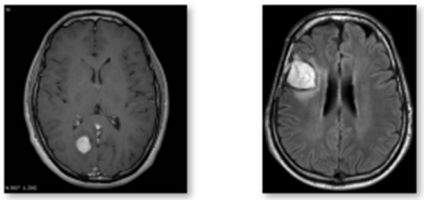


Fig: Working of XGBOOST

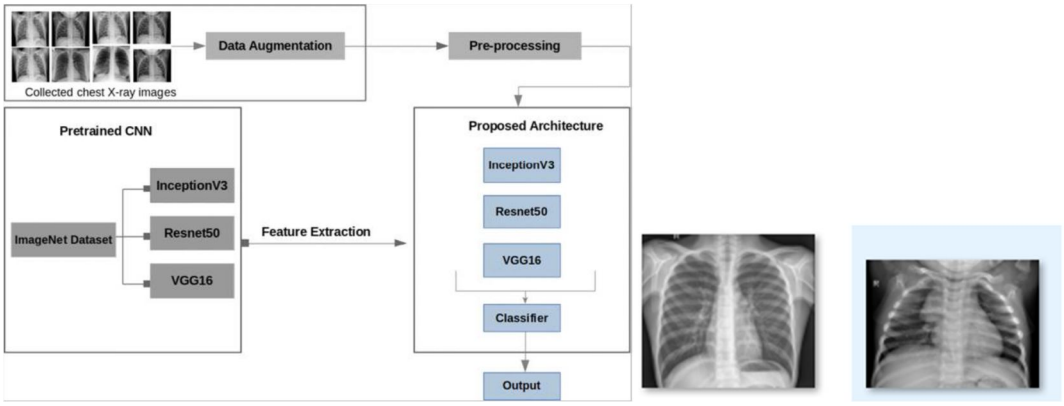
2) Brain Tumor

A brain tumour is a growth of abnormal cells in your brain that is similar to a heart disease in that it is a severe disease. There are numerous varieties of brain tumours. Both benign (noncancerous) and malignant (cancerous) brain tumours can occur (malignant). Primary brain tumours are those that start in the brain; secondary (metastatic) brain tumours are those that start in other parts of the body and spread to the brain. A brain tumor's rate of growth can vary significantly. Your nervous system's ability to function depends on the growth rate and location of a brain tumour. The type of brain tumour you have, as well as its size and location, all affect your treatment options. We use the CNN algorithm for prediction, and the user must supply an MRI of the brain for evaluation and prediction.



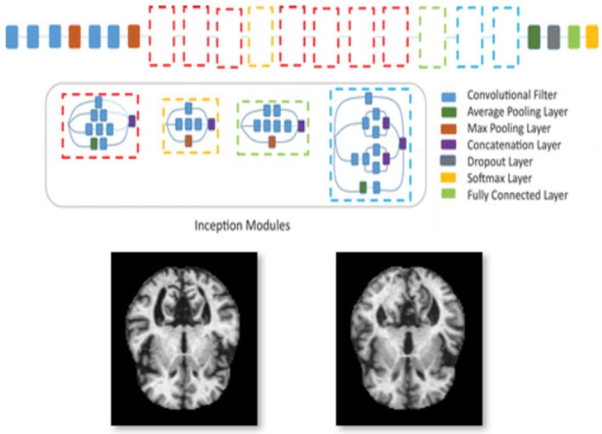
C. Covid-19

The SARS-CoV-2 virus is the infectious disease known as coronavirus disease (COVID-19). The majority of virus-infected individuals will experience a mild to moderate disease and will recover without the need for special care. However, some people will develop serious illnesses and need to see a doctor. Serious illness is more likely to strike older people and those with underlying medical conditions like cancer, diabetes, cardiovascular disease, or chronic respiratory diseases. COVID-19 can cause anyone to become seriously ill or pass away at any age. operating covid-19 When an infected person coughs, sneezes, speaks, sings, or breathes, the virus can spread from their mouth or nose in minute liquid particles. From larger respiratory droplets to smaller aerosols, these particles are diverse. The CNN algorithm is used for Covid-19 model prediction, and users must upload their chest X-rays or scans for analysis. Working of covid-19-



D. Alzheimer

A brain disorder that worsens over time is Alzheimer's disease. Changes in the brain that result in protein deposits are what define it. The brain shrinks as a result of Alzheimer's disease, and eventually, brain cells pass away. The most frequent cause of dementia, which is characterised by a slow loss of memory, thinking, behaviour, and social abilities, is Alzheimer's disease. The ability to function is impacted by these modifications. The disease's early symptoms include forgetting recent conversations or events. It eventually leads to severe memory issues and a loss of the ability to carry out daily tasks. The symptoms' progression may be slowed or improved by medications. Programs and services can support those who have the disease and those who care for them. In this case, we use the CNN algorithm to train the model in accordance with the MRI scan that the user provided.

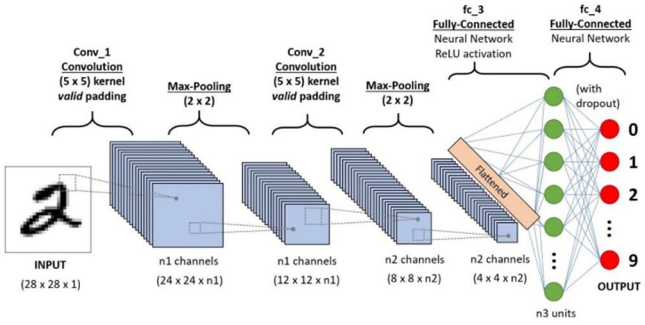


3) Convolutional Neural Network (CNN)

A deep learning algorithm known as a convolutional neural network (CNN) is particularly effective at processing and recognising images. Convolutional layers, pooling layers, and fully connected layers are among the layers that make up this structure.

The central part of a CNN is its convolutional layers, where filters are used to extract features like edges, textures, and shapes from the input image. The output of the convolutional layers is then passed through pooling layers, which are employed to down-sample the feature maps and retain the most crucial data while lowering the spatial dimensions. One or more fully connected layers are then applied to the output of the pooling layers in order to predict or categorise the image.

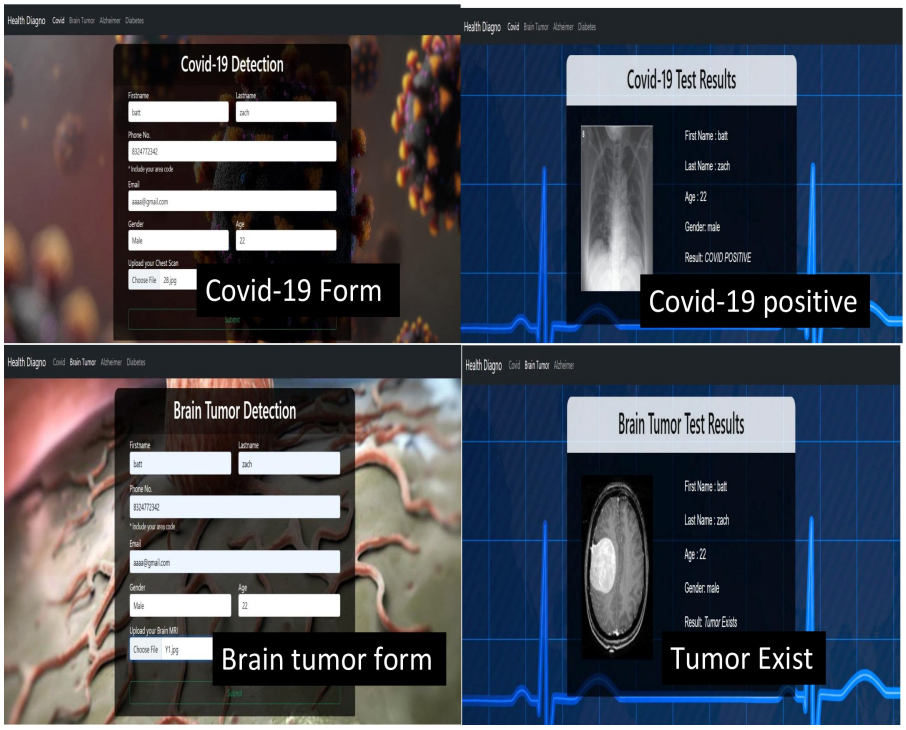
With the help of a sizable dataset of labelled images, CNNs are trained to identify patterns and features that are connected to particular objects or classes. A CNN can be trained to classify new images and can also be used to extract features for other tasks like object detection or image segmentation.

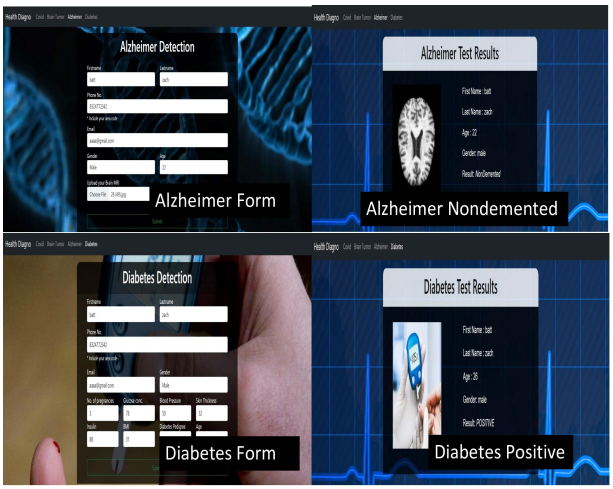


IV. RESULTS

In this system numerous types of diseases are analyzed and predicted according to the input values provided by the user. We use multiple algorithms like XGBOOST, Random Forest, CNN and other algorithms to train that specific disease model to provide accurate as well as appropriate results with efficiency and ease.

Following are some results of diseases with their labelling:





V. CONCLUSION

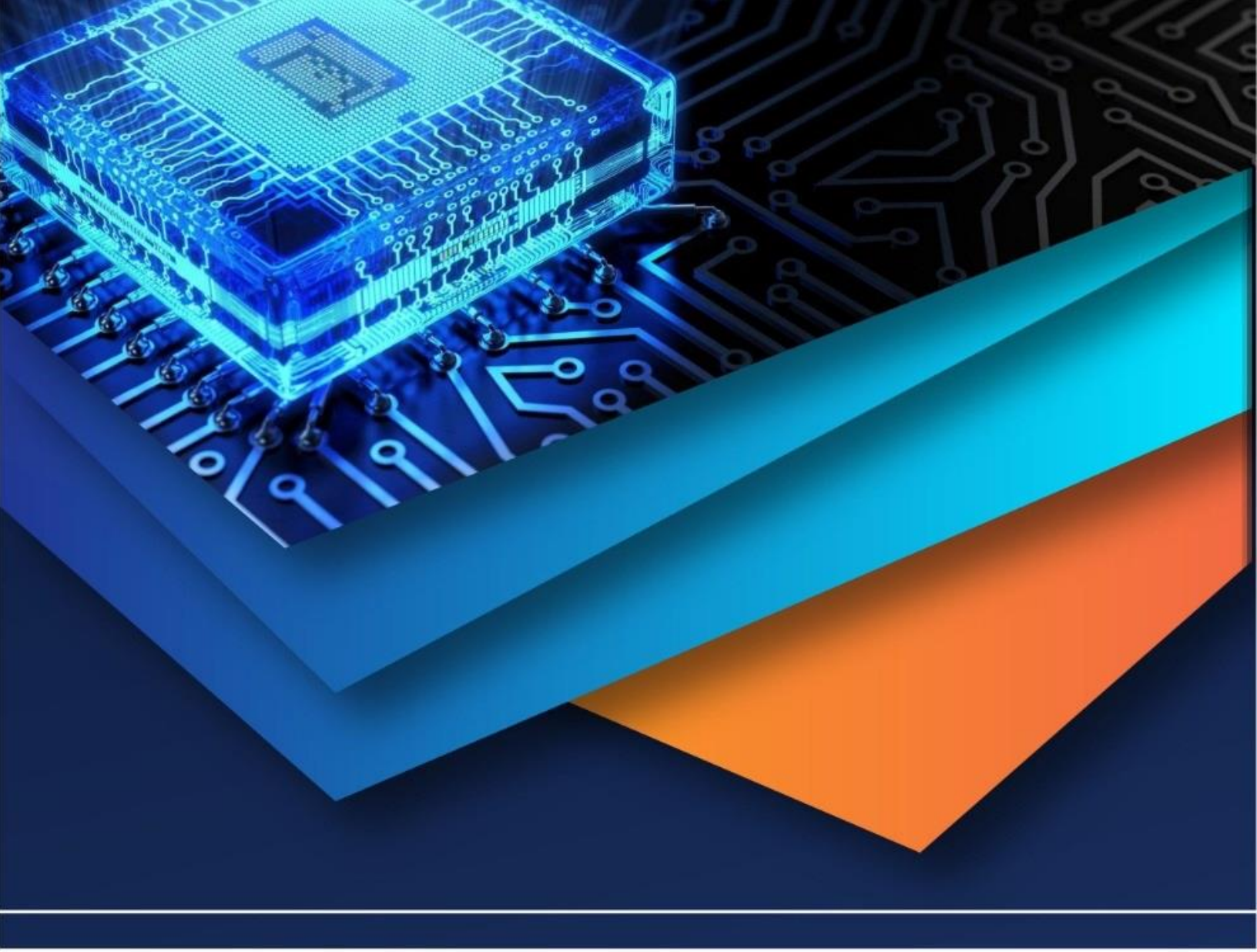
A multiple diseases prediction model is used to predict several diseases simultaneously. Here, a disease prediction is based on user input. The user will be given a choice. A corresponding disease model will be activated and predicted based on the user's x-ray input if they wish to predict a specific disease. The advantage of using a multi-disease prediction model in advance is that it can estimate the likelihood that numerous diseases will manifest themselves while also lowering the mortality rate.

Accuracy Comparison between Health Diagno & Previous Projects:

Health Diagno Diseases	Algorithm	Accuracy (%)	Previous Projects /Papers			
			Paper	Author	Algorithm	Accuracy
Covid19	Convolutional Neural Network(CNN)	93%	Covid19 Prediction	Auburn, Amman	CNN	92%
Diabetes Detection	Random Forest	90%	Diabetes Type-2 Prediction	Asif Hassan Syad,Trej Khan	Pearson's Square Test,Binary Logistic Regression	82%
Breast Cancer	Random Forest	91.81%	Breast Cancer	Pahulpreet Singh	SVM, Decision Tree	78%
Heart Disease	XGBoost	86.96%	Heart Disease	Pahulpreet Singh	SVM, Decision Tree	78%
Brain Tumour Detection	VGG-16,CNN	99%	Brain Tumour Detection using Machine Learning	Manav Sharma ,Pramanshu Sharma,Ritika Mittal,Kamakshi Gupta	CNN	97.79%
Alzheimer Detection	Trained using CNN	73.54%	Alzheimer Detection Early stage Detection Using Machine Learning Techniques	Roobaea Aloobaea,Seifeddine Mechti,Sudhakar Sengan	Logistic Regression,Spport Vector, Random Forest	62.13%
Pneumonia Detection	Convolutional Neural Network	83.17%	Pneumonia Detection Using CNN	V.Sirish Kaushik,Anand Nayyar,Gaurav Kataria,Rachna Jain.	CNN	76.67%

REFERENCES

[1] Y. Chen, H. Jiang, C. Li, X. Jia and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on Convolutional Neural Networks," IEEE Transactions on Geoscience and Remote Sensing, pp. 6232-6251, 2016.
[2] Yaguchi, Y., Omura, M., & Okumura, T. (2017). Geometrical mapping of diseases with calculated similarity measure.
[3] Abebe Tadesse, G., Javed, H., Thanh, N. L. N., Ha Thai, H. D., Le Van, T., Thwaites, L., … Zhu, T. (2020). Multi-modal Diagnosis of Infectious Diseases in the Developing World.
[4] P. Gokila Brindha, M. Kavinraj, P. Manivasakam and P. Prasanth, "Brain tumor detection from MRI images using Deep Learning Techniques," IOP Conference Series: Materials Science and Engineering, p. 012115, 2021.
[5] Jo, T., Nho, K., and Saykin, A. J. (2019). Deep learning in alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. Frontiers in aging neuroscience, 11:220.
[6] Kumar, U. (2019). Applications of machine learning in disease pre-screening. In Pre-Screening Systems for Early Disease Prediction, Detection, and Prevention, pages 278–320. IGI Global.
[7] Chaitrali S. Dangare, Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications (0975 888)Volume 47No.10, June 2012.
[8] Amudhavel, J., Inbavalli, P., Bhuvaneswari, B., Anandaraj, B., Vengattaraman, T., Premkumar, K., "An effective analysis on harmony search optimization approaches", (2015) International Journal of Applied Engineering Research, 10 (3), pp. 2035-2038. [9] Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients Mai Shouman, Tim Turner, and Rob Stocker International Journal of Information and Education Technology, Vol. 2, No. 3, June 2012.
[9] Valensi P.,Prévost G.,Schnell O.,Standl E.,Ceriello A. Targets for blood glucose: what have the trials told us.Eur J Prev Cardiol. 2019; 26: 64-
[10] Ed-daoudy, A., Maalmi, K., "Application of machine learning model on streaming health data event in real-time to predict health status using spark" In: 2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT), IEEE, 2018 1-4.
[11] K. Lee, A. Ankit, C. Alok,"Real-time disease surveillance using twitter data: demonstration on flu and cancer," In: Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining, 2013, pp. 1474-1477.
[12] Available from: https://archive.ics.uci.edu/ml/datasets/heart+Disease Online, accessed December 2017.
[13] Available from http://cassandra.apache.org Online, accessed December 2017
[14] Y. H. Hu, W. J. Tompkins, J. L. Urrusti, and V. X. Afonso, "Applications of artificial neural networks for ECG signal detection and classification," J. Electrocardiol, pp. 66–73, 1994.
[15] S. Osowski, L. T. Hoai, and T. Markiewicz, "Support vector machine based expert system for reliable heartbeat recognition," IEEE transaction biomedical engineering, vol. 51, no. 4, pp. 582–589, Apr. 2004.
[16] Shin-Chi Lai, Chien-Sheng Lan and Sheau-Fang Lei, "An Efficient Method of ECG Signal Compression by Using a DCT," International Conference on Communications, Circuits and Systems (ICCCAS),vol.1, no.46-49,2013.
[17] Muhammad Zubair, jinsul kim, changwoo yoon, "An Automated ECG beat classification system using convolutional neural network," International Conference on IT Convergence and Security (ICITCS), pp. 1-5,2016.
[18] Serkan Kiranyaz, Turker Ince, Moncef Gabbouj,"Real-time patient specific ECG classification using 1D convolutional neural network",IEEE transaction biomedical engineering,vol.63,pp.664-675,2016.P



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY