**LUNG CANCER PREDICTION**

**ABSTRACT**

Identifying the lung cancer through the machine learning techniques is challenging task. Prediction of lung cancer helps the medical-technological field factors that cause lung cancer and application of ML algorithms are discussed up to date and also draws special attention to their relative strengths and weaknesses. have been utilized because of its accurate outcomes. Our aim is to predict the lung cancer through the various machine learning algorithms for the given input dataset and the predict the accurate result. The prominent cause of cancer-related mortality throughout the globe is "Lung Cancer". Hence beforehand detection, prediction and diagnosis of lung cancer has become essential as it expedites and simplifies the consequent clinical board. To advance the progress and medication of cancerous conditions machine learning techniques have been utilized because of its accurate outcomes., have been applied in the healthcare sector for analysis and prognosis of lung cancer. In this review, factors that cause lung cancer and application of ML algorithms are discussed up to date and also draws special attention to their relative strengths and weaknesses.

**INTRODUCTION**

Lung cancer is a harmful disease that causes a huge number of deaths globally. The primal encounter of lung cancer is necessary to decrease the mortality rate of patients. Thus it is a great challenge encountered by doctors and researchers to detect and diagnose lung cancer. Detection of lung cancer can be done by using medical images such as computed tomography, chest X-ray; MRI scans, etc., ML approaches recognize the main characteristics of complex lung cancer datasets. A CAD (Computer-Aided Diagnosis) was developed in the early 1980s to enhance the survival rate and efficiency that aid the doctors in interpreting medical images. Some of the machine learning algorithms that have a profound impact in health care are SVM, K-nearest neighbors and ds so on. We have also discussed the deep learning methods techniques and algorithms that can be implemented for diagnosis, detection, and prediction of various cancers.

The preeminent intent of this research work is to present a concise vision of present work on different cancers and mainly lung cancer prediction using deep learning and machine learning models. Symptoms are categorized based on the location and size of the tumor . During the early stages, it's difficult to analyze and detect as it will not any cause any pain and symptoms in some cases. Lung cancer diagnosed patient may suffer through Cough, Chest pain, Shortness of breath, Wheezing, Hemoptysis i.e. coughing up blood, Pancoast syndrome (shoulder pain), Hoarseness (paralysis of vocal cords), Weight loss, Weakness, and Fatigue. Cancer, neurology, cardiology are the major parts of medicinal studies where AI is implemented. As this disease are superior in the mortality rate. Apart from these diseases, AI is even applied to other medicinal areas for prediction, analysis, and curing. Predominantly renowned ML algorithms extended in the healthcare sector are SVM,DS and KNN

**LUNG CANCER**

Lung cancer, also known as lung carcinoma, since about 98–99% of all lung cancers are carcinomas, is a malignant [lung tumor](https://en.wikipedia.org/wiki/Lung_tumor) characterized by uncontrolled [cell growth](https://en.wikipedia.org/wiki/Cell_growth) in [tissues](https://en.wikipedia.org/wiki/Tissue_(biology)) of the [lung](https://en.wikipedia.org/wiki/Lung). Lung carcinomas derive from transformed, malignant cells that originate as epithelial cells, or from tissues composed of epithelial cells. Other lung cancers, such as the rare sarcomas of the lung, are generated by the malignant transformation of connective tissues which arise from mesenchyme cells. Lymphomas and melanomas (from lymphoid and melanocyte cell lineages) can also rarely result in lung cancer.

In time, this uncontrolled [growth](https://en.wikipedia.org/wiki/Neoplasm) can spread beyond the lung – either by direct extension, by entering the lymphatic circulation, or via the hematogenous, blood borne spread – the process called [metastasis](https://en.wikipedia.org/wiki/Metastasis) – into nearby tissue or other, more distant parts of the body. Most [cancers](https://en.wikipedia.org/wiki/Cancer) that start in the lung, known as primary lung cancers, are [carcinomas](https://en.wikipedia.org/wiki/Carcinomas). The two main types are [small-cell lung carcinoma](https://en.wikipedia.org/wiki/Small-cell_lung_carcinoma) (SCLC) and [non-small-cell lung carcinoma](https://en.wikipedia.org/wiki/Non-small-cell_lung_carcinoma) (NSCLC).The most common [symptoms](https://en.wikipedia.org/wiki/Symptom) are coughing (including [coughing up blood](https://en.wikipedia.org/wiki/Hemoptysis)), weight loss, shortness of breath, and [chest pains](https://en.wikipedia.org/wiki/Chest_pain).The vast majority (85%) of cases of lung cancer are due to long-term [tobacco smoking](https://en.wikipedia.org/wiki/Tobacco_smoking). About 10–15% of cases occur in people who have never smoked.These cases are often caused by a combination of [genetic factors](https://en.wikipedia.org/wiki/Genetics) and exposure to [radon](https://en.wikipedia.org/wiki/Radon) gas, [asbestos](https://en.wikipedia.org/wiki/Asbestos), [second-hand smoke](https://en.wikipedia.org/wiki/Passive_smoking), or other forms of [air pollution](https://en.wikipedia.org/wiki/Air_pollution). Lung cancer may be seen on [chest radiographs](https://en.wikipedia.org/wiki/Chest_radiograph) and [computed tomography](https://en.wikipedia.org/wiki/Computed_tomography) (CT) scans.[]](https://en.wikipedia.org/wiki/Lung_cancer#cite_note-Merck-15) The [diagnosis](https://en.wikipedia.org/wiki/Medical_diagnosis) is confirmed by [biopsy](https://en.wikipedia.org/wiki/Biopsy), which is usually performed by [bronchoscopy](https://en.wikipedia.org/wiki/Bronchoscopy) or CT-guidance.

The major method of prevention is the avoidance of risk factors, including smoking and air pollution. Treatment and long-term outcomes depend on the type of cancer, the [stage](https://en.wikipedia.org/wiki/Staging_(pathology)) (degree of spread), and the person's overall health.Most cases are not curable. Common treatments include [surgery](https://en.wikipedia.org/wiki/Surgery), [chemotherapy](https://en.wikipedia.org/wiki/Chemotherapy), and [radiotherapy](https://en.wikipedia.org/wiki/Radiation_therapy). NSCLC is sometimes treated with surgery, whereas SCLC usually responds better to chemotherapy and radiotherapy.Worldwide in 2020, lung cancer occurred in 2.2 million people and resulted in 1.8 million deaths. It is the most common cause of cancer-related death in men and second-most common in women after [breast cancer](https://en.wikipedia.org/wiki/Breast_cancer). The most common age at diagnosis is 70 years. In the United States, [five-year survival rate](https://en.wikipedia.org/wiki/Five-year_survival_rate) is 20.5%,while in Japan it is 41.4%.Outcomes typically are worse in the developing world.

Many of the symptoms of lung cancer are not specific. In many people, the cancer has already spread beyond the original site by the time they have symptoms and seek medical attention.Symptoms that suggest the presence of metastatic disease include weight loss, bone pain, and neurological symptoms Common sites of spread include the brain, bone, [adrenal glands](https://en.wikipedia.org/wiki/Adrenal_gland), opposite lung, [liver](https://en.wikipedia.org/wiki/Liver), [pericardium](https://en.wikipedia.org/wiki/Pericardium), and [kidneys](https://en.wikipedia.org/wiki/Kidney). About 10% of people with lung cancer do not have symptoms at diagnosis; these cancers are [incidentally](https://en.wikipedia.org/wiki/Incidental_imaging_finding) found on routine [chest radiography](https://en.wikipedia.org/wiki/Chest_radiograph).

Depending on the type of tumor, [Para neoplastic phenomena](https://en.wikipedia.org/wiki/Paraneoplastic_syndrome) – symptoms not due to the local presence of cancer – may initially attract attention to the disease.In lung cancer, these phenomena may include hyperkalemia, [syndrome of inappropriate antidiuretic hormone](https://en.wikipedia.org/wiki/Syndrome_of_inappropriate_antidiuretic_hormone) (abnormally concentrated urine and diluted blood), ectopic [ACTH](https://en.wikipedia.org/wiki/ACTH) production, or [Lambert–Eaton my asthenic syndrome](https://en.wikipedia.org/wiki/Lambert%E2%80%93Eaton_myasthenic_syndrome) (muscle weakness due to [autoantibodies](https://en.wikipedia.org/wiki/Autoimmune_disorder)). Tumors in the [top of the lung](https://en.wikipedia.org/wiki/Apex_of_lung), known as [Pan coast tumors](https://en.wikipedia.org/wiki/Pancoast_tumor), may invade the local part of the [sympathetic nervous system](https://en.wikipedia.org/wiki/Sympathetic_nervous_system), resulting in [Horner's syndrome](https://en.wikipedia.org/wiki/Horner%27s_syndrome)  as well as damage to the [brachial plexus](https://en.wikipedia.org/wiki/Brachial_plexus).

**MACHINE LEARNING**

Machine learning (ML) is the study of computer [algorithms](https://en.wikipedia.org/wiki/Algorithm) that can improve automatically through experience and by the use of data.It is seen as a part of [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence). Machine learning algorithms build a model based on sample data, known as [training data](https://en.wikipedia.org/wiki/Training_data), in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, [email filtering](https://en.wikipedia.org/wiki/Email_filtering), [speech recognition](https://en.wikipedia.org/wiki/Speech_recognition), and [computer vision](https://en.wikipedia.org/wiki/Computer_vision), where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.A subset of machine learning is closely related to [computational statistics](https://en.wikipedia.org/wiki/Computational_statistics), which focuses on making predictions using computers; but not all machine learning is statistical learning. The study of [mathematical optimization](https://en.wikipedia.org/wiki/Mathematical_optimization) delivers methods, theory and application domains to the field of machine learning. [Data mining](https://en.wikipedia.org/wiki/Data_mining) is a related field of study, focusing on [exploratory data analysis](https://en.wikipedia.org/wiki/Exploratory_data_analysis) through [unsupervised learning](https://en.wikipedia.org/wiki/Unsupervised_learning). Some implementations of machine learning use data and [neural networks](https://en.wikipedia.org/wiki/Neural_networks) in a way that mimics the working of a biological brain. In its application across business problems, machine learning is also referred to as [predictive analytics](https://en.wikipedia.org/wiki/Predictive_analytics).

The term machine learning was coined in 1959 by [Arthur Samuel](https://en.wikipedia.org/wiki/Arthur_Samuel), an American [IBMer](https://en.wikipedia.org/wiki/IBMer) and pioneer in the field of [computer gaming](https://en.wikipedia.org/wiki/Computer_gaming) and [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence). Also the synonym self-teaching computers was used in this time period.A representative book of the machine learning research during the 1960s was the Nilsson's book on Learning Machines, dealing mostly with machine learning for pattern classification. In 1981 a report was given on using teaching strategies so that a [neural network](https://en.wikipedia.org/wiki/Neural_network) learns to recognize 40 characters from a computer terminal.

[Tom M. Mitchell](https://en.wikipedia.org/wiki/Tom_M._Mitchell) provided a widely quoted, more formal definition of the algorithms studied in the machine learning field: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E. This definition of the tasks in which machine learning is concerned offers a fundamentally [operational definition](https://en.wikipedia.org/wiki/Operational_definition) rather than defining the field in cognitive terms.Modern day machine learning has two objectives, one is to classify data based on models which have been developed, the other purpose is to make predictions for future outcomes based on these models. A hypothetical algorithm specific to classifying data may use computer vision of moles coupled with supervised learning in order to train it to classify the cancerous moles. Whereas, a machine learning algorithm for stock trading may inform the trader of future potential predictions.

**SVM**

In [machine learning](https://en.wikipedia.org/wiki/Machine_learning), support-vector machines are [supervised learning](https://en.wikipedia.org/wiki/Supervised_learning) models with associated learning [algorithms](https://en.wikipedia.org/wiki/Algorithm) that analyze data for [classification](https://en.wikipedia.org/wiki/Statistical_classification) and [regression analysis](https://en.wikipedia.org/wiki/Regression_analysis). Developed at [AT&T Bell Laboratories](https://en.wikipedia.org/wiki/AT%26T_Bell_Laboratories) by [Vladimir Vapnik](https://en.wikipedia.org/wiki/Vladimir_Vapnik) with colleagues Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-[probabilistic](https://en.wikipedia.org/wiki/Probabilistic_classification) [binary](https://en.wikipedia.org/wiki/Binary_classifier) [linear classifier](https://en.wikipedia.org/wiki/Linear_classifier). SVM maps training examples to points in space so as to maximize the width of the gap between the two categories. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.In addition to performing [linear classification](https://en.wikipedia.org/wiki/Linear_classifier), SVMs can efficiently perform a non-linear classification using what is called the [kernel trick](https://en.wikipedia.org/wiki/Kernel_method#Mathematics:_the_kernel_trick), implicitly mapping their inputs into high-dimensional feature spaces.

When data are unlabeled, supervised learning is not possible, and an [unsupervised learning](https://en.wikipedia.org/wiki/Unsupervised_learning) approach is required, which attempts to find natural [clustering of the data](https://en.wikipedia.org/wiki/Cluster_analysis) to groups, and then map new data to these formed groups. The support-vector clustering[[2]](https://en.wikipedia.org/wiki/Support-vector_machine#cite_note-HavaSiegelmann-2) algorithm, created by [HaveSiegel Mann](https://en.wikipedia.org/wiki/Hava_Siegelmann) and [Vladimir Vapnik](https://en.wikipedia.org/wiki/Vladimir_Vapnik), applies the statistics of support vectors, developed in the support vector machines algorithm, to categorize unlabeled data, and is one of the most widely used clustering algorithms in industrial applications. More formally, a support-vector machine constructs a [hyper plane](https://en.wikipedia.org/wiki/Hyperplane) or set of hyper planes in a [high-](https://en.wikipedia.org/wiki/High-dimensional_space) or infinite-dimensional space, which can be used for [classification](https://en.wikipedia.org/wiki/Statistical_classification), [regression](https://en.wikipedia.org/wiki/Regression_analysis), or other tasks like outliers detection.[[3]](https://en.wikipedia.org/wiki/Support-vector_machine#cite_note-3) Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training-data point of any class since in general the larger the margin, the lower the [generalization error](https://en.wikipedia.org/wiki/Generalization_error) of the classifier.

**KNN**

The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. While it can be used for either regression or classification problems, it is typically used as a classification algorithm, working off the assumption that similar points can be found near one another.

For classification problems, a class label is assigned on the basis of a majority vote—i.e. the label that is most frequently represented around a given data point is used. While this is technically considered “plurality voting”, the term, “majority vote” is more commonly used in literature. The distinction between these terminologies is that “majority voting” technically requires a majority of greater than 50%, which primarily works when there are only two categories.

**DS**

Data structures are essential tools for efficient data storage, manipulation, and retrieval in computer programming. Here are some common data structures and algorithms used in programming:Arrays are a collection of elements of the same data type stored in contiguous memory locations. They are commonly used to store and manipulate large amounts of data.Linked lists are a collection of nodes that contain data and a pointer to the next node. They are used to store data that can be dynamically resized and are often used to implement other data structures such as stacks and queues.Stacks are a collection of elements that can be added and removed only at one end. They follow the Last In First Out (LIFO) principle and are used in solving problems that involve recursion or nested function calls.Queues are a collection of elements that can be added at one end and removed from the other end. They follow the First In First Out (FIFO) principle and are used in solving problems that involve processing data in the order it is received.Trees are hierarchical data structures consisting of nodes connected by edges. They are used to represent hierarchical relationships between data elements. Graphs are a collection of nodes (vertices) and edges that connect them. They are used to represent complex relationships between data elements, such as social networks or road networks.Sorting algorithms are used to arrange data elements in a specific order, such as ascending or descending. Common sorting algorithms include Bubble Sort, Selection Sort, Insertion Sort, Quick Sort, and Merge Sort.Searching algorithms are used to find a specific element in a data structure. Common searching algorithms include Linear Search, Binary Search, and Depth-First Search.In summary, data structures and algorithms are essential tools for efficient programming. Some common data structures include arrays, linked lists, stacks, queues, trees, and graphs. Sorting and searching algorithms are also essential for manipulating and retrieving data efficiently.

**LITERATURE REVIEW**

**OPTIMAL DEEP LEARNING MODEL FOR CLASSIFICATION OF LUNG CANCER ON CT IMAGES**

SachiNandanmohanty et.al., has proposed. In this paper Lung cancer is one of the dangerous diseases that cause huge cancer death worldwide. Early detection of lung cancer is the only possible way to improve a patient's chance for survival. A Computed Tomography (CT) scan used to find the position of tumor and identify the level of cancer in the body. The current study presents an innovative automated diagnosis classification method for Computed Tomography (CT) images of lungs. In this paper, the CT scan of lung images was analyzed with the assistance of Optimal Deep Neural Network (ODNN) and Linear Discriminate Analysis (LDA). The deep features extracted from a CT lung images and then dimensionality of feature is reduced using LDR to classify lung nodules as either malignant or benign. The ODNN is applied to CT images and then, optimized using Modified Gravitational Search Algorithm (MGSA) for identify the lung cancer classification The comparative results show that the proposed classifier gives the sensitivity of 96.2%, specificity of 94.2% and accuracy of 94.56%.Medical image analysis has extraordinary supremacy in the field of health sector, particularly in noninvasive treatment and clinical examination .

The acquired restorative images such as X-rays, CT, MRI, and ultrasound imaging are used for specificdiagnosis . In medical imaging, CT is one of the filtering mechanism which use attractive fields to capture images in films . Lung cancer is one-of-its-kind of cancer that leads to 1.61 million deaths per year. In Indonesia, lung cancer is ranked in the third position among the prevalent cancers, for the most part, found in the MIoT centers. The survival rate is higher if the cancer is diagnosed at the beginning stages. The early discovery of lung cancer is not a simple assignment. Around 80% of the patients are diagnosed effectively only at the center or propelled phase of cancer. Lung cancer is positioned second among males and tenth among females globally. The information given in these studies is a general portrayal of lung cancer location framework that contains four basic stages. The lung cancer is the third most frequent cancer in women, after breast and colorectal cancers. Feature extraction process is one of the simplest and efficient dimensionality reduction techniques in image processing. One of the striking features of CT imaging is its non-obtrusive character. The rise of angles, which might be viewed, is odd when compared to parallel imaging modalities. The selected or extracted features set will extract the relevant information from the input data to the reduction process . The reduced features are assigned to a support vector machine for the purpose of training and testing. The models used for lung cancer image classification are neural network models with finalization image pre-processing . The existing research work for lung cancer classification was performed using a neural network model which provided 80% accuracy. Various investigations have been conducted regarding lung cancer classification and Classifiers, for example, ‘SVM, KNN and DS. The SVM is a universal useful learning method based on statistical learning hypothesis. However, these techniques are expensive and detect lung cancer at its advanced stages due to which the chance for survival is very low. The early detection of cancer can be helpful in curing the disease completely. So, the requirement of developing a technique to detect the occurrence of cancerous nodule in the early stage is increasing. The contribution of the current work considers two important phases: First phase is the CT lung cancer classification processes where the selected features are extracted to LDA reduction process and in the second phase, optimal deep learning classifier with MGSA optimization algorithm is used to classify the CT lung cancer images.[1]

**MULTI-STAGE LUNG CANCER DETECTION AND PREDICTION USING MULTI-CLASS SVM CLASSIFIER**

JaneeAlamet.al., has proposed. In this paper Recognition and prediction of lung cancer in the earliest reference point stage can be very useful to improve the survival rate of patients. But diagnosis of cancer is one the major challenging task for radiologist. For detecting, predicting and diagnosing lung cancer, an intelligent computer-aided diagnosis system can be very much useful for radiologist. This paper proposed an efficient lung cancer detection and prediction algorithm using multi-class SVM (Support Vector Machine) classifier. Multi-stage classification was used for the detection of cancer. This system can also predict the probability of lung cancer. In every stage of classification image enhancement and segmentation have been done separately. Image scaling, color space transformation and contrast enhancement have been used for image enhancement. Threshold and marker-controlled watershed based segmentation has been used for segmentation. For classification purpose, SVM binary classifier was used. Our proposed technique shows higher degree of accuracy in lung cancer detection and prediction.

Lung cancer is the major cause of cancer death in the world. The symptoms of lung cancer come into light at the final stage. So it is very tough to identify in its beginning stage. For this reason, the death percentage is very high for lung cancer in comparison with all other types of cancer. The two kind of lung disease which develop and spread in an unexpected way, are little cell lung malignancies (SCLC) and non-little cell lung tumors (NSCLC) . The phase of lung disease alludes to the degree to which the growth has spread in the lung. According to a statistics conducted by world health organization that every year more than 7.6 million people died of lung cancer. Moreover, the death rates of lung cancer are expected upon to keep rising, to wind up around 17 million worldwide in 2030. We found that lung cancers deaths in Bangladesh reached 9,660 or 1.33% of total deaths, according to the latest WHO data published. In year of 2005, around 1,362,825 new cancer cases are expected and around 571,590 deaths are expected to happen due to cancer in the United States. It was evaluated that there will be 162,921 deaths from lung cancer, which occurs 30% of all cancer deaths . There have been reported researches for cancer cell detection in recent year. Murphy et al. built up a CAD framework, where lungs pictures were divided by utilizing the region growing technique and morphological smoothing. The algorithm had an accuracy of 84%. Ye et al.proposed another algorithm to improve the location of knobs with ground-glass opacity. Messay, Hardie and Rogers displayed a CAD algorithm utilizing thresholding, morphological handling and Fisher Linear Discriminant to fragment, recognize patients nodules and take out of false positives. The framework got an accuracy of 82.66% with 3 FP per case being validated with 143 knobs. Gomathi and Thangaraj utilized image processing algorithm, Fuzzy CMean calculation and neural classifier in the phases of preprocessing, fragmentation identify patients nodules and respectively. This algorithm had an accuracy of 76.9%. Kumar et al. proposed a CAD algorithm that utilized Biorthogonal Wavelet Transform, region growing and fuzzy based framework in preprocessing, fragmentation and identification of nodules. The algorithm had an accuracy of 86%. In our proposed algorithm we have tried to solve these problems. Our developed algorithm can detect cancer affected cell and the corresponding stage such as initial, middle, or final stage. If no cancer affected cell is found in the input image then it checks the probability of lung cancer.[2]

**AUTOMATIC LUNG CANCER PREDICTION FROM CHEST X-RAY IMAGES USING THE DEEP LEARNING APPROACH**

ArjareeThirachet.al., has proposed. In this paper Since, cancer is curable when diagnosed at an early stage, lung cancer screening plays an important role in preventive care. Although both low dose computed tomography (LDCT) and computed tomography (CT) scans provide greater medical information than normal chest x-rays, access to these technologies in rural areas is very limited. There is a recent trend toward using computer-aided diagnosis (CADx) to assist in the screening and diagnosis of cancer from biomedical images. In this study, the 121-layer convolutional neural network, also known as DenseNet121 by G. Huang et. al., along with the transfer learning scheme is explored as a means of classifying lung cancer using chest xray images. The model was trained on a lung nodule dataset before training on the lung cancer dataset to alleviate the problem of using a small dataset. The proposed model yields 74.43±6.01% of mean accuracy, 74.96±9.85% of mean specificity, and 74.68±15.33% of mean sensitivity. The proposed model also provides a heatmap for identifying the location of the lung nodule. These findings are promising for further development of chest x-ray-based lung cancer diagnosis using the deep learning approach. Moreover, they solve the problem of a small dataset. As reported by WHO, cancer caused approximately 8.8 million deaths in 2015 . Almost 20% or 1.69 million of these deaths were due to lung cancer . Cancer screening plays an important role in preventive care because it is most treatable when caught in the early stages. This study shows that the appearance of malignant lung nodules more commonly demonstrate a spiculated contour, lobulation, and inhomogeneous attenuation .

Presently, low dose computed tomography (LDCT) plays an important role in lung cancer screening. LDCT screening has reduced lung cancer deaths and is recommended for high-risk demographic characteristics. Results from LDCT screening may be further evaluated with standard dose computed tomography (CT) . However, there are many barriers to implementing LDCT screening, such as providers’ anxiety concerning the access to LDCT equipment and the potential financial burden on rural populations . Moreover, rural populations have limited access to both primary care physicians and specialists . On the other hand, chest xrays are readily available in rural areas. Nonetheless, chest xrays produce lower quality images compared to LDCT or CT scans and, therefore, a lower quality diagnosis is generally expected. This study explores the use of chest x-rays with a computer-aided diagnosis (CADx) system to improve lung cancer diagnostic performance. The convolutional neural network (CNN) is proven to be very effective in image recognition and classification tasks. The development of CNNs starts from, LeNet,AlexNet , ZFNet , VGG , Inception ResNet, Inception-ResNet , Xception , DenseNet , and NASNet. There are many studies on the use of deep CNNs to detect abnormalities in chest x-rays. For instance, M. T. Islam et al., use several CNNs to detect abnormalities in chest x-rays. There is also a study by X. Wang et al., on the use of CNNs to detect thoracic pathologies from chest x-ray images. Their study also provides a large dataset as is the case in this study. Among current research, some studies on the application of Densely Connected Convolutional Networks (DenseNet) [14] to detect thoracic pathologies such as ChexNetand the Attention Guided Convolutional Neural Network (AG-CNN) . Both studies train the neural network on a very large chest x-ray image dataset.[3]

**IMAGE-BASED SURVIVAL PREDICTION FOR LUNG CANCER PATIENTS USING CNNS**

ChristophHaarburgeret.al., has proposed. In this paper Traditional survival models such as the Cox proportional hazards model are typically based on scalar or categorical clinical features. With the advent of increasingly large image datasets, it has become feasible to incorporate quantitative image features into survival prediction. So far, this kind of analysis is mostly based on radiomics features, i.e. a fixed set of features that is mathematically defined a priori. To capture highly abstract information, it is desirable to learn the feature extraction using convolutional neural networks. However, for tomographic medical images, model training is difficult because on the one hand, only few samples of 3D image data fit into one batch at once and on the other hand, survival loss functions are essentially ordering measures that require large batch sizes. In this work, we show that by simplifying survival analysis to median survival classification, convolutional neural networks can be trained with small batch sizes and learn features that predict survival equally well as end-to-end hazard prediction networks.

The medical image computing (MIC) community has been influenced strongly by advancements in machine learning and computer vision. Public availability of large annotated datasets has highly improved applicability and reproducibility of deep learning in MIC. As a result, the state of the art in computer aided diagnosis and detection as well as segmentation of medical images is currently dominated by convolutional neural networks (CNNs). A MIC subfield that has not seen such a strong benefit from these methods yet is survival prediction based on medical images. Survival analysis and prediction have been influenced mostly from biostatistics, i.e. statistical modeling based on nonimage data. Motivated by the recent success of radiomics there has been increasing interest in image-based survival analysis.Survival analysis refers to the study of the time-to-event data for an individual or the study of the distribution of those times for a cohort. Typical events in a medical context are death, disease incidence or relapse from remission. Usually, regression modeling strategies cannot be applied to survival data since although for each patient, a time-to-event is specified, those events may be qualitatively different. For some patients, the time indicated is the time-to-event, for others it indicates the time of the last follow-up before leaving the study. This is referred to as right-censoring and indicated by the event indicator δi that equals 1 if the event occurred and 0 for censoring.

We presented a method for survival prediction based on tomographic medical images. Our method can leverage trainable CNN features from CT image data, capturing abstract image information as well as clinical features in a single model. We show that by simplifying survival analysis to median survival classification, CNNs can be trained with small batch sizes and learn features that predict survival equally well as end-to-end hazard prediction networks and outperform the previous radiomics approach. This is a crucial step towards large scale image-based survival analysis that will allow survival prediction for more complex image data such as 3D+t images in the future.[4]

**LUNG CANCER SCREENING WITH LOW-DOSE CT SCANS USING A DEEP LEARNING APPROACH**

Jason L. Causey et.al., has proposed. In this paper Lung cancer is the leading cause of cancer deaths. Early detection through low-dose computed tomography (CT) screening has been shown to significantly reduce mortality but suffers from a high false positive rate that leads to unnecessary diagnostic procedures. Quantitative image analysis coupled to deep learning techniques has the potential to reduce this false positive rate. We conducted a computational analysis of 1449 low-dose CT studies drawn from the National Lung Screening Trial (NLST) cohort. We applied to this cohort our newly developed algorithm, Deep Screener, which is based on a novel deep learning approach. The algorithm, after the training process using about 3000 CT studies, does not require lung nodule annotations to conduct cancer prediction. The algorithm uses consecutive slices and multi-task features to determine whether a nodule is likely to be cancer, and a spatial pyramid to detect nodules at different scales. We find that the algorithm can predict a patient’s cancer status from a volumetric lung CT image with high competition, based on the challenge datasets. We report here the application of Deep Screener on an independent NLST test set. This study indicates that the deep learning approach has the potential to significantly reduce the false positive rate in lung cancer screening with low-dose CT scans.

Lung cancer is the leading cause of cancer deaths and the second most common cancer in both men and women in the United States . Because lung cancer is most often diagnosed at an advanced stage, the overall 5-year survival is poor Early detection is key to early intervention leading to improved survival. Compared with radiographs, low-dose CT can provide more detailed information and has been reported to lead to a 20% lower mortality rate . Low dose CT has been recommended for lung cancer screening by the US Preventive Services Task Force. Lung cancer screening studies based on analysis by human experts have reported false positive rates as high as 58% . This not only increases the cost for further tests and surgical procedures, but also causes unnecessary anxiety for patients and their families. The development of powerful computer-aided approaches for lung cancer early screening is critical to improve the current 3 clinical practice of CT imaging assessment. Computer-aided approaches are expected to produce automated solutions for early lung cancer screening and a reduced false positive rate in diagnosis. Numerous computer-aided approaches have been developed for chest image analysis in the past fifty years.

Ginnekenreviews computer analysis in chest imaging and illustrates how the three types of approaches — rule-based image processing, machine learning, and deep learning — have been applied, and how deep learning is currently becoming the dominant approach with very promising results . Most computational approaches to date such as , focus on finding and analyzing nodules in lung CT images. This dependency on predefined objects of interest requires detection and segmentation steps that are difficult to automate and limit the applicability of these approaches for automated screening. We have developed a deep learning approach we call Deep Screener, based on a convolutional neural network (CNN) model. The algorithm uses a pseudo-3-D model considering context information of consecutive slices of a participant’s lungs. Deep Screener is an automated end-to end solution for lung cancer screening that does not require a-priori lung nodule annotations. The preliminary framework of the algorithm ranked 16th of 1972 teams (top 1%) in the DSB2017 competition [14]. We evaluated the algorithm’s ability to generalize beyond the competition dataset by applying it to an independent cohort drawn from the National Lung Screening Trial (NLST) [5].

**NONPROPORTIONALITY OF SCINTILLATOR DETECTORS. V. COMPARING THE GAMMA AND ELECTRON RESPONSE**

Patrick R. Beck et.al., has proposed. In this paper This paper is the fifth in a series of articles on the basic physics of light yield no proportionality in scintillators. Here, we compare and contrast the no proportionality as registered by gamma rays and high-energy electrons. As has been noted in the past, these two types of data have different curve shapes (for plots of the light yield against electron or gamma energy). Herein, we show how the experimental gamma no proportionality curve can be calculated from the electron response by accounting for the distribution of high energy electrons created by the gamma photon via the photoelectric interaction. Similarly, we measure and model the gamma-induced resolution as a function of energy and compare this data to predictions from our model. The utility of the model is explored using data acquired with the scintillators (Eu), GYGAG(Ce) and CsI(Na). R ESEARCHERS most often report the gamma no proportionality of scintillators in the literature, since this data can be generated through the use of a number of different radioactive sources that are generally available . This data is a plot of the gamma-induced light output as a function of the incident gamma energy. On the other hand, Valentine and co-workers reported on the electron nonproportionality, based on their use of the Compton Coincidence Technique. Here they noted that the electron response had a related but somewhat different shape than the gamma response, and they accounted for this difference in their analysis . The difference between the two types of nonproportionality data sets is expected, since the gamma ray generates a specific distribution of high energy electrons upon photoelectric absorption by the medium while, in contrast, the electron nonproportionality data is acquired by generating a single high energy electron.

Herein, the experimental measurements are performed with a second-generation instrument, built with our colleagues at Lawrence Berkeley National Laboratory and known as the SLYNCI (Scintillator Light Yield No proportionality Characterization Instrument); this Compton spectrometer was designed to acquire data considerably more rapidly . The principle of operation involves the measurement of coincidences between scintillation events in the test scintillator and a Compton-scattered gamma ray detected in the surrounding germanium detectors, where the energy of the deposited mono-energetic electron is the difference between the scattered and incident gamma photons (usually generated by a source). More recently, additional instruments that measure no proportionality have been set up by the research groups of Moszynski ,Dorenbos and Melcher . The physics of scintillator no proportionality has been a pressing issue in the scientific literature of radiation detectors for many years , as this phenomenon generally limits the achievable energy resolution. As mentioned above, most of the nonproportionality data is deduced through gamma spectroscopy (“gamma nonproportionality”), while there are fewer reports of “electron nonproportionality”. The electron and gamma nonproportionality were measured side-by-side for the same scintillator in some cases , and the differences between the two curves have been observed and acknowledged to be an expected physical effect. Despite these measured differences, they are often spoken of in an interchangeable manner, even though they arise from a significantly different physics basis. As noted, the difference between gamma and electron no proportionality is a result of several electrons being created by the gamma ray in contrast to the single mono-energetic electron generated in instruments such as the SLYNCI. The main purpose of this paper is to measure and relate the electron and gamma light yield no proportionality curves using an analytical formulation, and to calculate the gamma light yield no proportionality curve from the measured electron response. We then proceed to also model the gamma energy resolution, although the resolution model must necessarily include an estimate of the contribution from photon statistics. This work is performed for three scintillator materials: (Eu),GYGAG(Ce) and CsI(Na) , selected for their different no proportionality curve shapes. The model reasonably reproduces the data of all three scintillators, with the exception that it poorly replicates the resolution of CsI(Na); we attempt to identify the physics issue that may have been overlooked for CsI(Na).[6]

**A SURVEY OF MEDICAL IMAGE ANALYSIS USING DEEP LEARNING APPROACHES**

AasiaRehmanet.al., has proposed. In this paper With the expanding development of Deep Learning techniques Medical Image Analysis have become an active field of research. Medical Image Analysis typically refers to the utilization of various kinds of image modalities and techniques to obtain images of the human body which in turn can be used by medical experts for diagnosis along with treatment of patients. This paper provides a survey of various improvements that have been made in Medical Image Analysis using DL techniques related to different pattern recognition tasks. These pattern recognition tasks include Classification, Detection/Localization, Segmentation, and Registration. . The paper discusses several recently published research papers related to different pattern recognition tasks including liver lesion classification and segmentation, lung nodule detection & classification, lung nodule segmentation, brain tumor classification, and detection, brain tumor segmentation, Breast cancer detection, etc. Comparative description of these papers is also provided in terms of organ, modality, dataset, model used and limitation/improvements needed. This survey briefly describes several medical imaging modalities used in medical image. Also, the proposed research work has evaluated various challenges encountered in the Medical Imaging domain and have discussed about the current trends for new researchers/ medical instrument experts encouraging them to take full advantage of DL techniques in the future. Deep Learning has attained promising outcomes in a wide variety of scientific fields including Image Recognition , Natural Language Processing , Robotics Speech Recognition , etc. A deep learning framework permits the machine to learn complex computational models for feature representation which is further used to perform data analytics accurately. Linear and non-linear activations are carried out on the input data hierarchically by adjusting various parameters (weights). Using numerous hierarchical layers with non-linear activations the term Deep Learning was coined in which the main aim of the machine is to learn the parameters of a complex model utilizing the training samples in such a way that the machine constantly improves in performance. The concept of DL originated from the study of artificial neural networks (ANNs) made up of numerous layers of neurons and various weights (parameters) which indicate the robustness of the connections among neurons of different layers of ANN. The Deep Learning model is trained on a specific task using a particular type of training samples after which the model is able to fulfil the same task utilizing the test data which is new to the trained model. Deep learning models are trained using Backpropagationthat enables some sort of popular Gradient Descent algorithm that iteratively arrives at the optimal parameter values. These Deep learning machines are typically trained for hundreds of epochs prior to being deployed. Deep Learning is a subfield of Machine Learning which is basically a subfield of Artificial Intelligence. A lot of research has been already done in Machine Learning and Artificial-Intelligence like in. Medical Image Analysis has recently become an active area of research particularly with the expansion of Deep Learning. Although Machine Learning techniques have been applied in Medical Imaging since the 1960s the first notable contributions relating to modern Deep Learning methodologies appeared in Medical Imaging literature in the 1990s. These techniques are related to deep learning in the sense they applied ANN for performing Medical Imaging tasks however due to the limited amount of training data available, they have trained models having only two to three layers depth which is mostly not considered deep network nowadays. In the case of deep learning models the number of layers typically ranges from few to over hundreds. Computer Vision has developed such deep models for image processing. There is a close relationship between the computer vision field and Medical Imaging in analyzing digital images, as this field already has a long history of taking advantage of the findings in computer vision [7].

**A DEEP LEARNING-BASED APPROACH FOR THE DETECTION AND LOCALIZATION OF PROSTATE CANCER IN T2 MAGNETIC RESONANCE IMAGES**

RubaAlkadiet.al., has proposed. In this paper We address the problem of prostate lesion detection, localization, and segmentation in T2W magnetic resonance (MR) images. We train a deep convolutional encoder-decoder architecture to simultaneously segment the prostate, its anatomical structure, and the malignant lesions. To incorporate the 3D contextual spatial information provided by the MRI series, we propose a novel 3D sliding window approach, which preserves the 2D domain complexity while exploiting 3D information. Experiments on data from 19 patients provided for the public by the Initiative for Collaborative Computer Vision Benchmarking (I2CVB) show that our approach outperforms traditional pattern recognition and machine learning approaches by a significant margin. Particularly, for the task of cancer detection and localization, the system achieves an average AUC of 0.995, an accuracy of 0.894, and a recall of 0.928. The proposed mono-modal deep learning-based system performs comparably to other multi-modal MR-based systems. It could improve the performance of a radiologist in prostate cancer diagnosis and treatment planning. On a worldwide scale, prostate cancer (CaP) has been reported as the second most diagnosed cancer and the fifth leading cause of cancer deaths in men. In 2012, an estimated 1.1 million men were diagnosed with prostate cancer worldwide . Several screening and diagnostic tests are carried out in daily clinical routines to ensure early detection and treatment. Particularly, an increased Prostate Specific Antigen (PSA) level is usually followed by MR screening, and an ultra-sound (TRUS)-guided biopsy, respectively. Besides its noninvasive nature, magnetic resonance (MR) screening has shown high potential in early diagnosis, monitoring, and treatment planning of prostate cancer.

Interestingly, the detection rate of prostate cancer in MRI has been recently reported to be in the range of 0.44 to 0.87, which is higher than the reported finding of blind TRUS biopsy making it a suitable noninvasive alternative . Other key advantages of MR screening lies in its ability to provide information about the tumors’ location, volume, and level of malignancy. These characteristics are especially essential for the active surveillance, where indolent lesions that are confined in the prostate are regularly monitored to ensure that cancer cells are not developing in an irregular manner that necessitates medical intervention or other therapies . Basically, there are four main MRI modalities that are used in CaP diagnosis. These include T2 weighted (T2W), dynamic contrast enhanced (DCE), diffusion weighted (DW), and magnetic resonance spectroscopy (MRS). T2W MRI is the basic MRI modality that uses the transverse relaxation time T2 to construct a grayscale image of the scanned object. Due to its increasing popularity and availability by many health providers, T2W-MR images have been an effective tool for noninvasive CaP diagnosis. . The main advantage of this modality is that it allows to visually differentiate between normal prostatic tissues and cancerous tissues by means of intensity and homogeneity . More precisely, malignant tissues are characterized by lower signal intensity in the peripheral zone (PZ) of the prostate and a more homogeneous appearance in both the central gland (CG) and the PZ compared to the surrounding healthy tissues . This mainly results from the presence of a foci of congested glands (cancer) surrounded by less dense benign cells, which in turn translates to a region of low signal intensity in the image.

The other main advantage of this modality is that it encodes details of the zonal anatomy of the prostate gland. That is, the CG is well-distinguished from the PZ of the prostate and the surrounding non-prostatic tissues . In practice, MRI images are interpreted by an experienced radiologist who produces a report of findings for each case. Typically, a radiologist spends 14–15 years of post-high-school education and training before commencing his work as a radiologist . Because of the wellstructured nature of radiologists’ work which mainly relies on image analysis and interpretation, limited interaction exists between the radiologist and patients. Particularly, this sufficiently isolated and independent nature of work makes it an attractive candidate for computerized image processing. Mazurowski et al.have recently suggested that some existing computer vision algorithms could considerably: reduce the efforts required by radiologists, improve the quality of the interpretation by assessing new features in images that have not been previously assessed by humans, improve the repeatability of the decision, and reduce the time needed for image interpretation. These facts have fueled the need for developing an expert computer-aided MRI-based CaP detection system that competes well with experienced radiologists.[8]

**DEEP NEURAL NETWORKS FOR ANATOMICAL BRAIN SEGMENTATION**

Giovanni Montana et.al., has proposed. In this paper We present a novel approach to automatically segment magnetic resonance (MR) images of the human brain into anatomical regions. Our methodology is based on a deep artificial neural network that assigns each voxel in an MR image of the brain to its corresponding anatomical region. The inputs of the network capture information at different scales around the voxel of interest: 3D and orthogonal 2D intensity patches capture a local spatial context while downscaled large 2D orthogonal patches and distances to the regional centroids enforce global spatial consistency. Contrary to commonly used segmentation methods, our technique does not require any non-linear registration of the MR images. To benchmark our model, we used the dataset provided for the MICCAI 2012 challenge on multi-atlas labeling, which consists of 35 manually segmented MR images of the brain. We obtained competitive results showing the potential of our approach. To our knowledge, our technique is the first to tackle the anatomical segmentation of the whole brain using deep neural networks.

Quantitative research in neuroimaging often requires the anatomical segmentation of the human brain based on magnetic resonance images (MRIs). Quantitative research in neuroimaging often requires the anatomical segmentation of the human brain using magnetic resonance images (MRIs). For instance, abnormal volumes or shapes of certain anatomical regions of the brain have been found to be associated with brain disorders, including Alzheimer’s disease and Parkinson.The analysis of MR images is therefore essential to detect these disorders, monitor their evolution and evaluate possible treatments. The anatomical segmentation of the brain requires a segmentation protocol defining how each region should be delineated so that the resulting segmentations are comparable between brains. However, manually segmenting the brain is a time consuming and expensive process that cannot be performed at a large scale. Its full automation would enable systematic segmentation of MRIs on the fly as soon as the image is acquired. These potential benefits have encouraged an active field of research, which is today dominated by multi-atlas based and patch-based methods . Machine learning methods consist in training classifiers to assign each voxel (a 3D pixel) to its anatomical region based on various input features describing it, such as its neighborhood intensities or location. Recently, deep neural networks, and in particular convolutional neural networks, have proven to be the state of the art in many computer vision applications Contrary to traditional shallow classifiers in which feature engineering is crucial, deep learning methods automatically learn hierarchies of relevant features directly from the raw inputs .[9]

**PROSTATE SEGMENTATION IN MR IMAGES USING ENSEMBLE DEEP CONVOLUTIONAL NEURAL NETWORKS**

HaozheJiaet.al., has proposed. In this paper The automated segmentation of the prostate gland from MR images is increasingly used for clinical diagnosis. Since deep learning demonstrates superior performance in computer vision applications, we propose a coarse-to-fine segmentation strategy using ensemble deep convolutional neural networks (DCNNs) to address prostate segmentation in MR images. First, we use registration-based coarse segmentation on preprocessed prostate MR images to define the potential boundary region. We then train four DCNNs as voxel-based classifiers and classify the voxel in the potential region is a prostate voxel when at least three DCNNs made that decision. Finally, we use boundary refinement to eliminate the outliers and smooth the boundary. We evaluated our approach on the MICCAI PROMIS12 challenge dataset and our experimental results verify the effectiveness of the proposed algorithms. Statistics from National Cancer Institute indicate that greater than 220,800 men were diagnosed with prostate cancer in USA in 2015 and they identify prostate cancer as a major health threat. Magnetic resonance (MR) imaging is the principal imaging technology for evaluation of the prostate gland due to its superior spatial resolution and tissue contrast .

Determination of the prostate volume and extension through the capsule of the gland is important for diagnosis, management and prognosis. Currently, MR prostate segmentation is mainly done by radiologists based almost entirely on visual inspection on a slice-by-slice basis, which is time-consuming, requires a high degree of skill, concentration, and is prone to intra- and inter-operator bias. Many semi or fully automated methods have been proposed for segmentation of various organs or tissues from medical images, however, automated MR prostate segmentation remains a challenging task. The challenges are largely related to the variability in size/shape/contours of the gland, heterogeneity in signal intensity around end rectal coils (ERCs) when used, imaging artifacts and low contrast between the gland and adjacent structures. Recently, deep learning has been applied to a wide variety of problems, most prominently in computer vision, and it has been demonstrated that convolutional networks are advancing recognition. The availability of large annotated medical imaging data now makes it feasible to use deep convolutional neural networks (DCNNs) for medical image segmentation and classification . In this paper, we propose a coarse-to-fine prostate segmentation approach that is based upon two components – an atlas-based coarse segmentation and an ensemble DCNNbased fine segmentation. We first perform the DRAMMS algorithm to construct a probabilistic prostate atlas for coarse segmentation, and then train a set of DCNNs using the patches extracted from the boundary region obtained in coarse segmentation for fine segmentation. We evaluated our approach against several other state-of-the-art methods on the MICCAI PROMIS12 dataset.[10]

**ARCHITECTURE DIAGRAM**

Input dataset

Machine learning

Test score prediction

Roc curve

SVM,DS AND KNN

**SYSTEM SPECIFICATION**

**HARDWARE SPECIFICATION:**

* Processor Type : Pentium i3
* Speed : 3.40GHZ
* RAM : 4GB DD2 RAM
* Hard disk : 500 GB
* Keyboard : 101/102 Standard Keys
* Mouse : Optical Mouse

**SOFTWARE SPECIFICATION**:

* Operating System : Windows 10
* Front End : MATLAB

**SOFTWARE DESCRIPTION**

**ABOUT MATLAB**

Key Features

* Numeric Computation
* Data Analysis and Visualization
* Programming and Algorithm Development
* Application Development and Deployment
* Contact Sales
* Product Trial
* Pricing and Licensing

**Key Features**

High-level language for numerical computation, visualization, and application development Interactive environment for iterative exploration, design, and problem solving Mathematical functions for linear algebra, statistics, Fourier analysis, filtering, optimization, numerical integration, and solving ordinary differential equations Built-in graphics for visualizing data and tools for creating custom plots Development tools for improving code quality and maintainability and maximizing performance Tools for building applications with custom graphical interfaces Functions for integrating MATLAB based algorithms with external applications and languages such as C, Java, .NET, and Microsoft® Excel® Analyzing and visualizing data using the MATLAB desktop. Enlarge Analyzing and visualizing data using the MATLAB desktop. The MATLAB environment also lets you write programs and develop algorithms and applications.

**Numeric Computation**

MATLAB provides a range of numerical computation methods for analyzing data, developing algorithms, and creating models. The MATLAB language includes mathematical functions that support common engineering and science operations. Core math functions use processor-optimized libraries to provide fast execution of vector and matrix calculations.

**Available methods include:**

* Interpolation and regression
* Differentiation and integration
* Linear systems of equations
* Fourier analysis
* Eigenvalues and singular values
* Ordinary differential equations (ODEs)
* Sparse matrices

MATLAB add-on products provide functions in specialized areas such as statistics, optimization, signal analysis, and machine learning. Refinement of gridded data using 2-D cubic interpolation. Enlarge Refinement of gridded data using 2-D cubic interpolation.

Data Analysis and Visualization MATLAB provides tools to acquire, analyze, and visualize data, enabling you to gain insight into your data in a fraction of the time it would take using spreadsheets or traditional programming languages. You can also document and share your results through plots and reports or as published MATLAB code.

**ACQUIRING DATA**

MATLAB lets you access data from files, other applications, databases, and external devices. You can read data from popular file formats such as Microsoft Excel; text or binary files; image, sound, and video files; and scientific files such as netCDF and HDF. File I/O functions let you work with data files in any format.

Using MATLAB with add-on products, you can acquire data from hardware devices, such as your computer's serial port or sound card, as well as stream live, measured data directly into MATLAB for analysis and visualization. You can also communicate with instruments such as oscilloscopes, function generators, and signal analyzers.

A mixed numeric and text file for import into MATLAB using the Import Tool.

**ENLARGE**

A mixed numeric and text file for import into MATLAB using the Import Tool. MATLAB automatically generates a script or function to import the file programmatically.

**ANALYZING DATA**

MATLAB lets you manage, filter, and preprocess your data. You can perform exploratory data analysis to uncover trends, test assumptions, and build descriptive models. MATLAB provides functions for filtering and smoothing, interpolation, convolution, and fast Fourier transforms (FFTs). Add-on products provide capabilities for curve and surface fitting, multivariate statistics, spectral analysis, image analysis, system identification, and other analysis tasks.

Fitting a surface to data with a custom model using MATLAB and Curve Fitting Toolbox. Enlarge Fitting a surface to data with a custom model using MATLAB and Curve Fitting Toolbox.

**VISUALIZING DATA**

MATLAB provides built-in 2-D and 3-D plotting functions, as well as volume visualization functions. You can use these functions to visualize and understand data and communicate results. Plots can be customized either interactively or programmatically.

The MATLAB plot gallery provides examples of many ways to display data graphically in MATLAB. For each example, you can view and download source code to use in your MATLAB application**.**

**Features of MATLAB**

Following are the basic features of MATLAB −

* It is a high-level language for numerical computation, visualization and application development.
* It also provides an interactive environment for iterative exploration, design and problem solving.
* It provides vast library of mathematical functions for linear algebra, statistics, Fourier analysis, filtering, optimization, numerical integration and solving ordinary differential equations.
* It provides built-in graphics for visualizing data and tools for creating custom plots.
* MATLAB's programming interface gives development tools for improving code quality maintainability and maximizing performance.
* It provides tools for building applications with custom graphical interfaces.
* It provides functions for integrating MATLAB based algorithms with external applications and languages such as C, Java, .NET and Microsoft Excel.

**Uses of MATLAB**

MATLAB is widely used as a computational tool in science and engineering encompassing the fields of physics, chemistry, math and all engineering streams. It is used in a range of applications including −

* Signal Processing and Communications
* Image and Video Processing
* Control Systems
* Test and Measurement
* Computational Finance
* Computational Biology

### Structures

MATLAB has structure data types.[[15]](https://en.wikipedia.org/wiki/MATLAB#cite_note-15) Since all variables in MATLAB are arrays, a more adequate name is "structure array", where each element of the array has the same field names. In addition, MATLAB supports dynamic field names (field look-ups by name, field manipulations, etc.). Unfortunately, MATLAB JIT does not support MATLAB structures, therefore just a simple bundling of various variables into a structure will come at a cost.

### Functions

When creating a MATLAB function, the name of the file should match the name of the first function in the file. Valid function names begin with an alphabetic character, and can contain letters, numbers, or underscores. Functions are also often case sensitive.

### Function handles

MATLAB supports elements of lambda calculus by introducing function handles, or function references, which are implemented either in .m files or anonymous/nested functions.

### Classes and object-oriented programming

MATLAB supports object-oriented programming including classes, inheritance, virtual dispatch, packages, pass-by-value semantics, and pass-by-reference semantics.However, the syntax and calling conventions are significantly different from other languages. MATLAB has value classes and reference classes, depending on whether the class has *handle* as a super-class (for reference classes) or not (for value classes).

**EXISTING SYSTEM**

In existing system that presented the comparative analysis in the prognostic of hepatitis data using Rough set technique over Multi- layer Neural Network using back-propagation algorithm. The prediction of the outcome is more specific and accurate using Rough set technique. Performance and time taken to run the hepatitis data is fast in Naive Bayes algorithm. The results obtained were compared with other algorithms like, Naive Bayes up-datable algorithm, **FT Tree algorithm, Kstar algorithm, J48 algorithm, LMT algorithm and neural network**. Attributes were fully classified and the result obtained was of average performacne. Based on the experimental results the classification accuracy is found to be better using Naïve Bayes algorithm compared to other algorithms. In this work which cannot effectively distinguish between normal and anomalous instances on breast cancer dataset.Lung data distance measures between instances can be challenging when the data are complex.Cannot guaranteed critical nuggets information that belong to a data set may not be at a great “distance” from the other “normal” points, and may end up being classified as “normal.

**PROPOSED SYSTEM/METHODOLOGY**

Identification of the lung cancer using the machine learning techniques like **SVM,DS and KNN classifiers** all the algorithms are utilized in our project.So that the input lung cancer data set is given in the data visualization which consists of age, smoke, area-q and the alcohol four parameters will be tested.Based on data preprocessing training and testing values will be used and in logistic regression confusion Matrix will be classified so these algorithms produces the better results than previously existing models.

Improves the classificationaccuracy. It can provide to very close to the class boundary and are sensitive to small changes in attribute values.Best accuracy to classify test data information.High performance.Highest accuracy in the cross validation and machine learning is doneThe maximum separation for the machine learning techniques is made possibleThe res-net in the matching dimensions are increased to provide the better result.The effectiveness of cancer prediction system helps the people to know their cancer risk with low cost and it also helps the people to take the appropriate decision based on their cancer risk status. The data is collected from the website online lung cancer prediction system.Lung cancer dataset is given as the input.Certain algorithm such as SVM with KNN classifiers are used in our project.

**MODULEDESCRIPTION**

* **DATA PREPROCESSING**
* **MACHINE LEANING MODEL IMPLEMENTATION**
* **SVM, DS AND KNN**

**DATA PREPROCESSING**

* Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format.
* The data can have many irrelevant and missing parts. To handle this part, data cleaning is done. It involves handling of missing data, noisy data etc.

**MACHINE LEANING MODEL IMPLEMENTATION**

* Data sturucture, support vector machines, and KNN classifiers.
* All these machine learning models are given as the implementation model.
* Each model provide the specific output result for the each individual classifiers.

**SVM, DS AND KNN**

SVM, DS and KNN are popular machine learning algorithms that can be used for lung cancer prediction. In the context of lung cancer prediction, SVM, DS and KNN can be used to classify patients into either a lung cancer or non-lung cancer group based on their clinical data. To implement SVM, DS and KNN for lung cancer prediction, you would need to first gather a dataset of patient clinical data. This dataset would need to be preprocessed to extract relevant features and convert them into a suitable data format such as arrays or matrices. Once the data is in a suitable format, you can use SVM, DS or KNN modules in a machine learning library such as Scikit-learn to train and evaluate the models. For example, in Sickie-learn, you can use the SVC class to create an SVM model and the KNeighborsClassifier class to create a KNN model. These classes can be used to set hyper parameters such as the kernel function and number of neighbors, fit the models to the training data, and predict the class labels for new data points. In terms of data structures, both SVM, DS and KNN require efficient data structures for storing and manipulating the clinical data. For example, SVM may require feature scaling to ensure that all features are on the same scale, while KNN may require a distance matrix to efficiently compute the distances between data points. In summary, SVM, DS and KNN can be used for lung cancer prediction by classifying patients into either a lung cancer or non-lung cancer group based on their clinical data. To implement SVM, DS and KNN for lung cancer prediction, you would need to gather and preprocess a dataset of patient clinical data, use machine learning libraries such as Scikit-learn to train and evaluate the models, and use suitable data structures for storing and manipulating the clinical data.

**INPUT DESIGN**

The goal of designing input data is to make data entry as easy, logical and error free from errors as possible. In entering data, operators need to know the following: The allocated space for each field. Field sequence, which much match that in the source document. The format in which data fields are entered for example, filling out the date field is required through the edited format of parameters. When we approach input data design, we design the source document. Let us elaborate on each step.

Needless to say, therefore, that the input data is the lifeblood of a system and have to be analyzed and designed with at most case and consideration.

The decisions made during the input design are

* To provide effective method of input
* To achieve the highest possible level of accuracy
* To ensure that input is understand by the user.

Input design is a process of converting a user-oriented description of the input to the computer-based system. This design is important to avoid errors in the input process and show the correct direction to the management for getting the correct information from the computerized system.

A source document differs from a turnaround document in that the former contains data that change the status of a resource while the latter is a machine readable document. Transaction throughput is the number of error-free transactions entered during a specified time period. A document should be concise because longer documents contain more data and so take longer to enter and have a greater chance of data entry errors.

**OUTPUT DESIGN**

Output design generally refers to the results and information that are generated by the system for many end-users; it should be understandable with the enhanced format. The Output of the software is used to make the remote installation of the new software in the system and, it is awake the immediate alert to the system that should be enhanced it as the input to the system. Output is the main reason for developing the system and the basis on which they evaluate the usefulness of the application.

Computer output is the most important direct source of information to the user output design deals with form design efficient output design should improve the interfacing with user. The term output applies to any information produced by an information system in terms of displayed. When analyst design system output, they Identify the specific output that is needed to meet the requirements of end user. Previewing the output reports by the user is extremely important because the user is the ultimate judge of the quality of the output and, in turn, the success of the system

As the outputs are the most important sources of information to the users, better design should improve the system’s relationships with user and also will help in decision-making. Form design elaborates the way output is presented and the layout available for capturing information

**INPUT DATASET**

Lung cancer dataset is given as the input

The input can be splitted in to the training and the testing.

The testing dataset is used to provide the highest accuracy.

**EXPERIMENTAL SETUP**

For classification of the malignance of the detected lung nodules, traditional computer aided diagnosis (CAD) frameworks are widely employed, via image processing techniques. **SVM, DS AND KNN** classifier provides the highest classification accuracy.The machine learning feature-based CAD systems have plenty of flaws which restrict the further improvements. Old existing algorithm is also shown in the table below the results are mentioned for theoretical purpose only.

|  |  |
| --- | --- |
| **ALGORIHTM** | **ACCURACY** |
| DS | 96 |
| SVM | 65 |
| KNN | 50 |

**CONCLUSION**

In this project, we demonstrate a survey on lung cancer, its causes, symptoms, mortality rate due to cancer in India and throughout the Globe and deliberates the machine learning techniques, its applications in healthcare and cancer prognosis and detection. Most of the researchers developed the cancer prediction systems based on a supervised learning technique of ML and classification algorithms to produce an accurate outcome. Deep learning in health care and algorithms are emphasized. Prediction and diagnosis of the Lung cancer system can be embellished and extended further by employing deep learning techniques to enhance the accuracy of both identification and prediction of lung cancer.

**REFERENCES**

1. G. Lakshmanaprabu S.K., SachiNandanMohanty, Shankar K., Arunkumar N., Gustavo Ramirez., 2021. Optimal Deep Learning Model for Classification of Lung Cancer on CT Images.Future Generation Computer System 2021.

2. JaneeAlam, S., &Hossan, A. Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifier. 2020 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)...

3.A.Worawate. Thirach, A., Marukatat, S., &Wilaiprasitporn, T. 2018.Automatic Lung Cancer Prediction for Chest X-ray Images Using the Deep Learning Approach. Biomedical International Conference (BMEiCON-2020

4. Haarburger, C., Weitz, P., Rippel, O., &Merhof, D. 2020. Image-Based Survival Prediction for Lung Cancer Patients Using CNN.2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2020). doi:10.1109/isbi.2019.875949

5. Casuey,l, J., Guan, Y., Dong, W., Walker, K., Qualls, A, J., Prior, F., & Huang X. 2020 Lung Cancer Screening With Low-Dose CT Scans Using A Deep Learning Approach. ArXiv:1906.00240 [eess.

6. Hu, Z., Tang, J., Wang, Z., Zhang, K., Zhang, L., & Sun, Q. (2021). Deep learning for image-based cancer detection and diagnosis − A survey.Pattern Recognition, 83, 134– 149. doi:10.1016/j.patcog.2018.05.01

7.Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Sánchez, C. I. (2021). A survey on deep learning in medical image analysis.Medical Image Analysis, 42, 60– 88. doi:10.1016/j.media.2017.07.00

8.Alkadi, R., Taher, F., El-baz, A., &Werghi, N. (2021). A Deep Learning-Based Approach for the Detection and Localization of Prostate Cancer in T2 Magnetic Resonance Images.Journal of Digital Imaging.doi:10.1007/s10278-018-0160-1

9.M. Havaei., A. Davy., D. Warde-Farley., A. Biard., A. Courville., Y. Bengio., C. Pal, P.-M. Jodoin, H. Larochelle. 2021. Brain Segmentation With Deep Neural Networks. Medical image analysis, 35 (2017) 18-31.

10. Z. Tian., L. Liu & B. Fei. 2021. Deep Convolutional Neural Network for Prostate MR Segmentation. SPIE Medical Imaging, International Society for Optics and Photonics2021, pp. 101351L-101351L-101356.