DOI:10.33266/1024-6177-2025-70-3-83-89

Muaayed F. Al-Rawi, Izz K. Abboud, Nasir A. Al-Awad

ИСПОЛЬЗОВАНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АВТОМАТИЧЕСКОГО ВЫЯВЛЕНИЯ ОНКОЛОГИЧЕСКИХ ЗАБОЛЕВАНИЙ

College of Engineering, Mustansiriyah University, Baghdad, Iraq

Контактное лицо: Muaayed F. Al-Rawi, e-mail: muaayed@uomustansiriyah.edu.iq

РЕЗЮМЕ

Число онкобольных растет во всем мире. За последние двадцать лет количество таких пациентов в Ираке увеличилось вдвое, что привело к увеличению числа смертей от рака. Помимо этого, именно опухолевые поражения являются второй по частоте причиной смертей госпитализированных пациентов. Пути решения сложившейся проблемы заключаются в уменьшении времени диагностики онкозаболевания, увеличении ее точности, правильности алгоритмов маршрутизации пациентов с симптомами рака, а также в улучшении систем мониторинга. Рассматриваемый в статье подход к ведению онкобольных подразумевает использование программного обеспечения на основе алгоритмов машинного обучения, позволяющего пациенту самостоятельно распознать симптомы онкологического заболевания и направляющего его к профильному специалисту, что в свою очередь обеспечит выявление рака на ранней стадии. Помимо этого, рассматриваемое ПО призвано обеспечить мониторирование состояния пациента на протяжении лечения. В имеющихся исследованиях применительно к ранней онкодиагностике рассматривается лишь один метод машинного обучения. В данной работе проанализировано применение сверточных нейронных сетей (CNN), классификаторов Random Forest и XGBoost, которые представляют собой алгоритмы машинного обучения, применяемые к структурированным и табличным данным, используемым для выявления наличия рака молочной железы, опухолей головного мозга, рака кожи и рака легких. Использование данных программ обеспечит более быструю и более точную диагностику рака. Создание облачного сервера с таким ПО сделает предлагаемую методику ранней онкодиагностики общедоступной и более удобной в использовании.

Ключевые слова: *лучевая диагностика, машинное обучение, Random Forest, классиикатор XGBoost, обнаружение рака, рак кожи, рак головного мозга, рак легкого*

Для цитирования: Al-Rawi Muaayed F., Abboud Izz K., Al-Awad Nasir A. Использование алгоритмов машинного обучения для автоматического выявления онкологических заболеваний // Медицинская радиология и радиационная безопасность. 2025. Т. 70. № 3. С. 83–89. DOI:10.33266/1024-6177-2025-70-3-83-89

DOI:10.33266/1024-6177-2025-70-3-83-89

Muaayed F. Al-Rawi, Izz K. Abboud, Nasir A. Al-Awad

Using Machine Learning Algorithms to Detect Cancer Automatically

College of Engineering, Mustansiriyah University, Baghdad, Iraq

Contact person: Muaayed F. Al-Rawi, e-mail: muaayed@uomustansiriyah.edu.iq

ABSTRACT

The number of people diagnosed with cancer is growing all around the world. During the last twenty years, the overall cancer incidence in Iraq has doubled, leading to an increase in the number of diagnosed cancer fatalities. When it comes to deaths that occur in hospitals, cancer is the second-biggest cause. Therefore, a remedy to the issue should be an arrangement to decrease time waste, the right technique of directing the patient to notice symptoms, extremely accurate cancer detection, and a better monitoring system. The proposed method is an arrangement that lets and leads a patient to identify symptoms on their own, guiding them to a proper healthcare professional, correctly diagnosing cancer in its initial stages, and monitoring the patient throughout therapy. Currently, research into cancer detection systems only employs a single machine learning approach to identify cancer. The study that is being presented makes use of Convolutional Neural Networks (CNN), Random Forest, and the XGBoost Classifier, which are a machine learning algorithms that are applied to structured and tabular data in order to identify the existence of breast cancer, brain tumors, skin cancer, and lung cancer. These methods provide findings more quickly while also achieving a greater level of accuracy. Hosting this suggested solution in the cloud with a cutting-edge program will make it available to the public, providing an improved user experience and easier operation.

Keywords: *radiation diagnostics, machine learning, CNN, Random Forest, XGBoost classifier, Cancer detection, Brain cancer, Skin cancer, Lung cancer*

For citation: Al-Rawi Muaayed F., Abboud Izz K., Al-Awad Nasir A. Using Machine Learning Algorithms to Detect Cancer Automatically. Medical Radiology and Radiation Safety. 2025;70(3):83–89. DOI:10.33266/1024-6177-2025-70-3-83-89

1. Introduction

The state sector in Iraq mostly supplies cancer services, offering them at no cost to the general population. Following the introduction of a national cancer strategy on cancer prevention and control, there has been a discernible increase in the quality of cancer care provided throughout the country. Stopping smoking and getting a Human papillomavirus vaccine (HPV) are two important preventive measures. The NCEDC center makes scanning strategies available of some cancers, including breast, oral, and cervical cancer. However, during the year, countless cases of skin, lung, brain, and breast cancer are discovered and treated.

2. Literature

The number of people affected by cancer worldwide is increasing. Over the past twenty years, the amount of confirmed instances of cancer has increased, coinciding with a rise in the number of deaths attributed to cancer. Cancer has become the second most prevalent cause of death in hospitals in Iraq [1], behind the causes of death from other diseases. This study aims to present an overview of the current situation and suggest a system that can quickly recognize cancer in its early stages using machine learning algorithms.

2.1. Breast Cancer

By referencing pertinent information derived from a variety of sources, the authors investigate the identification of breast cancer. In their research, the authors have used the most widely employed techniques: Random Forest, KNN (k-Nearest Neighbor), Naive Bayes, Support Vector Machines (SVM), and Bayesian Networks (BN) are among the methods used to diagnose breast cancer [2, 3]. The SVM classifier approach merges RFE and SVM into a single step. Recursive feature extraction (RFE) is a method that selects features from a dataset by determining which features have the lowest value. In each cycle of SVM-RFE, the algorithm eliminates the features with the lowest weight, which are considered the wrong features. The support vector machine (SVM) fulfills its job by selecting important examples from all classes, which it refers to as support vectors. Support vectors are then used to separate the samples, creating a linear function that maximizes the separation between them [2]. Support vector machines (SVM) map an input vector to a high-dimensional space to locate the hyper plane that best classifies the data set. The Random Forest methodology claims that a single decision tree can generate either a simple or a specific model. Random Forest (RF) is used in cancer diagnosis due to its effective handling of data from minorities, as demonstrated in the evidence presented in the paper [2]. For example, even if a tumor comprises only 11 % of the entire data set analyzed, it can still be classified as either benign or malignant. The term "naive Bayes classifier" refers to probabilistic classifiers that use Bayes theory as their foundation. Naive Bayes has demonstrated utility for a wide range of machine learning tasks, which is uncommon in real-world scenarios, even though it assumes that all characteristics are independent of one another [4]. The k-nearest neighbor algorithm is a kind of algorithm that may be defined as an algorithm that determines the location of a data set by analyzing the data sets that are located in its immediate vicinity. supervised learning technique that incorporates regression and classification. In order to analyze a new data point, KNN first collects all of the data points that are nearby. Key criteria in the process of measuring the distance are attributes that exhibit a significant degree of fluctuation. The researchers determined that the KNN algorithm accuracy is 94.72 %, with only one observation incorrectly labeled as benign and five as malignant. Based on the findings, the researchers came to the conclusion that the KNN algorithm is superior to the Random Forest Method [4]. To create an image that can be utilized by machine learning algorithms, several image processing functions are employed, as demonstrated by research [5]. Machine learning algorithms employ a sequence of image processing functions to build an image. There have been a number of studies that have sought to use machine learning for the detection and diagnosis of breast cancer. These studies have utilized a variety of approaches or a mix of algorithms in order to achieve a higher level of accuracy. Reviewing a number of research papers can provide insights into the limitations of these methodologies. For instance, the support vector machine (SVM) classifier fails when applied to high-end computer vision applications with big datasets. The Naive Bayes classifier does not produce adequate results when the training data is poorly represented [6].

2.2. Brain Tumor Cancer

It is not possible to overstate the significance of receiving a diagnosis and discovering brain tumors at an early stage. Medical professionals commonly use computer-aided diagnostic (CAD) methods in a methodical and specialized way to identify brain tumors. According to medical professionals, a brain tumor is a development of tissue that has the potential to disturb the normal function of the brain. In the United States of America alone, medical professionals are expected to diagnose 84,870 people with brain and other neurological cancers in 2021. The IARC cancer reports a 78 % death rate associated with brain tumors. It emphasizes how crucial it is to detect brain cancer as soon as possible and help patients adhere to the prescribed course of treatment to avoid more difficulties [7].

Technological developments have made it possible to employ computerized systems to automatically identify brain cancers through the use of computed tomography scans and magnetic resonance imaging (MRI). Convolutional neural networks (CNN) and other machine learning and deep learning methods have gained popularity in medical research for the identification and categorization of brain cancers. Moreover, identifying malignancies necessitates executing procedures with remarkable accuracy and speed [8]. Magnetic resonance image segmentation isolates questionable regions from complex medical images. It is possible to diagnose a brain tumor using manual procedures. Aberrant blobs or regions in the brain are detected within MRI scans. These blobs, or regions of the brain, have a distinct illumination pattern from the brain relief and have a higher level of brightness than the brain relief. The process of segmenting tumors in MRI, on the other hand, is highly difficult. There is a wide range of variations in the morphology, size, texture, and even location of tumors. In the process of distinguishing the tumor based on characteristics such as light, it is possible to encounter complications such as pixel intensities that overlap with those of normal tissues. Identifying and segmenting brain tumors in MRI is crucial, as it reveals the presence of abnormal tissues that can be utilized for treatment or patient follow-up [9, 10].

2.3. Skin cancer

The impact of cancer on healthcare systems across the globe is one of the most significant. In the year 2022, it is estimated that cancer will be responsible for over 9 million fatalities worldwide. The two most frequently diagnosed types of cancer in females are breast cancer and lung cancer. Cancer attributed to malignancies of the lungs, liver, and stomach is the most common cause of mortality. A common type of cancer among Caucasians, skin cancer – which encompasses both malignant melanoma and nonmelanoma skin cancer (NMSC) – is expanding in incidence. There are more people in the United States who are diagnosed with skin cancer each year than there are with all other types of cancer combined, as stated by the United States Skin Cancer Foundation [11]. Researchers have observed a 51 percent increase in the yearly incidence of melanoma cases. The increased exposure to ultraviolet (UV) particles may contribute to a portion of this rise. Early identification may significantly raise the chance of survival, even though melanoma is one of the most deadly types of skin cancer [12]. The World Cancer Research Fund (WCRF) reports that melanoma, which is a kind of skin cancer, ranks as the nineteenth most prevalent form of both male and female cancer. In 2018, 332,000 new cases were reported. Skin cancer ranks fifth among all cancers for men and women, including melanoma and non-melanoma. As of 2019, over a million cases had been reported worldwide. The American Academy of Dermatology Association (AADA) conducted a study which found that skin cancer is the most common form of cancer in the United States. There was a 138 % increase in the overall incidence of basal cell carcinoma (BCC) between 1975–1985 and 2000–2010, whereas the overall incidence of squamous cell carcinoma (SCC) increased by 253 % over the same time period. Nonmelanoma basal cell carcinoma and squamous cell carcinoma patients were more likely to be female than male [13]. The American Cancer Association reports a steady rise in the incidence of skin cancer over the last three decades, with 83,000 new cases of melanoma diagnosed each year in the United States alone. Ferris et al. (2017) found that when it comes to cutaneous melanoma, the most difficult aspect for medical personnel to deal with is the diagnosis rather than the treatment [14].

2.4. Lung cancer detection

Considering the technical context of this study, we can determine that the machine learning models are the most crucial aspect of these systems. To achieve a higher level of accuracy, researchers should train these models using a more extensive dataset. The mentioned study demonstrates the performance of those studies, taking into consideration that they indicate certain particular technologies that may be utilized for diagnosing lung tumors using machine learning, as stated by several researchers who have been doing their work over the last few years [15]. Specifically, some researchers have achieved an accuracy level of only about thirty percent. The findings of that study reveal a gap in the researchers' utilization of technologies and collection of datasets. The study provides them with a glimpse into the produced outcomes. On the other hand, some of them have conducted their studies using just a single kind of dataset. When taking into account lung cancer nodules and the research that goes along with them. This provides information on how to diagnose lung cancer using images from CT scans. They have utilized several preprocessing approaches to clean up the datasets in preparation for the analysis, after which the model was trained to utilize those datasets. Furthermore, they employed these procedures in addition to the facts supplied by the study. During the course of the study, they also improved the accuracy of the training process. Those researchers use machine learning or deep learning algorithms the majority of the time. Taking into account the history and the works that are associated with it, the study demonstrates an accuracy of around 92 % [16, 17].

3. Research Approach

3.1. Identification of Breast Cancer

The utilization of two different datasets is one of the methodologies used in this research. The first dataset contains information on the results of the tests, such as the diagnosis of the breast tissue (whether it is benign or malignant), the thickness of the clumps, the uniformity of the cell size, and so on. All of the breast tissues' ultrasound images are included in the second dataset. The main data source for execution is a dataset made up of breast cancer case parameters. These factors include the radius, texture, perimeter, area, and other properties of the tumor in the breast as well as the breast tissue. The radius, texture, perimeter, area, and other characteristics of the tumor in the breast are among these factors. Model 1 training utilizes this dataset. The process of getting clean, diagnosable breast cancer data involves a number of different procedures all working together. Initially, researchers need to study and analyze the selected dataset. We obtained the original dataset in Wisconsin to train the model for this investigation. Before beginning the process of feature extraction, it is necessary to investigate the dataset. Initially, we will carry out both pre-processing and investigation of the dataset. After conducting an exploration of the dataset and doing some preliminary processing on it, the third stage consisted of picking the most suitable model to train before separating the dataset. There are a few processes that need to be finished. These include loading the data, encoding the descriptive data, showing the category data, and eliminating the blank columns in the extracted data. We need to divide a section of the dataset, which consists of 75 % training data and 25 % testing data, into two halves. Using the K-Neighbors Classifier, Random Forest Classifier, Decision Tree Classifier, and Logistic Regression as classifiers during the model training process [5], 99.7 % training accuracy was attained by the decision tree classifier and 99.6 % by the random forest classifier.

These accuracies were taken into consideration for each training approach. The results showed that the Decision Tree Classifier achieved an accuracy of 95.3 % when the models were employed on test data and recorded on a confusion matrix. The accuracy of 95.3 % for the Decision Tree Classifier and the success rate of 96.7 % for the Random Forest Classifier were determined by operating the models on test data. In comparison, the Random Forest Classifier had a success rate of 96.7 % when taking into account the accuracy of both the training and testing processes, the Random Forest Classifier is the one that is most suited for the testing of the model. The study makes use of a dataset that contains ultrasound images. of breast cancer tissues in order to carry out the training for Model 2. This collection of data contains ultrasound images. of breast tissues that are either benign or cancerous, as well as normal. Model 2 accomplishes its training by utilizing this relevant data collection. Information pre-processing, extraction of characteristics, and the division of the data set into test and train data were among the procedures we applied to the data set during the training phase. To increase accuracy, the procedure makes use of the Convolutional Neural Networks (CNNs) approach in combination with the Keras sequential model. It is feasible to increase accuracy after the two models, which eventually leads to a high level of accuracy, by using the ensemble technique, which combines the accuracy of the two models. Fig.1 shows sample of breast non-cancer (A) versus cancer (B) using MRI.

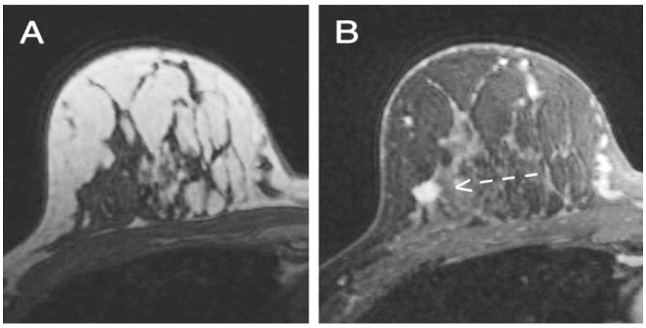

Fig.1. Sample of breast non-cancer (A) versus cancer (B) using MRI

3.2. Identification of Brain Tumor

This suggested approach will yield results faster and more accurately by using CT and MRI image data sets in combination with other tests to detect some of the extreme typical signs of a brain tumor, such as diminished hearing (acoustic neuroma) and altered vision (a lack of vision). The goal is to identify brain tumors as soon as possible and stop growing them before they become apparent. Convolutional neural networks (CNNs) with the LeNet-5 architecture were chosen as the first model. The model was composed of six layers: the Flatten, Conv2D, Dropout, Activation Functions, Dense, and Max Pooling layers. All of these layers also included the Max Pooling Layer. The categorical cross-entropy combined with ReLU along with soft-max activation techniques was applied to both the activation and the loss function. Fig. 2 shows sample of brain tumor (A) versus non-tumor (B) using MRI.

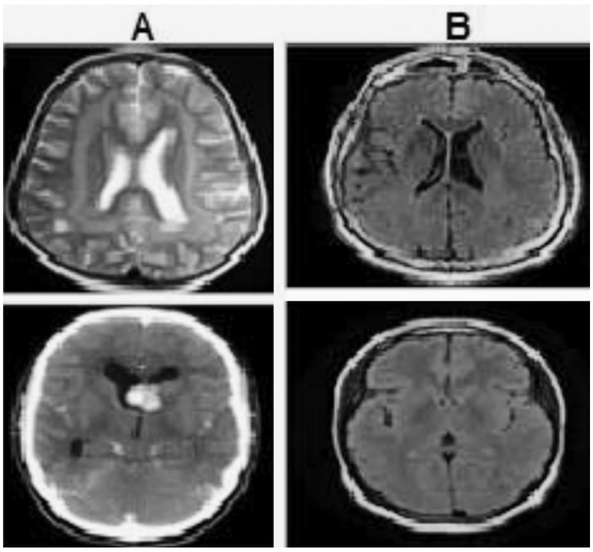

Fig. 2. Sample of brain tumor (A) versus non-tumor (B) using MRI

3.3. Identification of Skin Cancer

Skin cancer has emerged as one of the most significant challenges in the field of healthcare worldwide, leading to the development of several technological approaches for its diagnosis and prevention. Approaches based on machine learning can be found, but they solely focus on identifying whether or not the user has been diagnosed with skin cancer. Aside from such functionality, those systems do not have any other features. Therefore, the primary objective of putting this system into operation is to take cancer detection to a higher level by not only assessing whether or not there is skin cancer but also identifying the specific kind of skin cancer that may be present. The reason for this is that, in contrast to other forms of cancer, skin cancer may be classified into multiple subtypes, and the symptoms and therapies for each subtype of skin cancer are significantly different.

Therefore, determining whether or not the person has skin cancer is not sufficient at this point. As a result, regarding skin cancer, it is essential to know the precise form of cancer that the person has been detected with. The three cancer types that have the highest chance of being identified for that particular submitted image of the skin tumor will be identified by this model if it concludes that the provided image is skin cancer. Upon completion of that stage, none of the existing systems contain any features. The user can keep track of their cancer's progress and show patients the cancer's current state as it spreads or contracts. This new capability is available in this system when the user has recognized the sort of cancer that they have. Monitoring the cancer consistently until it is entirely treated is an essential component, as the danger persists even after identification.

This approach will be used most of the time to achieve four primary goals as follows:

- In order to determine the level of risk, it is recommended that users provide answers to certain questions. Here are the results of the skin cancer diagnosis. In addition, you should give comments and directions based on the points that customers get.
- In order to evaluate whether a tumor is skin cancer or not, it is necessary to do a scan of the region of the tumor. If the tumor is considered cancer, it is necessary to identify the three forms of cancer that have the greatest likelihood of being identified.
- When the user is diagnosed with skin cancer, it is necessary to regularly monitor the region of the cancer that has been identified by comparing the scanned photos. Additionally, it is important to provide the patients with the live status of their disease, which includes whether it has grown or diminished.
- Here is a separate part where you may learn more about skin cancers, including their symptoms and the ways to prevent them. Table 1 shows the skin cancer types with a number of images.

Table 1

Skin cancer types with a number of images

Types of skin cancer	Number of images
Actinic Keratoses	351
Basal cell carcinoma	521
Benign keratosis	1000
Dermatofibroma	125
Melanocytic nevi	6700
Melanoma	1100
Vascular	242

Overcoming the imbalanced data for each form of cancer was the main hurdle when utilizing this dataset. Constructing the model using data such as this would skew it toward a single form of cancer. Additionally, the limited quantity of data presented a barrier as well. We used data augmentation to overcome these challenges. The dataset has been increased to include approximately 40,000 photos in total, with nearly 6,000 images for each form of cancer. After augmenting the data, this expansion took place. The following is an example of a snapshot of the dataset. Fig.3 shows sample of skin cancer types.

The constructed model divides skin lesions into seven distinct categories. The construction of the model utilized MobileNet CNN as the architecture. A recently developed

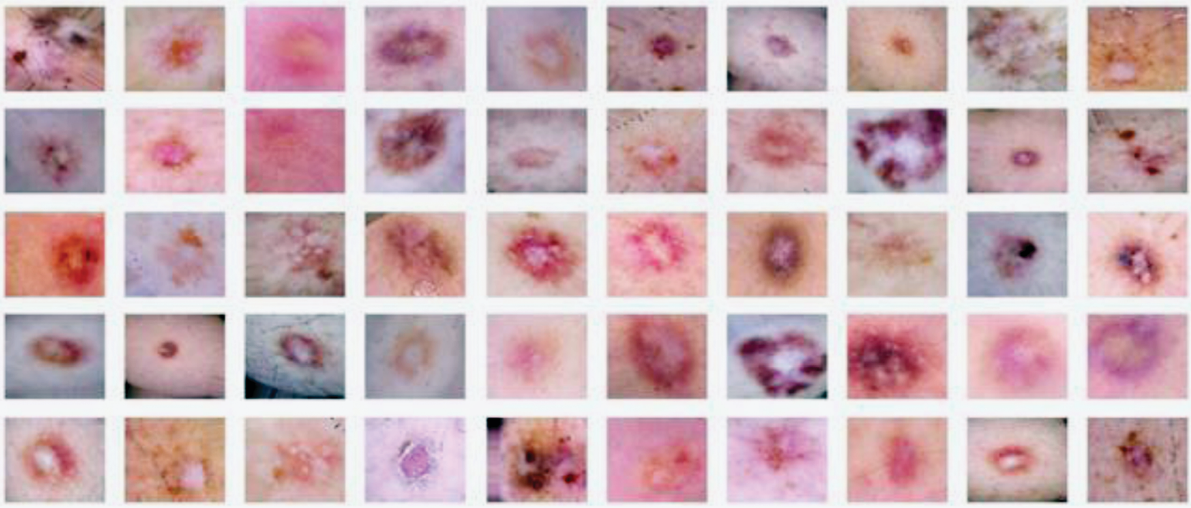

Fig. 3. Sample of skin cancer types

library known as Tensorflow.js was used to convert and run this model in the browser. The model includes multiple layers such as the ZeroPadding2D Layer, Conv2D Layer, Batch Normalization Layer, ReLu Activation DepthwiseConv2D Layer, and Dense Layer. The Adam Optimizer and the Categorical Cross-entropy Loss Function have been used in this process. Categorical cross-entropy has been used in this process. Dense and SoftMax activation have also been utilized.

3.4. Identification of Lung Cancer

Together with the location in which they are located, cancers may take on a variety of forms. During the research process, researchers separate the classified portion by categorizing tumors based on their features. The four most common types of malignancies seen all over the globe are lung cancers. Lung cancer is responsible for a significant number of fatalities that are associated with cancer. The time it takes to discover cancer indicates that delayed treatment may be a major factor in the rise in the death rate. This is a shocking reality. The great prevalence of smoking behaviors and the widespread pollution of the air across the globe are two additional factors that contribute to the development of lung cancer. As part of the investigation into the disease, there are some symptoms that are particular to lung malignancies that indicate that the individual has a tumor in their lung. Chest discomfort, dyspnea, exhaustion, supraclavicular lymphadenectasis, pain from metastasis, and fever are some of the indications linked to lung cancer. It is possible that the symptoms described above are more prevalent among individuals who are attempting to determine whether or not they are likely to be lung cancer patients. Individuals in stage IV may exhibit symptoms that include a greater proportion of chest discomfort, shortness of breath, dyspnea, weight loss, and exhaustion. These symptoms may differ from those experienced by individuals of various ages. Along with sexual behavior and certain smoking behaviors, these symptoms may take on a variety of forms. Constructing machine learning models using the specified features can predict the presence or absence of lung cancer. The accuracy of this model would be superior to that of older models since it would concentrate just on certain forms of data, such as numerical or visual data. Both kinds of datasets were discovered throughout the process of collecting data for this procedure by the study data gathering effort. Fig. 4 below shows sample of lung non-cancer (A) versus cancer (B) using MRI and the flow of the data process is shown in the Fig.5 that can be seen below.

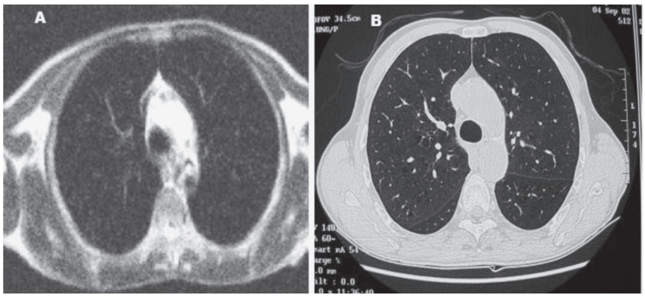

Fig. 4. Sample of lung non-cancer (A) versus cancer (B) using MRI

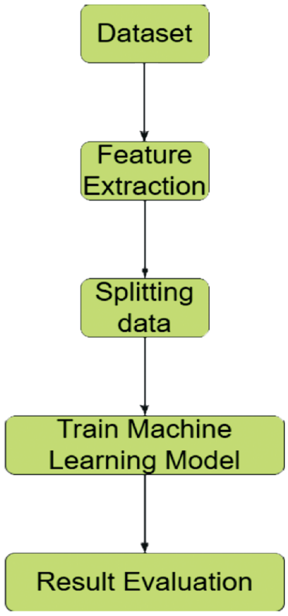

Fig.5. Flow chart of lung cancer detection

The recognized properties that may be used in the development of the model are the features that have been extracted. When you have finished picking characteristics, you should next continue with preprocessing the data in order to eliminate any error or null value. This will result in a reduction in the accuracy of the machine learning model. A number of different categorization strategies have been used for this numerical dataset. There are a number of different classification methods, some of which are linear regression, random forest classification, K-neighbors Neighbors Classifier, decision tree classification, gradient boosting classification, XGB classification, and support vector classification. When these classification approaches are taken into consideration, the Random Forest Classifier and the XGB Classifier produce the best results. From this point on, XGB will choose it since its performance is superior to that of the random forest classifier. Following this, give careful consideration to the alternative model that is dependent on the image dataset. Convolutional neural networks, sometimes known as CNNs, are techniques that are utilized for the categorization of images [9]. In addition to this approach, the following are some alternative algorithms that will be assessed: Past research indicates that the VGG-3 version will provide more accurate findings. The VGG-3 version will be used for further study. There is also the possibility of using the Keras sequential model for this image classification model. The Keras sequential model can achieve accuracy. Using ensemble learning models that incorporate both image and numerical data can achieve the goal of accurately recognizing lung malignancies. Both models and the approach that has been followed must be considered to accomplish this target. During the discussion of the challenges that need to be overcome, it was revealed that there might be some problems with the pre-processing of the data, but these problems were resolved. To determine the correctness of these models, it is necessary to analyze the predictions made by these models using actual data. To determine whether or not the system is capable of being used in medical settings, the predicted data should be compared from a relevant viewpoint.

4. Results and discussions

4.1. Breast Cancer Detection

The accuracy of the decision tree classifier was in the range of 95.3 %. When the models were applied to test data and recorded on a confusion matrix, the Random Forest Classifier achieved an accuracy rating of 96.7 %. As a result of comparing the accuracy of the random forest classifier to that of the training classifier, it is evident that the random forest classifier is the superior option for testing models. Table 2 below reveals the testing accuracies for breast cancer.

Table 2

Testing Accuracies

Algorithm	Accuracy (%)
Decision Tree Classifier	95.3
K-Neighbors Classifier	94.2
Logistic Regression	94.0
Random Forest Classifier	96.7

4.2. Brain Cancer Detection

A combination of binary and categorical cross-entropy functions was used in order to arrive at the following consequences: Based on the findings, the categorical cross-entropy function with 23 epochs produced the greatest accuracy (98 %) of all of the methods tested. Table 3 below reveals the accuracy attained by varying the number of epochs and loss functions; moreover, Fig. 6 and Fig. 7 have been seen as metrics of the brain tumor model's accuracy and model loss metrics for brain tumors, respectively.

Table 3

Accuracy attained by varying the number of epochs and loss functions

Loss Function	Epochs	Accuracy (%)
Binary Cross-entropy	10	96.2
Categorical Cross-entropy	20	97.6
Categorical Cross-entropy	30	98.2

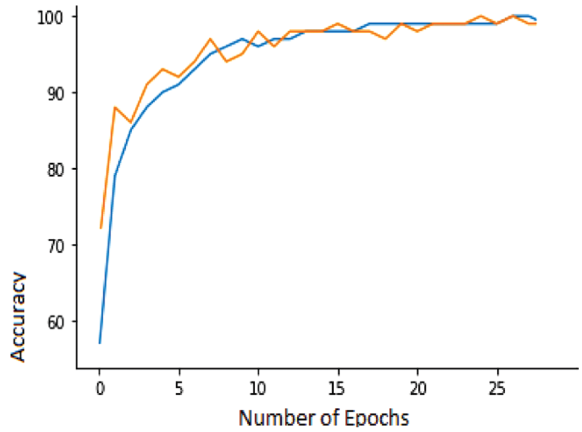

Fig. 6. Metrics of the brain tumor model's accuracy

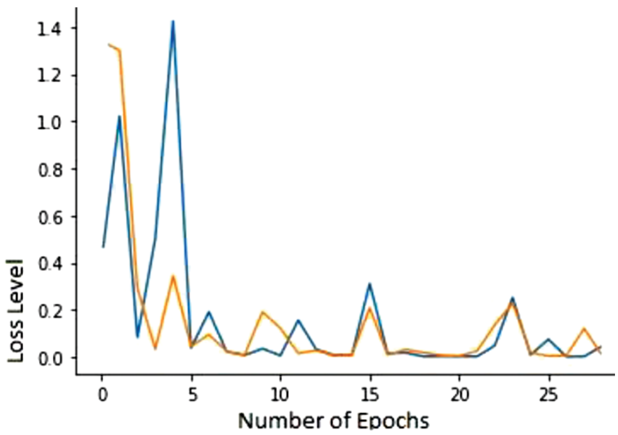

Fig. 7. Model loss metrics for brain tumors

4.3. Skin Cancer Detection

By the time it reached 30 epochs, the model had an outstanding accuracy rate of 97.8 %, and this was accomplished by utilizing categorical cross-entropy functions. Table 4 below shows the accuracy of the loss function.

Table 4

Accuracy of loss function
Accuracy of loss function

Loss Function	Epochs	Accuracy (%)
Categorical Cross-entropy	10	91.7
Categorical Cross-entropy	20	95.2
Categorical Cross-entropy	30	97.8

4.4. Lung Cancer Detection

The development of machine learning models yielded the findings shown in Table 5 below, with the subsequent table displaying the accuracy of each model. Random Forest Classifier and XGB Classifier are the two algorithms that research may choose to employ because, This study can choose the best and most precise algorithms based on this and the dataset that was utilized as the numerical dataset.

Table 5

Accuracy achieved by the use of several algorithms

Algorithm	Accuracy (%)
Linear Regression	62.0
K-Nearest Neighbors	92.6
XGB Classifier	99.2
Random Forest Classifier	99.2

5. Conclusion

The primary focus of this study is to develop prediction models that achieve a high level of accuracy in predicting the results of real diseases using supervised machine learning methods. The examination of the results suggests that combining multidimensional data with various classification, feature selection, and dimensionality reduction approaches might offer advantageous tools for inference in this particular area. To enhance the grouping methods and increase their ability to anticipate additional aspects, researchers should conduct further research in this area.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Izz K. abboud, Muaayed F. Al-Aawi, Nasir A. Al-Awad. Digital Medical Image Encryption Approach in Real-Time Applications. System Research & Information Technologies. 2024;1:26-32.
2. URL: http://www.breastcancer.org/symptoms/understand_bc/what_is_bc.
3. Hotko Y.S. Male Breast Cancer: Clinical Presentation, Diagnosis, Treatment. Exp Oncol. 2022;35:303-10.
4. URL: https://www.biospectrumindia.com/views/21/15300/statistical-analysisof-breast-cancer-in-india.html.
5. Malvia S., Bagadi S.A., Dubey U.S., Saxena S. Epidemiology of Breast Cancer in Indian Women. Asia Pac J Clin Oncol. 2019;13;4:289-295.
6. Devi R.D.H., Devi M.I. Outlier Detection Algorithm Combined with Decision Tree Classifier for Early Diagnosis of Breast Cancer. Int. J. Adv. Eng. Tech. 2021;5;2:251-259.
7. Muaayed F. Al-Rawi, Izz K. Abboud, Nasir A. Al-Awad. Novel Approach Using Transfer Deep Learning for Brain Tumor Prediction. Medical Radiology and Radiation Safety. 2021;69;3:81-85.
8. Miller K.D., Ostrom Q.T., C Kruchko., Patil N., Tihan T., Cioffi G., Fuchs H.E., Waite K.A., Jemal A., Siegel R.L., Barnholtz S..Brain and other Central Nervous System Tumor Statistics. A Cancer Journal for Clinicians. 2021;71;5:381-406.
9. Bienkowski M., Furtner J., Hainfellner J.A. Clinical Neuropathology of Brain Tumors. Handb Clin Neurol. 2022;145;477–534.
10. Lotlikar V.S., Satpute N., Gupta A. Brain Tumor Detection Using Machine Learning and Deep Learning: A Review. Current Medical Imaging. 2022;18;6:1-19.
11. Monika M.K., Vignesh N.A., Kumari C.U. Skin Cancer Detection and Classification Using Machine Learning. Materials Today: Proceedings. 2021;33;7:4266-4270.
12. Fransen M., Karahalios A., Sharma N., English D.R., Giles G.G., Sinclair R.D. Non-Melanoma Skin Cancer in Australia. Med J Aust. 2018;197:565–8.
13. Deinlein T., Richtig G., Schwab C., et al. The Use of Dermatoscopy in Diagnosis and Therapy of Nonmelanocytic Skin Cance. J Dtsch Dermatol Ges. 2021;14:144–51.
14. Ferris G.R., Treadway D.C., Perrewé P.L., Brouer R.L., Douglas C., Sean Lux. Political Skill in Organizations. Journal of Management. 2007;33:290-320.
15. Chaturvedi P., Jhamb A., Vanani M., Nemade V. Prediction and Classification of Lung Cancer Using Machine Learning Techniques. IOP Publishing Ltd, Jaipur, India. 2022;5;3:288-300.
16. Rahman S.P. a. H.Z. A New Method for Lung Nodule Detection Using Deep Neural Networks for CT Images. Int. Conf. on Electrical, Computer and Communication Engineering (ECCE). 2022:1-6.
17. Pehrson N.M. a. A.L.C. Automatic Pulmonary Nodule Detection Applying Deep Learning or Machine Learning Algorithms to the LIDC-IDRI Database. A Systematic Review Diagnostics. 2020;4;11:659-669.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.
Финансирование. Исследование не имело спонсорской поддержки.
Участие авторов. Статья подготовлена с равным участием авторов.
Поступила: 20.02.2025. Принята к публикации: 25.03.2025.

Conflict of interest. The authors declare no conflict of interest.
Financing. The study had no sponsorship.
Contribution. Article was prepared with equal participation of the authors.
Article received: 20.02.2025. Accepted for publication: 25.03.2025.