iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XII **Month of publication:** December 2024

DOI: https://doi.org/10.22214/ijraset.2024.66139

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

Cancer Detection Using Machine Learning

Ms. Poonam Chakravarty[1], Dhara Pandya[2], Mr. Jigar Pandya[3]
[1]*Assistant Professor, CSE/IT Department, Rai University, Ahmedabad*
[2]*B. Tech Student, CSE/IT Department, Rai University, Ahmedabad*
[3]*Assistant Professor, CSA Department, Rai University, Ahmedabad*

Abstract: Cancer has been classified into various subtypes. So, it is essential to detect cancer symptoms early on. Artificial Intelligence(AI) and Machine Learning(ML) have been used to classify cancer categories. There are certain datasets and models capable of estimating key features. There are many methods used for the development of cancer detection techniques, some of the methods are artificial neural networks(ANNs), and decision trees(DTs). While considering these methods and understanding them, a valid procedure must be considered for their implementation. In this study, various ML and Deep Learning(DL) approaches are explained which can be used in cancer detection.

Although there has been progress in the diagnosis and treating cancer victims with personalization, it is difficult to provide cancer victims with data-driven care. To improve patient outcomes and medical efficiency, the application of Artificial Intelligence(AI) has become an effective means.

Machine Learning provides an opportunity for systems to learn by gaining knowledge from learning models, and this approach is very successful at forecasting different forms of cancer, among them related to liver, lung, and other cancers. Professionals are not as precise in forecasting illness as machine learning and artificial intelligence are.

The recent advancement in deep learning has transformed medical imaging, providing tools for analyzing data automatically.

Convolutional neural networks, also known as CNNs, have been employed to identify several tumor categories such as MRI and CT scans. In this study, we shall learn about a structural framework that includes data preprocessing, model architecture design, and performance analysis.

Deep learning models can achieve higher accuracy and sensitivity for the identification of symptoms related to cancer with an efficient approach compared to traditional practices of treatment.

Keywords: Cancer detection, ML, Deep Learning, AI, Decision trees(DT), Artificial Neural Network, selection of treatment, cancer prediction and treatment

I. INTRODUCTION

Cancer has been a significant health issue globally, with increasing survival rates. Cancer is an illness in which body cells start growing and spread to the other body's parts. When the body's cells become ruptured, this process breaks down, and these ruptured cells start growing and multiply, leading to the formation of tumors. These are tissue lumps.

Tumors containing cancerous cells disperse and infect surrounding tissues and transfer farther away to develop into additional tumors. Metastasis is the spreading of cancer cells from one part to another of the body. Malignant tumors affect organs or tissues with rapid growth. On the other side, benign tumors don't migrate, and benign tumors do not grow back, when benign tumors start spreading in the brain, they can cause concerning issues.[1]

A. *Cancer Types*

Different types of cancer, more than 100 types are present. They have been named based on the organs and tissues of origin.[2]
While lung cancer develops in the lungs, brain cancer develops in the brain.
Here are a few categories:

1) *Carcinoma*

The most common one is Carcinoma. With the formation of epithelial cells, they build in different types of epithelial cells.[2]
Adenocarcinoma is a type of cancer that initiates forming in epithelial cells generated by mucus or fluids. Most cancers related to the colon, prostate, and breast, are adenocarcinomas. Transitional cell carcinoma belongs to a category of epithelial tissue known as transitional epithelium, or urothelium.[3] This tissue, consisting of multiple layers of flexible epithelial cells, is found in the linings of the renal pelvis, part of the kidney. It is also present in the ureters and bladder. Some bladder, ureter, and kidney cancers are carcinoma transition cells.

2) Sarcoma

Sarcomas is a category of cancer that forms in bone and soft tissues, including blood vessels, fat, and muscles.[3] Osteosarcoma is the most common bone cancer.

3) Leukemia

Cancers that start building in the tissue with blood formation are known as leukemias, which are present in bone marrow. No formation of tumors and large white blood cells with an abnormality can build up in the bone marrow and inside the blood.[2]

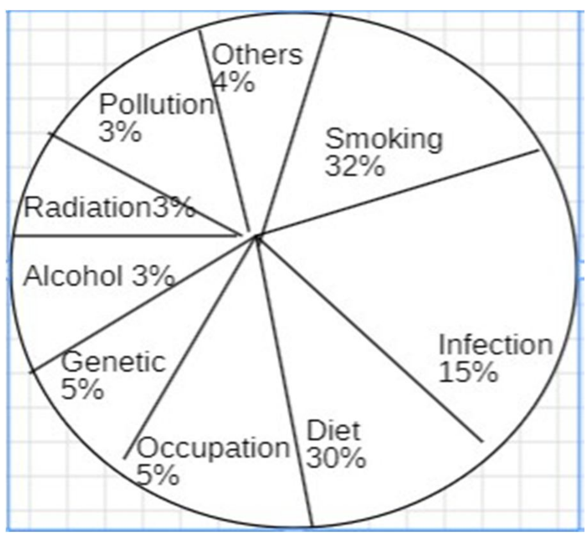
B. Brain and Spinal Cord Tumors

Many types of brain and spinal tumors are present. These are named on the type of cell in which they are formed. Example: An astrocytic tumor begins in star-shaped brain cells known as astrocytes. Brain tumors can be benign or malignant(cancer).
As of 2024, cancer continues to be a global health issue with an impactful and growing significance. In the year 2022, approximately 20 million new cases of cancer worldwide were reported. Lung cancer remains the most common cause of cancer-related mortality, with an approximate number of 1.8 million fatalities. Approximately 27 percent of all deaths related to cancer are attributed to cancer of the lungs.

C. Reasons

Normal cells undergo a multi-stage process that starts a cancerous lesion, after the growth of cancer, it might end up with a malignant tumor. Genetics play an important role and external agents also add up: Physical carcinogens, such as ultraviolet and radiation. Chemical carcinogens, such as components of tobacco smoke, alcohol Biological carcinogens, such as infections from certain viruses, bacteria, or parasites.
Following is the World Cancer Report by IARC,2008.

Figure 1



II. LITERATURE REVIEW

The literature review critically examines works on Artificial Intelligence(AI), Machine Learning (ML), and Deep Learning(DL), how AI is integrated with current medical facilities, usage of Machine Learning(ML) in Cancer Prediction, and Predicting cancer survival.

A. An Outline Of Artificial Intelligence

Artificial Intelligence is defined as computer-coded programs or algorithms that use data analysis and pre-programmed instructions to make predictions and decisions, thus detecting types of diseases.
Artificial Intelligence(AI) is a set of technologies that empower computers to perform different advanced operations, including the ability to see, understand translation, analyze data, and make recommendations.

1) How does AI work?

The core principle revolves around data, although there are many different AI techniques.

Systems of AI learn and improve with more exposure to data, and identifying patterns.

This learning process often involves algorithms, which are a set of instructions that direct AI's analysis and decision-making. Within machine learning, a subset of AI, algorithms are taught for making predictions and categorizing information.[5]

Deep learning is a specific form of Artificial intelligence (AI) that utilizes artificial neural networks to process information using multiple layers, coping with the structure and function of the human brain. AI systems can adapt to perform specific tasks through continuous learning and adaptation.

B. Models Of Training In Artificial Intelligence Training Models

Machine Learning is a subset of artificial intelligence which uses algorithms to train data to obtain specific outcomes.

There are three kinds of learning models used in machine learning:

1) Supervised Learning

This is a machine-learning model which has a process to associate a specific input with output-trained data. For instance, to train the algorithm to recognize pictures of cats, that have been labeled as cats.[6]

2) Unsupervised Learning:

It is machine learning that learns patterns based on unlabeled data. The result is not known. It is good at matching patterns along with descriptive modeling.

3) Reinforcement Learning

This is a machine-learning model. In this, an agent learns to perform a defined task by trial and error. Example: Learning would be teaching a robotic hand to pick a ball.[5]

4) Types of artificial neural networks

One of the most commonly used training models in AI is the artificial neural network, a model quite similar to the human brain.

An artificial neural network is a system of artificial neurons, called perceptrons, which are computational nodes to classify and analyze data. There are different layers, the first layer takes the data, and the perceptron makes decisions before passing the data to another layer. Deep neural networks, or deep learning, are training models with more than three layers This final layer is responsible for classifying an object and finding patterns.

Below are common categories of artificial neural networks:

- *Feedforward Neural Networks (FF)*

One of the oldest forms of neural networks involves the flow of data in one way within layers of artificial neurons for the output to be achieved. In modern days, most feedforward neural networks are known as deep feedforward with multi-layers.[6]

- *Recurrent Neural Networks (RNN)*

It is different from feedforward neural networks as it uses time series data or data involving sequences. It might use weights in each node of the network, these networks have memory of what happened in the previous layer. Example: While performing natural language processing, RNNs can remember words used in a sentence. These are used for speech recognition, translation, and captioning images.

- *Convolutional Neural Network (CNN)*

These include some of the most common neural networks in modern AI. For image recognition, CNNs use different layers(a convolutional layer and a pooling layer) to filter different parts of an image before putting it together. Firstly they look for simple features of an image, like colors and edges, and later look for complex features in additional layers.[5]

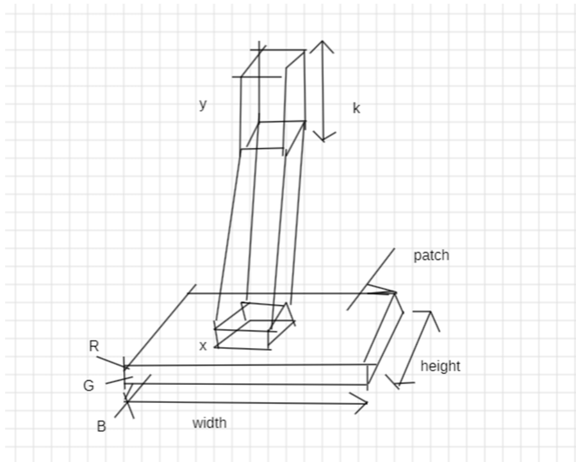
5) How the Convolutional layer works

Convolution Neural Networks are neural networks that share parameters. Example: Consider an image, which can be represented as a cuboid having its length, width (dimension of the image), and height(channels generally have red, green, and blue channels).

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

Imagine taking a small patch of this image and running a neural network, known as a filter or kernel, to it. The neural network will take K outputs and represent them vertically. Now sliding the neural network across the image, we will get another image with different widths, heights, and depths. This process is called Convolution.[6]

Figure 2
The convolutional process has been explained in the following image from Deep Learning Udacity.

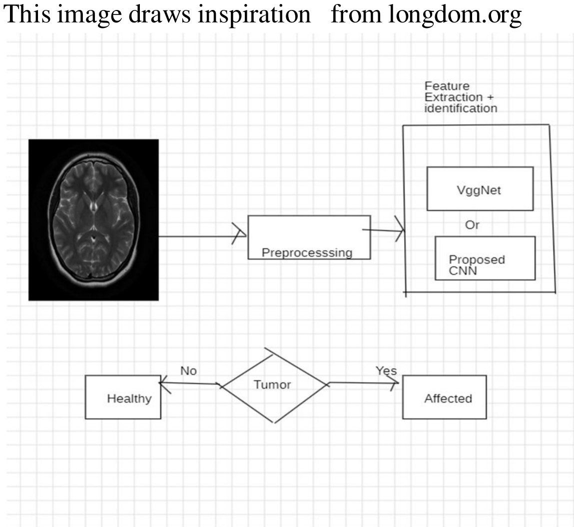


C. *Introduction To Machine Learning*

It is a branch of Artificial Intelligence, involving learning from data samples. It comprises two phases: i) making predictions about new output using estimated dependencies and ii) calculating unknown relationships in a system with a lot of data. There are two most prominent types of ML methods, known as i) supervised learning in which labeled training data is used to map input data for the desired output. unsupervised learning In unsupervised learning, which operates without labeled data there is no option for output. Thus, the learning scheme/model is responsible for finding patterns or discovering groups of input data. Regression and clustering are important tasks of ML. A learning function transforms data into real-value parameters for regression problems. The process can be utilized for calculating the value of the predicted sample for each new sample.[9][10][11]

Searching categories or clusters to describe data is a common unsupervised task called clustering. Each fresh sample is assigned to one of the discovered groups with shared characteristics using the method. Using data related to brain tumors, we attempt to determine if a tumor is malignant or not(labels 1 for Yes and 0 for No). The image illustrates the processing of data from MRI datasets of brain tumors and the process of identification and extraction of features, Yes indicates a healthy MRI and No is an affected image.

Figure 3
This image draws inspiration from longdom.org



While applying ML methods, data samples contribute to basic components. Each feature describes a specific value and feature. It is better to understand the quality of data in advance to make the right choice for data selection and analysis techniques.

There are some data processes required to make suitable datasets for ML using data preprocessing.

Problems with data quality include outliers, duplicated data, or missing data and noise; therefore the quality of data is better, then the resultant data analysis is also improved.

There are several techniques to convert raw data for analysis that concentrate on changing the data to better arrange the machine learning algorithms.

This approach includes: i) dimensionality reduction ii) feature extraction and feature selection. Lower dataset dimensionality improves the performance of the machine-learning algorithms.

Reducing dimensionality can deduce unrelated characteristics, reducing noise and creating learning models that are more efficient. In feature extraction, a new set of features is created from the initial set which has captured important information from a dataset. This new creation allows gathering information from dimensionality reduction.

ML methods create a model used for classification, prediction, and estimation. Classification is the most common learning process. When ML methods are used to develop a classification model, training and generalization errors might be produced.

Example: For brain tumor detection, the classification step includes a CNN model, to classify the MRI brain scans into two classes- tumor and non-tumor.

Estimating the classifier's performance is crucial after ML techniques have produced a classification model. Sensitivity, area under the curve(AUC), accuracy, and specificity, are used to measure the performance.

Testing sets are used to estimate generalized errors, which is a measure of the model's prediction accuracy.

Accuracy is one metric for estimating classification models. The accuracy of the model can be easily calculated by the confusion matrix.[7]

Consider the formula: Accuracy=(TP+TN)/(TP+TN+FP+FN)

Where,

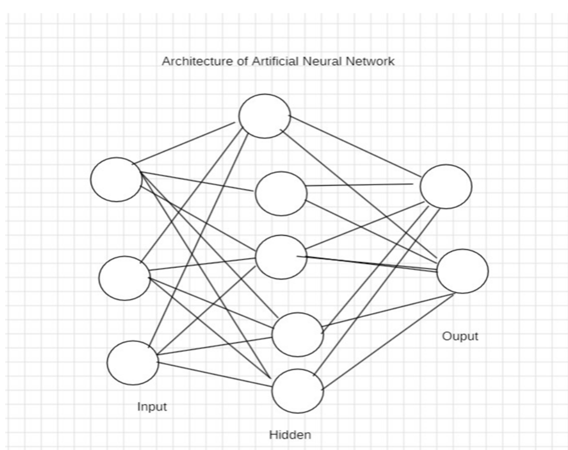
TP=True Positive FP=False Positive TN=True Negative FN=False Negative

Testing and training samples must be sufficient, large, and independent for the testing to be known, for obtaining reliable results, and for predicting the performance of a model's performance. Data processing involves different learning tasks. There are SVMs, BNs, ANNs, and DTs available. We shall understand ML techniques employed in studies for the detection of cancer using ML approaches, depending on the intention of this investigation. In this study, different classification techniques are examined, with the utilization of feature subsets such as SVM classifiers, and K-nearest neighbors neural networks.

ANN(Artificial Neural Networks) is the most widely used and reliable approach to make real-time predictions and forecasts. For various pattern recognition tasks, ANNs excel at classification. As a result, they are trained to produce an output in addition to specific factors. Utilizing several layers of hidden information is time-consuming and it uses interconnected nodes to generate classification processes.

Figure 4

The image draws inspiration from knoldus.com and illustrates the Architecture of Artificial Neural Networks.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

Decision Tree (DT) categorization is based on tree structure, with nodes representing input variables and leaves representing outcomes. DTs are quick to learn and easy to understand.

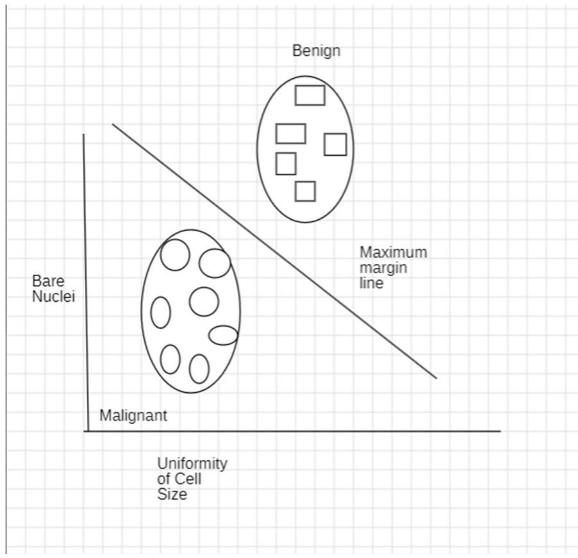
The particular architecture of DTs helps in reasonable outcomes, which makes it an appealing technique.

Support vector Machine(SVM) maps the input into a feature space of higher dimensionality. In order to classify new datasets with reliability, the resulting classifier must achieve generalizability. This illustration shows how SVM distinguishes between malignant and benign tumors, which is taken from researchgate.net, the classification is based on cell size.[12]

The margin line can be seen as a decision boundary separating two clusters.

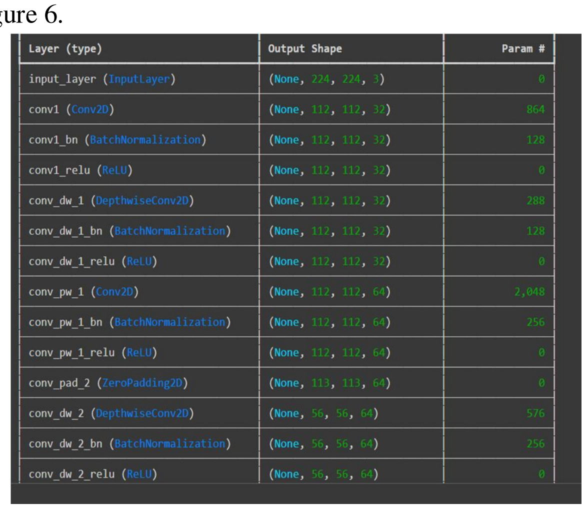
Figure 5

A simplified representation of SVM



BN stands for batch normalization. Instead of making predictions its classifiers produce probability estimates. They are used for classification tasks and also for knowledge representation and reasoning application.

Example: Consider the Resnet50 model to train the model. Microsoft introduced a residual neural network in 2015. Convolutional Neural Networks(CNN) are composed of three layers which are fully connected, convolutional layers, and pooling layers. The residual network is made of 6 types of layers- input layers, padding layer, convolution layer, batch normalization, activation(ReLu), and max pooling. Apart from these layers, additional layers are added to the model, which are Average Pooling, Flatten, Dense, and Dropout. The result is illustrated in Figure 6.



Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 224, 224, 3)	0
conv1 (Conv2D)	(None, 112, 112, 32)	864
conv1_bn (BatchNormalization)	(None, 112, 112, 32)	128
conv1_relu (ReLU)	(None, 112, 112, 32)	0
conv_dw_1 (DepthwiseConv2D)	(None, 112, 112, 32)	288
conv_dw_1_bn (BatchNormalization)	(None, 112, 112, 32)	128
conv_dw_1_relu (ReLU)	(None, 112, 112, 32)	0
conv_pw_1 (Conv2D)	(None, 112, 112, 64)	2,048
conv_pw_1_bn (BatchNormalization)	(None, 112, 112, 64)	256
conv_pw_1_relu (ReLU)	(None, 112, 112, 64)	0
conv_pad_2 (ZeroPadding2D)	(None, 113, 113, 64)	0
conv_dw_2 (DepthwiseConv2D)	(None, 56, 56, 64)	576
conv_dw_2_bn (BatchNormalization)	(None, 56, 56, 64)	256
conv_dw_2_relu (ReLU)	(None, 56, 56, 64)	0

D. Integration of AI with Existing Medical Facilities Is Crucial For Advancement

Artificial intelligence is utilized in healthcare and demonstrates the use of software for understanding and improving health and medical care data. AI applications are used to examine associations between patient outcomes and the healthcare system. The approach of Artificial intelligence differs from traditional practices as it involves the collection of data, processing and interpreting it, and estimating a predicted output. Machine learning and deep learning techniques are utilized to achieve this process as described above. This can be accomplished by identifying patterns and generating reasoning based on patterns.

Currently, AI-based information is used in medical advancements and development, customized medicine, and personalized healthcare, including monitoring of patients.

AI has represented great potential in clinical and laboratory protocols of infectious diseases and developing medicine.

Using samples from mass spectrometry, AI-based neural networks are introduced to accurately predict a host response to coronavirus.

Other uses of AI include support vector machine(SVM) for ML investigation of blood samples for malaria treatment and enhanced healthcare of treatment for diagnosis of Lyme disease.

The use of AI extends to the treatment of tuberculosis, meningitis, and various illnesses like hepatitis.

E. Machine Learning Applications For Cancer Prediction And Forecasting Cancer Survival

Early detection of cancer is very crucial for improving healthcare facilities. It is more challenging to treat cancer in the advanced stages as compared to the initial stages. ML algorithms have immense potential for improving healthcare's accuracy, personalization, and speed. ML Algorithms analyze imaging data to find trends that point to cancer in its early stages.

ML can classify different types of cancer cells helping in determining appropriate treatment approaches. ML algorithms can extract features from medical images that may not be visible to the human eye. It is important to improve current Artificial Intelligence(AI) and ML technologies for better results.

Support Vector Machine (SVM) is widely used in medical imaging to classify different types of tumors. SVM can differentiate between normal and abnormal cells with high accuracy. Natural Language Processing (NLP) can analyze patient histories, and NLP can extract individuals who may need further screening based on risk factors available in their records.

By researching a number of factors including age, lifestyle, and environmental factors, regression models are used to forecast the probability of developing cancer for early screening and treatment. In the past years, multiple methods of machine learning techniques along with particular algorithms have been utilized to treat and predict disease.

While considering the prediction of cancer, these tasks must be taken into consideration: i)cancer survival, ii) recurring prediction for cancer) cancer susceptibility prediction(risk control). First, we must find the likelihood of getting cancer and, second the the likelihood of getting a particular kind of cancer again. Lastly, we must predict a survival outcome after cancer treatment. Cancer prediction outcome consists of i) survivability ii)life expectancy) therapy progression and efficacy. [13][14][15][16][17]

For many years, ML techniques such as DTs and ANNs have been employed to recognize cancer. Numerous studies, involving research articles, have been published representing the use of NL algorithms and the integration of data from various sources for tumor detection, cancer type identification, prediction, and treatment. Family history, age, diet, and exposure to the environment are all crucial factors in predicting the development of cancer, these factors are integrated into data about population-based, clinical, and histological data.

The details provided are not sufficient to forecast cancer symptoms. The transition to evolving imaging technologies may result in new types of molecular information.

Accurate prediction of disease outcomes is a very challenging task for physicians.

Thus, ML methods have been used for research purposes.

These techniques can identify patterns and relationships among datasets, with effective output predictability and identification of cancer type.

A generalized approach can be used for identifying particular causes of disease, which makes the treatment less expensive and more efficient.

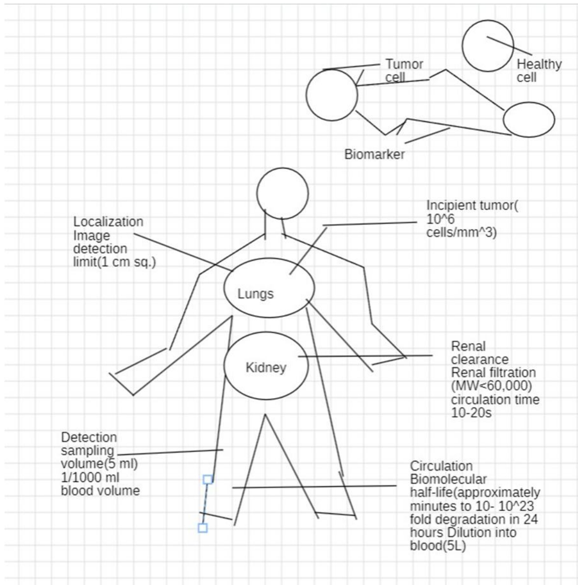
F. Concluding Remarks On Literature

There is extensive literature available in the field of machine learning usage for predicting disease with particular algorithms and specific disease diagnoses. Successful cancer treatment depends on early detection. Finding intrinsic biomarkers through blood examination is very important.[8]

Imaging systems also help in detecting tumors. Cancer that has been existing for ten years can be identified with current screening techniques.

Activity-based or genetically encoded mechanisms for early detection in biomarker studies solve several problems.

Figure 7(taken from mdpi.com) illustrates the problem related to the identification of tumors in their early stages, due to tiny size early-stage cancers are challenging to detect. Due to difficulties including biomarker transfer, dilution, and kidney function, only a few tumor-associated biomarkers can be found in a blood sample of a small part of blood volume.



As more providers adopt electronics, learning algorithms have the potential to grow in medical research and clinical enhancement. The implementation of machine learning techniques in the areas of diagnosis and outcome prediction using healthcare records may be helpful to the medical sector. This includes the possibility of identifying concerns for medical issues.

Skin cancer has been diagnosed using machine learning algorithms with a level of accuracy comparable to that of an experienced professional. Also, type 2 diabetes is predicted by utilizing patient medical reports from electronic healthcare. In medical sciences, the biggest advantage of ML approaches is automation which lets robots solve problems with less human input and get observations based on previous observations.

III. RESEARCH METHODOLOGY

A. *Overview Of The Proposed System*

Input data from imaging, clinical, genomic, histological, and other resources are included in the majority of literature on machine learning methods in cancer susceptibility, recurrence, prediction of cancer, and cancer survivability.

Regardless of ML techniques used, there is an increase in the prediction of cancer-type recurrence and risk evaluation.

Many researchers tried to predict the possibilities for re-developing cancer and enhancement of predictions with accuracy as compared to current techniques.

Vast publications have made predictions using clinical and molecular data.

Thus it is important to discuss the objectives, which include: i) survival rates ii) susceptibility; and iii) recurring with the help of ML techniques.It is a challenging task for AI to predict which medication is suitable for each patient based on many molecular, genetic, and tumor-based characteristics.[4] A large investigation has been made for the applications of AI in cancer risk possibility, diagnosis, cancer medicine development, and molecular tumor characterization. Different classification techniques have been examined to apply certain feature subsets, including K-nearest neighbors, SVM, and neural networks.[4]

SVM classifier models offer the highest level of overall accuracy in cancer treatment. Certain ML classification algorithms use stored historical data to learn from and predict new input categories, for example: benign and malign tumors.

The random forest model illustrated the highest accuracy of 96% in detecting different cancers. Finding the accuracy of SVMs and artificial neural networks(ANNs) was made simpler by observational research. Identifying potential models for predicting cancer is another advantage of using the Naive Bayes classifier.[4]

This research paper demonstrated that all ML approaches help in classification and high accuracy in cancer prediction. Also feature selection techniques can be used for more precise and better models with increased accuracy of prediction.

Clinicians realized the need for the usage of AI innovations such as DL and ML as the current arrival of the digital data era. Due to the complex nature of statistical analysis, it might be difficult to predict cancer progression.

Patients with more accurate prognoses can get more effective therapies, these approaches include Personalization care of the patient. Thus, AI can provide more precise information about patient survival, prognosis, and disease progress prediction for the prediction of cancer symptoms.

Algorithms based on AI have been represented to analyze unstructured data and estimating the possibilities for patients getting different illnesses. Certain AI algorithms can improve risk management and influence cancer screening outcomes with recommendations. Example: an artificial neural network design is accurate as compared to current screening guidelines.[25]

Thus, screening might separate patients into categories of a high risk of developing cancer. Personal risk prediction might help in early detection with increased treatment rates for malignant tumors for which there are no validated detection methods.

1) Evaluation of SVM, CNN, and Logistic Regression Models for Detecting Brain Tumor

Convolutional Neural Networks (CNN) CNNs are sub-categories of deep learning and it is based on layered architecture. They can extract information about characteristics in images. For classifying images, CNN architecture has been utilized.

2) Support Vector Machines (SVM)

This is a supervised training model. It is particularly implemented in classifying binary objects with algorithms and manages non-linear decision boundaries.

3) Logistic Regression

It helps in predicting of unifying results, which is structured with statistics.

4) Methodology

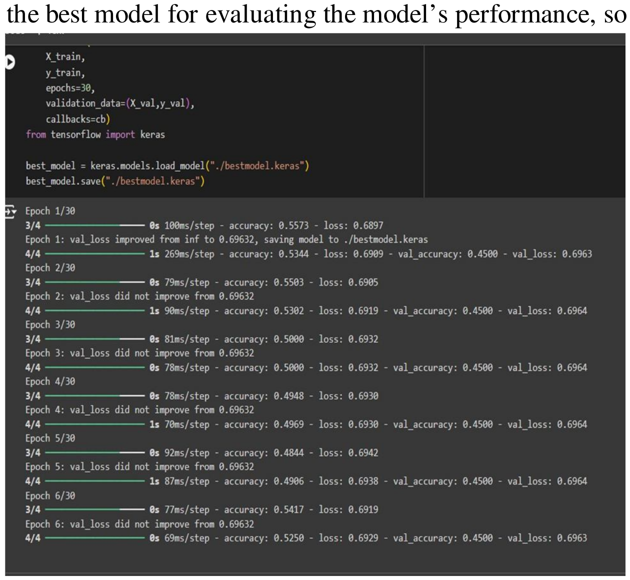
a) Data Collection and Processing

For the comparison, two datasets were implemented. MRI images of brain used as dataset labeled “yes” or “no” was used. Another dataset consists of testing as well as training folders.

b) Model Training

CNN algorithm is trained by splitting the information into 70% training data, 15% validation data, and 15% test data. The CNN architecture was designed by implementing layers like batch normalization.

Figure 8 illustrates how CNN utilizes the best model for evaluating the model’s performance, so predictions have been displayed.



c) Accuracy Comparision

Accuracy was considered a prime factor for model efficiency evaluation.
CNN model gained 33.333% accuracy.

Figure 9 illustrates the accuracy of model.

```
acc=model.evaluate(test_data)[1]
print(f"Our Model Accuracy is {acc*100} %")
```

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 3s 3s/step - accuracy: 0.3333 - loss: 0.7241
Our Model Accuracy is 33.33333432674408 %
```

SVM is used for training the model, a score helped in evaluating the efficiency. It demonstrated : SVM was better regarding particular dataset. While the testing score is trained, SVM has higher accuracy than logistic regression.

Figure 10 illustrates the training score and testing score of SVM and Random Forest.

```
#Evaluation
print("Training Score:",lg.score(xtrain,ytrain))
print("Testing Score:",lg.score(xtest,ytest))
```

```
Training Score: 1.0
Testing Score: 0.9551020408163265
```

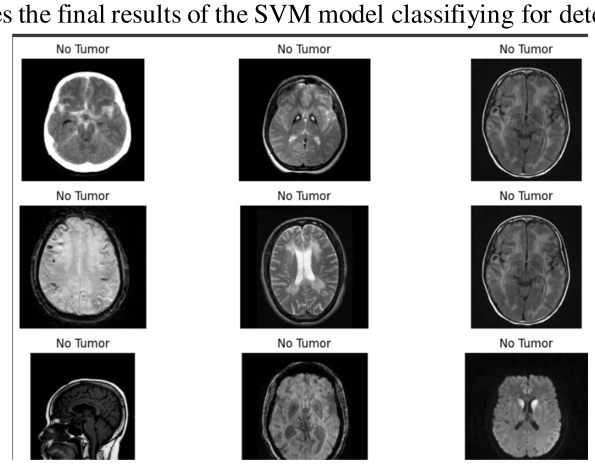
```
print("Training Score:",sv.score(xtrain,ytrain))
print("Testing Score:",sv.score(xtest,ytest))
```

```
Training Score: 0.9918116683725691
Testing Score: 0.9591836734693877
```

5) Conclusion of Results

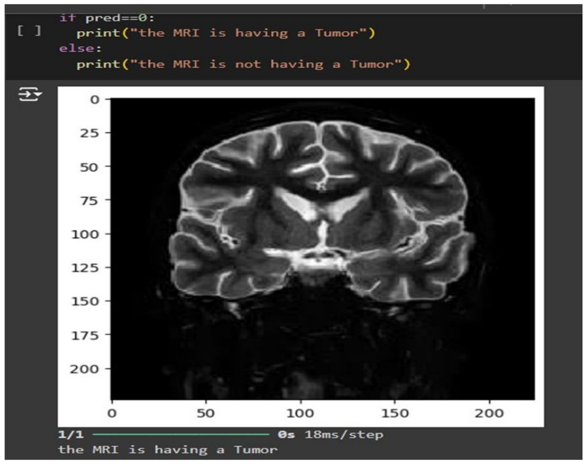
SVM can be effective regarding smaller datasets. For classifying images based on accuarate predictions, SVM might be less efficient.

Figure 11 illustrates the final results of the SVM model classifiying for detecting brain tumor.



International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

Figure 12 illustrates the classification of the MRI scans with the help of CNN.



CNN algorithm learns by extracting features from images and analyzing the small details. It has great efficiency with exact predicting of results.

The graphical representation of the CNN model can be described with accuracy and loss.

Figure 13 illustrates the graph for predicting accuracy and value accuracy.

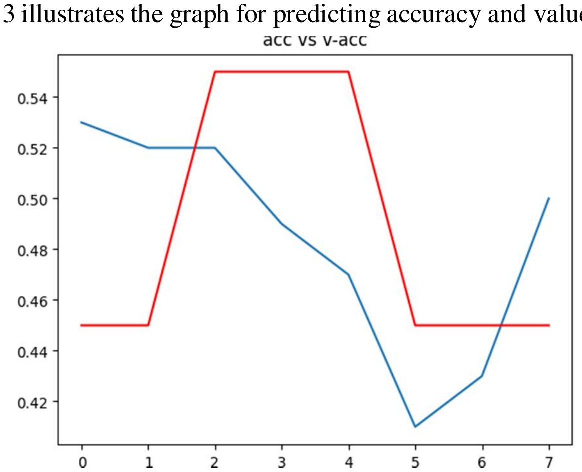
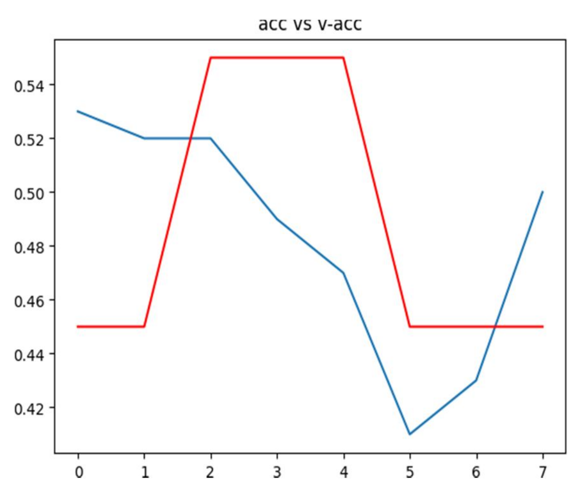


Figure 14 represents the graph of comparison between loss and value-loss, regarding the CNN model.



International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

B. Prediction Of Cancer Susceptibility

ANNs are the means to estimate cancer risk prediction, including decision-making tools to classify benign and malignant findings of cancer. This includes risk stratification (risk stratification involves grading of patients with incremental risk levels and it is defined as the process of medical decision-making with a collection of clinical and laboratory testing for determining the risk of a person for suffering a particular condition, need for prevention) and the construction of predicting models.[4][18][19][20]

Studies that depend on computer models, also use machine learning (ML) methods, such as artificial neural networks(ANNs), to calculate cancer risk. ANNs are employed for the construction of a predicting model for classifying the malignant findings from benign samples.

With the generalization of connections using small concealed nodes, the design can be built with large amounts of hidden layers. Factors related to risk issues and tumor characteristics, comprising information gathered can be reviewed by specialists and information can be obtained. The ANN model can be given the generated datasets.

Cross-validation can be used to estimate the performance.

The calculated Area Under the Curve(AUC) can be calculated using certain fold cross-validation. The model can thus estimate the risk assessment of cancer patients by integrating a large sample of data.

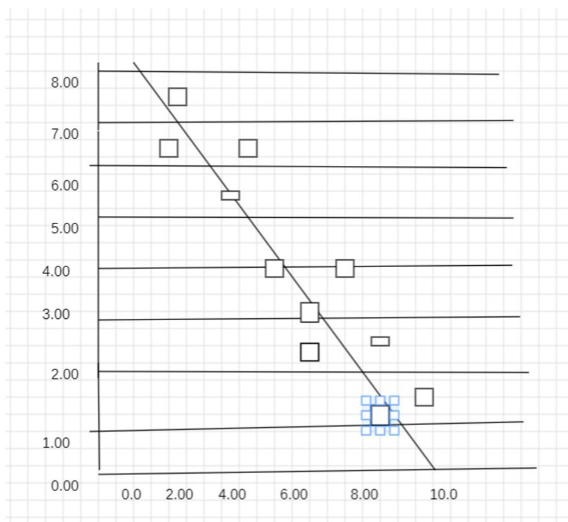
The two most frequently occurring elements of accuracy are calibration and discrimination.

Calibration is a metric used for risk management with the purpose of sorting patients into high or low-risk groups, in contrast to determination is a measurement

for calculating the separation of benign abnormalities from malignant ones.

Plotting a calibration curve to compare the model's calibration to the ideal calibration of the cancer prediction and a curve to determine the model's discriminative power. Utilizing a combination of diagnostic and screening datasets which are indistinguishable when giving ANNs with input. Pre-processing steps are therefore important to modify unprocessed data into appropriate formats for evaluation.

Figure 15 shows a calibration curve that is related to detection time(DT) obtained to log with base 10 of initial bacteria concentration.



C. Prediction Of Cancer Recurrence

For the purpose of analyzing the basis of oral squamous cell carcinoma(OSCC) development after treatment, a multiparametric decision-facilitating system was suggested in one study that was taken into account for predicting recurring OSCC. [21]

Thirteen of the 86 patients who were taken into consideration subsequently relapsed, while the rest were found to be free of illness. Wrapper algorithms and two feature selection algorithms are utilized after a particular selection process.[4]

While choosing informative characteristics from heterogeneous datasets, bias should be prevented. For the vectors that are submitted to certain classifications, selected input factors may be utilized. The total overall genomic, imaging, and clinical data comprised one of the characteristics selection methods. The CFS algorithm was applied to extract extra features such as more specific features of tumor spreading features.

This literature shows discrimination related to patients into categories of disease relapse and those with the absence of performance related to five classification algorithms. Employed algorithms are Artificial Neural Network(ANN), Batch Normalization(BN), Random Forest(RF), Support Vector Machine(SVM), and Decision Trees(DT) models. A ten-fold validation that was crossed, was utilized for evaluation resulting from the successful application of each machine learning technique.

Additionally, for the task of comparing different classification schemes, specificity, sensitivity, and accuracy are determined. Also, the analysis is considered for evaluation purposes.

The classification schemes used to collect data on data classification without data collection and feature selection after using a feature selection technique will be determined by the results.

Without using any feature selection strategies, the BN classifier produced output that was more efficient in terms of discriminating between imaging and clinical information. Similarly, genomic-based classification results revealed the best-performing classifier was BN coupled with CFS.

The classification results were very good, biased predictors and incorrect classification might come from the short sample size compared to data dimensionality.

D. Prediction Of Cancer Surviving

A predictive model is developed for evaluating cancer, addressing the importance of robustness under the model's parameter variation.[22]

The SEER (Surveillance, Epidemiology, and End Results) cancer database was used to compare the classification models SVM, ANN, and SSL(self-supervised learning).[4][23]

Additionally, a class characteristic named survivability, which refers to patients who have survived and those who have not survived the disease- was taken into consideration. Among the informative features were i) the patient's age at therapy ii) the size of the tumor, and iii) the number of nodes. After comparing the three models' performance the best performance is evaluated with their predetermination of accuracy.

Five-fold cross-validation was utilized to evaluate the efficiency of prediction designs. It should be noted that no processing regarding to collection of significant instructive aspects was stated. More robustness and reliable output under parameter conditions are depicted by a tiny box region on a particular diagram.

Example: To employ artificial neural networks (ANNs) to determine the survival of patients with non-small cell lung cancer(NSCLC). The databases consist of clinical data from the NCI array database and raw gene expression data from NSCLC patients. [4]

The most helpful survival-associated genes were extracted after the pre-processing stages and utilized to train the ANN network. The ANN model utilized four clinical characteristics as input variables: namely gender, age, and stages.

A variety of ANN architectures to see which is best to predict cancer survival. In terms of the model's predictive performance, the overall accuracy was much better, using classification approaches.

Classification of patients in different groups with their treatment protocol while 50% of them had not survived. The estimated survival of patients related to the test, validation, and training sets, with specific

p-values., they demonstrated that given certain p-values, patients in high-risk groups had lower median survival than in low-risk groups. [4]

This study is only peculiar to NSCLC, not other types of cancer. The idea that prediction models are not generalizable is considered an important drawback in research. [4]

E. Drawbacks Of Using Ai With Facilities Related To The Medical Field

Utilizing AI systems in any field, including healthcare, has limitations and drawbacks. The perspective must be changed to deal with the flaws of AI as it is an emerging technology in this digital era.

Firstly, data privacy is a major issue. Data accessibility and storing data is the first step in developing a system after problem selection and developing a solution approach. The requirement of a large quantity of effective data is necessary for the creation of an effective model. So, privacy concerns and data sharing issues as performed by many organizations. Also, data collection is a serious concern. For example: the patient's confidentiality limits the availability of data, which restricts the training of the model, thus the model is not efficient. Ethical and Legal Concerns are also mandatory for the implementation of AI as patient privacy and security of data are issues related to the use of ML in the detection of cancer. The solution for this is using encryption secure data sharing protocols to protect sensitive health information.[1]

Secondly, data quality is also an important factor in evaluating the drawbacks of AI.

AI systems rely on large datasets for training. Inaccurate and incomplete results can be produced with incomplete datasets. The solution for this issue is implementing bias detection algorithms and fairness constraints during the model development process.

Another concern is the lack of explainability, which shows that many algorithms make it difficult for doctors to understand how a diagnosis was made. The solution to this problem is developing and using models that offer transparency, allowing clinicians to understand how a diagnosis was made.

The other issue is the high initial cost as training staff, integrating AI, and acquiring technology is expensive.

The solution for this could be seeking public and private funding. Detailed cost analysis for long-term technology usage is important.

Another concern regarding the integration of ML is the emerging technology can be difficult for the adaptability of current medical techniques. With the existing technology, using AI can be difficult to integrate as environmental conditions are required.

The solution could be generating models that relate to current technological use is important.

Another issue is the generalization of Machine Learning can be a challenging task. Training is required for staff and using advanced technology in hospitals needs proper knowledge.

The solution can be simple and understandable, models can be generated and further training can be given. Lastly, the fragmentation of data is also a limitation for the deployment of artificial intelligence, it describes the design generation and deployment by one organization for a specific job such as natural language processing, regression, classification, and clustering cannot be transferred to another organization for immediate use without recalculation.

A tiered architecture has been defined by a team of researchers, which has access, transitions, quality, and socio-economic and environmental impact, thus it offers a framework to establish measurements that may be helpful to advance the equal resources for both patients and staff of the healthcare provider institution.

Blockchain technology has the potential to reduce fragmentation in the healthcare industry.

F. Future Prospects Of Ai Implementation

Obstacles must be resolved and AI approaches must be consented by future research, the integration of AI-based models will be possible.[1]

The data knowledge will replace the AI applications which is a better knowledge of tumors for more accurate therapy options, and ultimately, the decision-making process will be enhanced.

The treatment will be more specialized and patients will be getting frequent treatment.

Risk management tools integrated with smartphone applications will give the general public an intermediate cancer risk prediction. This will help drive patients to receive medical help and proper medical advice. Risk reduction estimates might help individuals to adopt healthier behaviors, such as quitting smoking, or being physically active.

The algorithms will assist physicians in determining healthcare facilities in a primary care setting. The electronic record system can help in providing an option for improved resource distribution based on subgroups with the reduction in greater risks of cancer growth or related consequences.

The rise of Chat GPT, by Open AI has great potential and knowledge of vast subjects. It is a large language model which is a faster-growing consumer application. According to the research estimates, the Chat GPT could transform the field of surgery by providing personalized and precise medical data, deducing errors, and improving outcomes. The rapid analysis of radiographic imaging and results based on tumor genomics are some opportunities in the processing of diagnosis using Chat GPT.[24]

IV. IMPLICATIONS FOR THEORY AND DISCUSSION

The application of AI and ML techniques and their relevance in cancer predicting cancer are discussed in this paper.

Following the explanation of the machine learning branch and the concepts of data processing techniques, feature selection methods and classification algorithms are employed, and case studies regarding the use of ML tools to predict cancer survival, recurrence, and susceptibility, recurrence are explained.

There are a variety of published ML studies that accurately align with results for particular prognostic outcomes of cancer and it is important for medical decisions such as design, appropriate data sample collection, and validation of classified results.

While decision-making is accurate and efficient, ML techniques have not been implemented in clinical practice. Semi-supervised ML techniques were widely used to predict the model of survival.

For prediction, the algorithm uses both labeled and unlabeled data. When compared to existing supervised ML techniques, the performance has improved.

The small amount of data samples is a disadvantage because the training dataset size is a fundamental prerequisite for classification in disease modeling, which must be sufficiently large.

The large datasets enable training and testing set splitting, which results in estimated satisfactory verification.

Inadequate datasets can result in misclassifications with the production of unstable and biased models. The generalization of the predictive model can be enhanced with richer sets related to patients used to predict illness survival.

Quality associated with the dataset with careful feature selection methods is very important for efficient machine learning and precise cancer forecasts. Additionally described are the structural sets with pathological histological evaluations.

The performance of validation tests of the learning algorithms must be utilized. The evaluation methods divided the original datasets into smaller groups. For the achievement of accurate results for the predictive model. Better validation can develop the selection of broad and independent features.

SVM and ANN classifiers are the most commonly used machine learning algorithms for predicting cancer patients. SVM constitutes the most recent approach for predicting cancer and is most commonly used related to accurate predictive performance.

To address the drawbacks, new models for cancer modeling must be researched for the future.

The better statistical analysis of heterogeneous datasets used will provide more accurate results and reasonable disease outcomes. Public databases must be built is needed to gather valid cancer datasets related to patients with disease diagnoses. Studies of these models will result in integrated decision-making.

V. CONCLUSION

The study explores the concepts of ML and its application in cancer-related estimation and prediction. Most studies were proposed focussing on creating predictive models to forecast disease outcomes utilizing supervised ML methods and classification techniques. Applying different feature selection and classification methods to multidimensional heterogeneous datasets can be enough to intervene in the cancer field.

VI. FUTURE WORK

As technology has advanced over time, AI-ML has emerged in this period. The innovations were used for non-medical purposes previously. But now these innovations are starting to be implemented for the enhancement of healthcare around the globe.

Every aspect of cancer research including diagnosis, prognosis, and treatment has applications showing the potential of AI. Cancer, a life-threatening disease, might have a treatment soon. Prevention is better than a cure. Early detection, cancer prediction, and disease prognosis prediction are important.

Early detection of cancer greatly increases the chances of successful treatment. Early diagnosis focuses on detecting symptomatic patients as early as possible.

Image Classification is possible with machine learning algorithms, particularly convolutional neural networks(CNNs) can classify medical images to detect tumors.

Natural Language Processing(NLP) as discussed above, NLP techniques can analyze unstructured clinical notes and research articles related to cancer diagnosis and treatment.

We can use several technologies like Support Vector Machine(SVM) for classification tasks. Random Forests for feature selection and classification in datasets Neural Networks can be used for complex pattern recognition in imaging and genomic data combining multiple models to improve predictive accuracy and robustness.

This helps in increased awareness of the first signs of cancer, among physicians, nurses, and other health care providers as well as among the general public, improved accessibility and tertiary levels of care.

The obstacles associated with disease will be overcome one day with the help of increasing algorithms, additional studies are required to continue to ensure clinical value and analytical and clinical validity.

REFERENCES

[1] "What Is Cancer?," National Cancer Institute, Oct. 11, 2021.

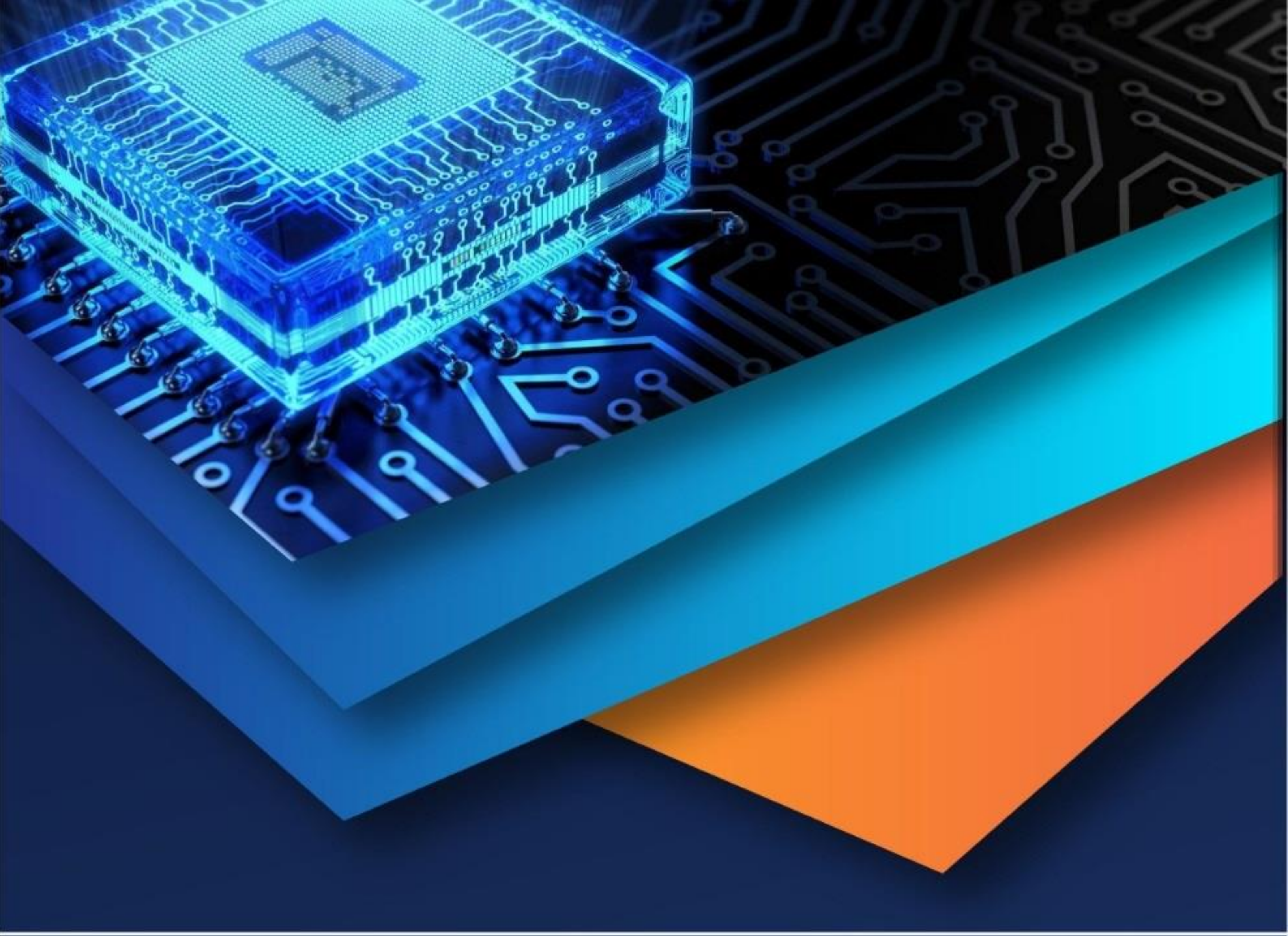
[2] B. Zhang, H. Shi, and H. Wang, "Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach," Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach, vol. Volume 16, no. 16, pp. 1779–1791, Jun. 2023.

[3] WHO, "Cancer," World Health Organization, Feb. 03, 2022.

[4] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," Computational and Structural Biotechnology Journal, vol. 13, no. 13, pp. 8–17, 2015, doi:

[5] Google, "What Is Artificial Intelligence (AI)?," Google Cloud, 2023.

[6] GeeksforGeeks, "Introduction to Convolution Neural Network," GeeksforGeeks, Mar. 14, 2024.
[7] S. Suman, "Brain Tumor classification and detection from MRI images using CNN based on ResU-Net Architecture," Medium, Mar. 03, 2021.
[8] S. Das, Mohan Kumar Dey, Ram Devireddy, and Manas Ranjan Gartia, "Biomarkers in Cancer Detection, Diagnosis, and Prognosis," Sensors, vol. 24, no. 1, pp. 37–37, Dec. 2023.
[9] "Google Scholar," Google.com, 2024. .
[10] "Google Scholar," Google.com, 2024.
[11] "Google Scholar," Google.com, 2024.
[12] "Google Scholar," Google.com, 2024.
[13] "Google Scholar," Google.com, 2024.
[14] D. V. Cicchetti, "Neural Networks and Diagnosis in the Clinical Laboratory: State of the Art," Clinical Chemistry, vol. 38, no. 1, pp. 9–10, Jan. 1992.
[15] "Google Scholar," Google.com, 2024.
[16] "Google Scholar," Google.com, 2024.
[17] "Google Scholar," Google.com, 2024.
[18] "Google Scholar," Google.com, 2024.
[19] "Google Scholar," Google.com, 2024.
[20] "Google Scholar," Google.com, 2024.
[21] "Google Scholar," Google.com, 2024.
[22] "Google Scholar," Google.com, 2024.
[23] "Google Scholar," Google.com, 2024.
[24] W. Li, Y. Zhang, and F. Chen, "ChatGPT in Colorectal Surgery: A Promising Tool or a Passing Fad?," Annals of Biomedical Engineering, May 2023.
[25] S. Yang, Q. Li, W. Li, X. Li, and A.-A. Liu, "Dual-Level Representation Enhancement on Characteristic and Context for Image-Text Retrieval," IEEE Transactions on Circuits and Systems for video technology, vol. 32, no. 11, pp. 8037–8050, Nov. 2022.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY