



Implementation of Artificial Intelligence Techniques for Cancer Detection

Darshan Patel¹ · Yash Shah¹ · Nisarg Thakkar¹ · Kush Shah¹ · Manan Shah²

Received: 16 September 2019 / Accepted: 10 October 2019 / Published online: 29 November 2019
© Springer Nature Singapore Pte Ltd. 2019

Abstract

Diseases like cancer have been termed as chronic fatal disease because of its deadly nature. The reason why cancer is termed as fatal is cancer progresses faster, and in most of the cases, these cells are detected at an advance stage. It is found that early detection of cancer is the key to lower death rate. In this study, overviews of applying AI technology for diagnosis of three types of cancer, breast, lung and liver, have been demonstrated. Various studies are reviewed for the different types of systems which are used for early detection of cancer. Automated or computer-aided systems with AI are considered as they provide a perfect fit to process a large dataset with accuracy and efficiency in detecting cancer. Diagnosis and treatment can be carried out with the help of these systems. Breast, lung and liver cancer studies have shown that some of these systems provide accurate precision in diagnosis and thus can solve the problem if these systems are implemented. However, these systems have to face a lot of hurdles to be implemented on a large scale. Image preprocessing, data management and other technology also need enhancement to be compatible with AI and machine learning algorithms to be implemented. Considering the experimental results, this study shows there is no doubt that the AI-implemented neural networks would be the future in cancer diagnosis and treatment.

Keywords Artificial intelligence · Health care · Cancer · Detection

Introduction

Artificial intelligence (AI) is the implementation of machines programmed in a way that they think and act like human. Artificial intelligence provides a step closer to automation in every sector. It provides efficient result with accuracy and without human error, making it one of the most important advancements in the coming technology. Artificial intelligence is based upon working of network of neurons in a human brain, it is an artificial replica of a neuron which is made to be learned to perform task of a biological neuron through various algorithms and methodologies. The biological neural network interact

through electrical impulses which act as a signal between the neuron to perform a specific task. The artificial neural network (ANN) works in a similar way, it uses several artificial neurons ranging from 10 to 10,000 which is much lesser than biological neurons in a brain which is in billions. ANN are computational models which are capable of pattern recognition and machine learning [20]. Artificial intelligence makes our day-to-day life efficient while minimizing human efforts. AI is implemented in various fields to ease our daily routines. AI technology handles several activities in banks and financial systems. It includes smarter chat bots for better and quick customer services. It also manages financial operations, smart investing in stocks, managing properties and much more. Manufacturing companies and industries also use AI, mostly in the production unit. They are used as robots to move an object from one place to another, provide different shapes to an object. Companies use AI for management of employee records, retrieval of specific information. It provides time management and provides accurate result. AI plays a vital role in Air transport systems, most activities performed in

✉ Manan Shah
manan.shah@spt.pdpu.ac.in

¹ Department of Computer Engineering, Indus University, Ahmedabad, Gujarat, India

² Department of Chemical Engineering, School of Technology, Pandit Deendayal Petroleum University, Gandhinagar, Gujarat, India

air transport include AI technologies. Different softwares based on AI provide a better flight experience for the customers. AI algorithms have provided bots against whose player could play in gaming console. AI technology has helped in advancement of playing against a superior competitor for better gaming experience.

AI plays a vital role in the field of medicine. Machines are capable of analyzing vast amounts of data and identifying patterns that humans cannot. Building of AI programs that carry out diagnosis and make recommendations is key concern of medical artificial intelligence. Some medical applications are based on entirely numerical and probabilistic procedures, unlike this, programs developed using AI are built on symbolic models of disease units and their association with patient factors and clinical demonstrations. Very early, the prospective of such a technology in healthcare and medicine fascinated scientists and doctors [30]. Increasingly, supercomputers and homegrown systems are being used by the doctors and hospitals to detect patients who might encounter heart disease, kidney failure or postoperative infections. Analysis of association between prevention or treatment methods and outcomes of patient is the key objective of artificial intelligence in medicine (AIM). It was expected that smart computers, with the ability to store and process broad quantity of knowledge, would provide aid or excel doctors with different tasks such as diagnosis and become “doctors in a box.” In 2014, the acquirement of AI start-ups in healthcare was about \$600 million, while in 2021, it is expected to be \$6.6 billion. Dr. Eric Horvitz, who works as director for Microsoft Research Labs and specializes in implementing AI in healthcare sector, said, “Electronic health records are like large quarries where there’s lots of gold, and we’re just beginning to mine them.”

One of the most catastrophic and lethal diseases to human beings universally among vast variety of diseases is cancer. 0.3 million deaths per year are caused by cancer, which is the second-largest disease responsible for maximum mortality [3, 35]. A total of 413,519 men died due to cancer and 371,302 women died of the same reason, which totaled to 784,821 deaths due to cancer [27]. In the USA, cancer ranks second in common causes of death and is responsible for nearly 25% of deaths (American Cancer Society), while in Britain, cancer will be diagnosed in more than one-third of people during their life span and one out of four will die from it [47]. Many different computer techniques are used for detection and prevention of this life-threatening disease. The most effective and efficient way is to use different AI techniques to identify the occurrence of cancer at early stages. The concept of machine learning can be taken into account for developing models for diagnosis of cancer. Ability of a computer to learn without being programmed is known as machine

learning [17]. Various models can be used to learn and predict the occurrence of cancer from CT scans, and image processing techniques would be helpful in this case. Models can also be trained on the basis of various factors such as genetic heritage, tobacco and alcohol consumption, obesity, exposure to radiation or even having poor and inactive lifestyle [2].

AI Techniques

There are several AI techniques available, but not all can be used for cancer detection. Each and every technique has its own feature which determines whether it can be used for a specific reason or not. There are several leading artificial techniques such as neural networks, genetic algorithms, fuzzy clustering, support vector machines (SVMs), particle swarm optimization (PSO), decision tree classifier, Bayesian networks, linear regression (LR) and computer-aided detection (CAD). These methods determine the classification or regression approach and may also determine a complete technique for cancer detection. Various techniques used for the detection of lung cancer, liver cancer and breast cancer are reviewed in this paper. It is observed that k-NN, SVM, fuzzy clustering and deep learning are few techniques which provide the best result for the identification of cancer.

Implementation of Various Techniques on Different Types of Cancer

Lung Cancer

Main reason for mortality universally among all varieties of cancers is lung cancer. Primary reason for this is the difficulty in detection in the initial stages, and it is very challenging to triumph against this malignant disease at later stages [37]. The main reason for the trouble in detection is the appearance of symptoms after long period of time [43]. Lung cancer was first recognized as a distinct disease in 1761 and its various aspects were described later in 1810. It was a rare disease before the invention of cigarette smoking [33]. In the USA, approximately 154,050 deaths were expected [27] due to lung cancer which accounted for 25% of cancer deaths (American Lung Association, lung.org). The fatal disease kills about 80% of patients within a span of 5 year after diagnosis [23]. Long-term exposure to tobacco smoke, which includes both active smoking and passive smoking, is responsible for about 85% of cases in lung cancer, while 10–15% cases occur due to combination of other factors such as exposure to radon gas, genetic factors, exposure to air pollutants or

asbestos [21]. There are two types of lung cancer: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC); NSCLC accounts for about 80% of the total cases [11]. There is substantial evidence which suggests that there will be a decrease in mortality rate by detecting lung cancer early [1, 36]. Thus, new techniques need to be implemented for discovery of lung cancer in initial stages as current available techniques fail to do so. Many procedures such as computed tomography (CT) scan, chest radiography (X-ray), sputum cytology and magnetic resonance imaging (MRI) scan are used to identify lung cancer. But all these can detect lung cancer at advanced stage. Thus, there is need for artificial intelligence in early detection of this cancer. Image processing can improve current scenario.

Pathan and Saptalkar [36] proposed a technique: 35 samples of images are taken which are used for the classification and they are compared with the actual image. The first step of the procedure is seed fill operation in which the given node and its area are detected in a multidimensional array. The next step is region of interest where data samples are collected for a given purpose. Then, image segmentation is done to differentiate parts of image to extract texture features. And the next step is color image processing where colors are used to improve the representation in terms of saturation, hue and intensity of image.

K-means algorithm is used for clustering of images where data point and clusters are defined. If a data point is found close to its own cluster, it is left where it is, and if not, it is moved to the closest cluster. Hence, the classification of these images is done by comparing them with samples and lung cancer is detected.

Zhou et al. [49] proposed a method for the detection of cancerous cell in lung based on ANN (artificial neural network) ensemble. Advancement in computer-aided diagnosis of lung cancer has been carried out in recent years, making it possible for the detection of lung cancer at initial stages. The aim of the paper is to implement ensemble neural network (NED) for the detection of cells which are cancerous. In the study, 522 cell images were considered out of which 75% belong to cancer cell. They divided the dataset of 522 images into five subsets where the proportion of different cells and their size remains same as the original dataset. They experimented five times using union of four subset as a training to set the module for the identification of lung cancer cell, and the remaining subset is used to test how well the module works. They found that single artificial neural network gives accuracy less than 60%. The two kinds of assemble of neural networks have resulted in lower error rates, and ensemble of two networks rather than single network has lowered all three error measures; however, the rate of false identification is still 7.9% which is not satisfactory for real-time implementation. Neural ensemble-based detection (NED) which

can be implemented in early stages of lung cancer diagnosis system (LCDS), in which NED is inputted, is used in cases where X-rays could not clearly diagnose cancer cells. Neural ensemble-based detection (NED) implements a specific two-level architecture: The first level judges a cell to be cancer cell or not on the basis of “full voting”; if all network gives output to be a healthy cell, then it is considered as one. The cells judged by first level to be cancer cell are passed on to second-level ensemble for detecting its type. NED comes out to be the overall best approach to be applied to lung cancer diagnosis system (LCDS) as it resolves all error measures to a greater extent. Bayi Hospital currently follows NED-implemented LCDS as routine examination followed by chest X-rays for diagnosing lung cancer.

Naresh and Shettar [35] presented an approach to use neural network for the early detection of lung cancer. Lung cancer will be diagnosed at initial stage using DICOM (.dcm) format CT scan images by the presented approach. SVM, ANN and k-NN are the three classifiers used for the detection of lung cancer and determine the severity: stage I or stage II. Results of these classifiers are compared on the basis of accuracy, sensitivity, precision and specificity. Gaussian white noise is removed from CT scan images which are used as input, and then, segmentation is performed. Enhancement is performed to get clear image so that nodule (tumor) can be detected.

CT scan images of National Lung Screening Trial (NLST) of stage I and stage II are used as dataset for this experiment. In total, 184 samples are taken: 111 samples for stage I and 73 samples for stage II of lung cancer. Four-fifth of the data were used for training the classifiers and remaining one-fifth for testing. The accuracy of SVM, ANN and k-NN was achieved as 95.12%, 92.68% and 85.37%, respectively. Precision was found to be 92.31%, 87.50 and 84.62% for SVM, ANN and k-NN, respectively. Sensitivity (recall) was 100% for SVM and ANN, while it was 91.67% for k-NN. Specificity for SVM was 88.24%, 100% for ANN and 76.47% for k-NN. The SVM classifier with RBF kernel achieved the best result with 95.12% accuracy and 92.31% precision. Thus, SVM classifier has been proposed for prediction of lung cancer at early stages.

Nancy and Kaur [34] presented a research to identify lung cancer in its primary stage which bestowed the use of neural network and GA algorithm. Lung cancer can be detected using manual mechanisms such as computed tomography (CT) scan, chest radiography (X-ray), sputum cytology and magnetic resonance imaging (MRI) scan. All these are used to diagnose cancer at very advance stage when it is too late. Thus, early detection automated approaches need to be adopted. Here neural network classifying algorithm is used to detect lung cancer at early stage and genetic algorithm is used for optimization.

The method is implemented fully in MATLAB, and tests are carried out on DICOM images. Firstly, CT scanned images are collected. Secondly, image preprocessing is done which is a crucial step in detecting lung cancer. To perform classification of patterns in neural networks, neural net has to be trained beforehand using data. Thirdly, important features are extracted which will be useful in classification. Finally, neural network is used for classification. Neural network can handle complex and dynamic situations and is very powerful in recognizing patterns present in the dataset given. Back-propagation is used to train neural network. This model back-propagates the errors. After that, genetic algorithm which is based on natural selection is used to optimize the results. It optimizes the set of potential solutions.

As a result, it is seen that after performing neural network classification algorithm, lung has been detected as abnormal lungs at very early stage. Neural network itself produces the best results, and to enhance its performance, it is passed on to genetic algorithm to optimize the obtained results.

Hussain et al. [25] conducted a research on the detection of lung cancer by the use of fuzzy clustering and ANN in the year 2015. The main objective of this research was to spot lung cancer at an initial stage so that there are higher chances of survivability. Their method of detecting the tumor was divided into six steps: (1) data collection—in this step, 100 data are collected in which 50 are cancerous and 50 are non-cancerous; (2) noise removal—since most of the images have noises, it is necessary to remove the noises using appropriate filters; (3) image enhancement—the image quality is enhanced for proper feature extraction; (4) Gray scale-to-binary conversion—in this step, the image is divided into segments prioritized according to pixels holding their threshold value; (5) removal of erroneous part—it is nothing but removing the final errors before getting the outcome; and (6) image segmentation—the step is performed by obtaining a gray-level co-occurrence matrix of the image which is used for feature extraction. The results will be very accurate and will tell whether the person has tumor or not. The testing image will be tested and it will be compared to the matrix which was formed during training. So to conclude, this method which is based on artificial neural network (ANN) will be very useful in detecting tumor in early stages to get proper treatment.

Sivakumar and Chandrasekar [42] carried out a study for the detection of nodule present in lung by the use of support vector machines and fuzzy clustering. The CT scan images are taken from Lung Image Dataset Consortium (LCID). The images are in the DICOM format with the size of 512×512 . A median filter was applied to eradicate the noise, which enriched the images. Afterward, the process

of enhancement segmentation is performed using weighted fuzzy possibilistic C-means (WFPCM) algorithm. Several other segmentation techniques are also available such as fuzzy C-means (FCM) and fuzzy possibilistic C-means (FPCM). After segmentation of images, several features like mean, contrast, entropy and standard deviation are calculated. SVM technique is used for the classification of images of lung nodules, which uses kernel function.

RBF, linear and polynomial are the different types of kernels used in SVM classifier. Results of these were compared on the basis of accuracy, specificity and sensitivity. RBF achieved 80.36% of accuracy, 76.47% of specificity and 82.05% sensitivity. In polynomial kernel, accuracy was 66.07%, specificity was 52.94% and sensitivity was 71.79%. In linear kernel, 71.43%, 75% and 70.83% were the results for accuracy, specificity and sensitivity, respectively. This study showed that classification performance was better for RBF kernel as compared to linear and polynomial kernels.

Liver Cancer

The largest internal organ in the body is liver and it is also the largest gland. Sixth most common cancer and third most common reason for cancer death is liver cancer, in which hepatocellular carcinoma (HCC) is the most common [9]. Second leading cancer responsible for death in men worldwide is liver cancer [45]. According to American Cancer Society, in 2017, more than 40,000 people in the USA were estimated to be diagnosed with liver cancer [8]. About 1 million new cases are diagnosed with primary liver cancer every year [31]. There are four ways of diagnosing liver cancer using image techniques, ultrasonic scan, angiography, MRI and CT scans. For early diagnosis, CT scan images are preferred due to its high resolution and innocuous to human body [26]. Lesions in liver can be identified in CT scans by difference in intensity in pixels. Segmentation of CT scans which is done manually is monotonous and consumes a lot of time [38]. Thus, automatic segmentation is a better choice as it will reduce the amount of time in the diagnostic process and help the patients. The following content provides different methods and techniques which can be used for the diagnosis of liver cancer.

Sharma and Kaur [40] carried out a research work on optimizing the detection of liver tumor and segmentation using neural network. In this research work, segmentation approach used is region based, while optimization techniques used are particle swarm optimization (PSO) and seeker optimization algorithm (SOA). CT scan images are used for the classification of tumors by comparing the mentioned optimization techniques. The primary aim of research is to identify the liver tumor and compare the

outcome of PSO and SOA on the basis of accuracy and elapsed time. The CT scan images are processed to remove noise and then enhanced to get better image quality. Segmentation differentiates tumor from liver. Region-based methods divide images that have similarities according to predefined criteria. It provides better results to contrast-enhanced image and is not affected by noise. The next step is optimization; here, PSO and SOA techniques are used. Optimization is the process of maximizing or minimizing a function.

In this research work, testing was done by using 15 different cases of patients having liver tumor. The outcomes of two optimization techniques are compared in terms of accuracy and elapsed time. The ability of classification test to identify a condition measures accuracy whereas elapsed time is the amount of time used for a particular process to be completed. Results of PSO gave 93.33% accuracy in classification and detection, while results of SOA provided 60% accuracy in classification and detection. On the basis of elapsed time, PSO took 42.429959 s, while SOA took 48.744537 s. Thus, it was concluded that results of PSO were better in terms of accuracy and elapsed time for processes as compared to results of SOA. For detection and classification accuracy, 15 cases were used for testing the neurons and 30 cases were used for training the neurons.

Das et al. [12, 13] carried out a research to detect cancer in liver by using decision tree classifier and modified fuzzy clustering in CT images. Manually detecting and characterizing CT scan images are difficult; thus, clustering method involving adaptive thresholding and spatial fuzzy is presented using multilayer perceptron (MLP) and C4.5 decision tree classifiers. As a result, it is seen that spatial fuzzy C-means (SFCM)-based segmentation with C4.5 decision tree classifier is an efficient way for automatic recognition of the liver cancer.

The data used in performing this experiment include 63 cases of hepatocellular carcinoma (HCC) and 60 cases of metastatic carcinoma (MET). Adaptive thresholding is used to segment the liver section from kidney and spleen, and SFCM is used to obtain cancerous part from the liver image. After this, LBP Fourier feature descriptor was used to obtain features from that image. And then, MLP and C4.5 decision tree classifier are used to differentiate the two types of cancer: (1) hepatocellular carcinoma (HCC) and (2) metastatic carcinoma (MET). Different measures taken for classifying are accuracy, precision, recall, specificity and sensitivity. C4.5 decision tree classifier is found to be more accurate and generate better results in all areas.

The proposed experiment shows that SFCM with C4.5 decision tree classifier is a very accurate way to detect cancer and classify the data without any manual process.

Through this process, cancer can be diagnosed at very early stage.

Gopal and Vanitha [19] conducted a research on liver tumor detection using neural network classifier in the year 2018. The whole research is performed in three steps which are preprocessing, feature extraction and classification. The first step, preprocessing, is mainly for removing the noises so that the liver image can be enhanced. Then in the next step texture features are extracted from the image and those features are classified using neural network classifier. Neural network classifier is said to be a replica of how the human brain works. Using the techniques properly by completing each and every step properly, the results were pretty good. The proposed method was proved to be the most accurate method among all the others that were performed earlier. The accuracy was almost near to 1 which is 0.982. Earlier the highest any method could reach was 0.957 by non-density-based method; particle swarm optimization had the second highest accuracy (0.93) until the proposed method came to highlight. The seeker optimization algorithm did not give promising results (0.60), making the proposed method the best and the most accurate method. Liver tumor segmentation is important for treating the tumor properly. The size of liver is relatively large, so segmenting it would be a good option, so that we get almost perfect results. The proposed method promises the best accuracy of tumor segmentation which is 98.2%.

Das et al. [13] proposed a study on liver cancer detection based on deep learning using Gaussian model techniques and watershed transform. It is time-consuming, and results are inaccurate when cancer tissues are manually detected. For appropriate therapy, a computer-based diagnosis is proposed which helps to decide and predict cancer. Their work was focused on the detection of the liver cancer precisely using their techniques. In their work, they have introduced a new system which involves various deep learning methods for the detection of the cancer lesion in using computer-based images of the liver. Two hundred and twenty-five images were used in their work for development of the model; these 225 images were from 75 patients (46 males and 29 females). Watershed segmentation process separates the liver using markers, and mixture model algorithm then segments the lessening. Texture features were taken out from the segmented region after tumor segmentation process is completed. Neural network is inputted with these segmented features for the classification of three types of liver cancer, i.e., hemangioma, metastatic carcinoma and hepatocellular carcinoma. They achieved efficient accuracy of 99%. They classified at one hundred epochs with neural networks with false reading of 0.06 during their process. This system is ready to aid radiologists in the detection of liver cancer using CT images.

Devi and Dabas [16] compared two liver tumor detection techniques. The back-propagation techniques using neural networks and support vector machines are compared in their study for the classification of liver cancer. The process of image segmentation is used, and a liver image is segmented and partitioned to detect tumor. The training of neural network is done with deciding the inputs to be feed into the network; then, the partitioning takes place. Normalization of data takes place, so the network has the common range; these data are then converted and given a specific range or values. These values represent the type of tumor a case has. Lastly the dataset is divided into: one that has to be tested and the other that trains the neural network. Both the models are compared if their rate of detection is true and false based on the study of five hundred and eighty-three cases out of which 418 cases are positive and the remaining negative. The back-propagation neural network shows accuracy of 73%, whereas the support vector machine shows 63% accuracy. This experimental results shows that the back-propagation neural networks is more efficient in detecting liver tumor than the rest of the techniques, and thus, it can be used in clinical approaches. However, the training for the network is a tedious task which can be simplified in the future.

Automation in identifying liver cancer is crucial task. The first step is image preprocessing [46]. Ultrasound liver cancer tumor images are used for this study. Images with high frequencies are obtained using filters for better research. In this proposed research, first of all seed point is rectified from liver cancer ultrasound images and is performed by co-occurrence matrix and run length method. Co-occurrence features considered here are energy and entropy.

After that, segmentation is done. Normally segmentation is difficult as it contains noise, non-uniform intensity, and by using region growing algorithm, noise is eliminated and spatial information is restored. After that, for further segmentation gray space map and Otsu method are used. Using Otsu method, textural features are extracted. For further classification of ultrasound images, SVM classifier is used which differentiates the liver cancer image and diagnoses the cancer images as normal, benign or malignant. Features taken for comparison are mean, variance, entropy, skewness and standard deviation. The results are 96.72% accurate. SVM outstands in comparison with other classifiers even with small number training data samples.

Breast Cancer

Among women which are aged between 35 and 64 years, the primary source of death is breast cancer and leading cause in cancer-related death among female population [18, 39]. Worldwide, after lung cancer, breast cancer is

responsible for most deaths related to cancer [15]. According to Principality of Asturias, among every 500 women, one will develop breast cancer at some time in her life [32]. 12.5% is the chance that a woman at some time in her life will develop breast cancer. Every year, about 182,000 new cases are diagnosed with breast cancer and causes death of about 46,000 women in the USA [10, 22]. Every one woman out of two women who are newly diagnosed with breast cancer dies in India. Mortality rate of breast cancer can be reduced if proper screening and diagnosis methods are followed at early stages before the physical symptoms appear in the body [44]. 15–35% of breast cancers are not detected during screening; this is because of error or is imperceptible to radiologists [24]. This is where the need for using computer-aided technology for the detection of breast cancer is helpful. It reduces the error and increases cancer detection rate.

Kumari and Singh [27] proposed a system for the prediction of breast cancer. The aim of the system is to scrutinize the smallest set of features and foretell the development of breast cancer at initial stage. The dataset used to conduct the proposed experiment is Wisconsin Breast Cancer Dataset (WBCD) which was obtained from UCI repository. 65% of the dataset are cancer samples, and 35% of the dataset are non-cancer samples. Linear regression (LR), SVM and k-NN classifiers are used to obtain enhanced accuracy. k-NN classifier produced highest accuracy in comparison with LR and SVM. The accuracy obtained in this study is highest as compared to other systems.

The dataset comprised 699 samples and 11 attributes; these 11 attributes could provide accurate information about development of breast cancer. In this study, filter method is used for the selection of relevant features from the available ones. The proposed study achieved a maximum classification accuracy of 99.28% as it surpassed a method proposed by Marcano-Cedeño et al. [29] which has a classification accuracy of 99.26%. With the help of proposed system, cost of treatment is reduced and quality of life is increased as breast cancer can be predicted at early stage of development. This work can further assist in creating fruitful prediction system that will reduce overall cost and time of patients and doctors, and will also reduce mortality rate.

Asri et al. [4] proposed an approach which performed prediction and diagnosis of breast cancer using algorithms based on machine learning (ML). The main objective of this research is to find the best possible data mining algorithm to classify data of breast cancer in terms of maximum accuracy and lowest error rate. The discrete algorithms of machine learning such as decision tree, k-nearest neighbors (k-NN), support vector machine (SVM) and naïve Bayes (NB) are tested in this experimental research. These

experiments use the Wisconsin Breast Cancer (original) datasets from the UCI Machine Learning Repository and are conducted in WEKA data mining tool.

For the evaluation of performance of these algorithms in terms of effectiveness, the evaluation criteria taken are time required to build a model, appropriately classified instances, wrongly classified instances and accuracy. According to results, the most accurate algorithm is SVM, with accuracy of 97.13% with highest of 678 instances being classified correctly and lowest of 21 instances being classified incorrectly. For better measurement of the performance, simulation error should also be taken under consideration. Hence, evaluation criteria now taken are root mean squared error, root relative squared error, mean absolute error, relative absolute error and kappa statistic. Results show that error rate of SVM algorithm is lowest compared to other algorithms with mean absolute error being 0.02, root mean squared error of 0.16, relative absolute error of 6.33% and root relative squared error being 35.58%.

For calculating performance in terms of efficiency, one needs to calculate accuracy in terms of true positive (TP) rate, false positive (FP) rate, precision and recall. As a result of comparison between all classifiers, it is seen that TP rate of SVM (97%) is highest and FP rate of SVM is lowest. Also to get better idea about efficiency, ROC curve is presented. It gives graphical performance of each classifier. ROC curve clearly shows that SVM is the best classifier. It shows that SVM is 99% sensitive and 99% specific.

Thus, we can summarize that SVM performs best in terms of effectiveness and efficiency when measures such as accuracy, error rate, precision and recall are taken into account while performing classification of breast cancer datasets. Here, the most important task was to identify the best algorithm to predict breast cancer datasets. After performance evaluation of four algorithms C4.5, NB, k-NN and SVM on breast cancer datasets, it is observed that SVM produces the most optimized results.

Watanabe et al. [48] carried out a retrospective study in which computer-aided detection (CAD) algorithm can be used for better screening and detection of breast cancer. The algorithm used was cmAssist, which is an AI-based CAD algorithm that includes deep learning. The aim of the study was to find techniques to detect breast cancer as early as possible. Mammograms of 122 patients were used for this study in which earliest was of February 7, 2008, and latest of January 8, 2016. A panel of seven radiologists was included, and cases were shown to each of them without any information. Everyone was asked to review the mammograms without markings of cmAssist and make a decision; later, they were given markings provided by the algorithm and an opportunity to alter their decision. There

was an increase in cancer detection rate (CDR) of average 11% when cmAssist was used. Overall CDR without the assistance was 25% to 71% with average of 51% and CDR with assistance of software was 41% to 75% with mean of 62%. There was a significant advantage to a radiologist who showed an increase of 64% in CDR when this technique was provided. This study identified 90 cases in which breast cancer could have been detected in the retrospective mammograms but instead declared as negative, false negative. Malignant lesions missed as early as 5.8 years (70 months) before the diagnosis were highlighted by this AI-CAD.

Bellaachia and Guven [6] conducted a research on using data mining techniques for the prediction of breast cancer survivability. They used the help of SEER database in finding all the information required. The data mining methods used were the C4.5, back-propagated neural network and the naïve Bayes. The results show that the first two techniques have superior performance. In the methodology after the procedure of preprocessing, a common evaluation would be conducted to determine the effect of the features on the prediction. The work is done using the weka toolkit. They used recall, precision and the performance metrics of accuracy to compare the three procedures. They also applied tenfold cross-validation. The data were divided into K groups. In this way, the rate of error can be expected in a neutral way. The results show that the C4.5 and the neural network method have given better results in comparison with the naïve Bayes method. According to the results classification rate of their approach (87%) is way higher than classification rate of approach used in Delen et al. [14] i.e. (81%). So, this study has discussed and solved the concerns, algorithms and methods for the problem of cancer survivability prediction in SEER database. The results of the experiments are very promising used in data mining. Records of future data are not included in the analysis. Finally using the survival time prediction, they would like to use it for respiratory cancer which is very dangerous and whose survivability is very low.

Badawy et al. [5] conducted a study on breast cancer detection using mammogram segmentation. Mammography is used to detect breast cancer in early stages. In this study, they approached with enhanced double thresholding method. They improvised by adding borders of the final segmented image to the original image which will help the physicians to detect cancer and make better diagnosis. They took four samples of mammogram images in their study. The method is performed in three steps. The first step is called double thresholding segmentation which is an important part of image segmentation for isolating the objects from background. The second step is masking and morphological operations; after the first step is completed,

the specified mask is used to remove unwanted borders; after that, morphological operation is performed. The third step is the applied segmentation approach which is further divided into three steps which will give the final result. The results not only show whether the person is having breast cancer or not, but it also shows the parts where the cancer is growing. This method has two great advantages: Firstly they reduce processing time and reduce processing storage area. So to conclude the study proved that enhanced double thresholding has given good results in breast cancer detection. Moreover, it can be also used for all biomedical images.

Singh and Gupta [41] conducted a study on breast cancer detection and segmentation in a mammogram. In this paper, they have introduced a simple and easy approach for detecting cancerous tissues. After the detection phase, segmentation is performed for the tumor region. It is helpful in detecting the cancer at an early stage which helps in reducing death possibilities. Their method is mainly divided into two steps: detection and segmentation. In the first step, they performed thresholding operation and averaging filter on the original image which shows the regions where cancer exists. In the next step which is known as segmentation, a tumor patch is discovered by the help of morphological closing operation and the region boundary is found using image gradient technique. The outcome of their experiment proved to be an accurate for detecting cancer as it is different from other methods. This method only detects the tumor region and locates it on original image unlike the other methods which also detects some non-tumor region. So to conclude the method performed is very simple and fast as they used basic imaging processing techniques. However, it has one drawback which is manual selection of threshold parameter and the size of averaging filter.

Challenges and Future Scope

The implementation of AI in cancer diagnosis has lately been recognized all over the globe for its accuracy and precision; however, there are several limitations with these existing systems. Clinical procedure generates a large amount of data; the management and storage of these data are an arduous task. Picture archiving and communication system and the digital imaging and communication in medicine have ensured that the data are organized for easy access and retrieval. The data available are not labeled, thus making it difficult to use in neural network. It is time-consuming and requires a lot of effort to convert it into labeled data. Unsupervised and self-supervised methods do not require data which are labeled; thus, developing such system could provide a better

option in future. Furthermore, professionals, private institutes and government bodies should be encouraged to share their data to increase the dataset. The efficiency of neural networks will increase as the amount of data used to train the model increases. The forthcoming new systems with better AI implementation can provide a solution to these problems.

The developers and engineers, who will design and develop the systems, should have the knowledge of the domains, computer technology and health care. They should be thorough with the latest techniques and problems occurring in these fields to make the best effective system. A team of experts may be hired to assist the developers to find variations in result and flaws in the system to increase the overall efficiency. Even the staff who will operate and access these systems should be well trained for proper functioning of the system. Training should be provided by the developer to the healthcare staff for better understanding and operation of the system. People will also not accept the diagnosis performed by a machine over human doctors, which will thwart progress of implementing these systems. People should be made aware about the benefits of the systems, which will result in betterment of mankind.

With AI being implemented for complex cancers as breast, liver and lung, cancer detection in early stages could result in improved healthcare facilities. Supercomputers like IBM Watson have partnered with Memorial Sloan Kettering Cancer Centre to develop computing systems for advanced research in cancer. Thirteen different cancer types are being analyzed by IBM Watson which includes breast cancer, in which breast cancer trials were boosted by 80%. Improvement in experimental neural networks would no doubt enhance the accuracy of the results produced in future [28, 7].

The current techniques mentioned in this paper have outperformed most classification and regression techniques till date. Although these techniques have been proved efficient theoretically, practically these techniques have not been implemented on a wide scale. Once these are implemented, their practical drawbacks can determine the further scope of research. Depending on the limitations, area of improvement can be identified and efforts to improve it can be made. Further research can be carried out to determine new techniques which have better performance than the existing ones. The new techniques should be more efficient, less time-consuming and able to process large amount of data with ease. According to a study by researcher at Well Cornell Medicine and New York-Presbyterian, there are also techniques which can determine the types of cancer cells with approximately 100 percent accuracy. CNN developed by the researchers can determine whether cells are malignant, and if they are, it can also identify the type of cancer. This can be further researched

to make it more efficient and able to handle large amount of data to reduce the work load of doctors on determining the type of cancer.

Conclusion

Research shows that numerous deaths take place due to breast, liver and lung cancers. As shown in this study, the reason behind them is detection of cancer at very advanced stage. Thus, early detection is necessary which can be done with the help of automation in diagnosing cancer. The proposed study achieved this with the help of data mining classifying algorithms or machine learning algorithms. With the help of artificial intelligence, early detection of cancer is possible which can save several lives. Various methods are proposed in this paper for detecting breast cancer at early stage which includes k-NN, C4.5, SVM, CAD and thresholding algorithm. Some study shows that k-NN is the best classifier for classifying breast cancer and some show that SVM is best classifier. Also CAD with deep learning produces efficient results. Several approaches were shown in this study which uses artificial intelligence algorithms for early detection of lung cancer. NED, SVM, neural network with genetic algorithm and fuzzy clustering algorithms are reviewed in this paper. Some study show that SVM is the best classifier with RBF kernel with 95.12% accuracy. Also neural network and fuzzy clustering produce efficient results. Genetic algorithm optimizes these results. PSO optimizing technique, back-propagation neural network, neural network classifier, SVM and SFCM are several techniques reviewed in this research for the detection of liver cancer. Accuracy results were as follows: PSO produces 93.33% accurate results, neural network classifier produces 98.2% accurate results, deep learning with Gaussian model and watershed transform produces 99% accurate results, and SVM produces 96.72% accurate results.

Acknowledgements The authors are grateful to Indus University and School of Technology, Pandit Deendayal Petroleum University, for the permission to publish this research.

Authors Contribution All the authors make substantial contribution in this manuscript. DP, YS, NT, KS and MS participated in drafting the manuscript. DP, YS, NT and KS wrote the main manuscript. All the authors discussed the results and implication on the manuscript at all stages.

Funding Not applicable.

Availability of data and material All relevant data and material are presented in the main paper.

Compliance with ethical standards

Competing interests The authors declare that they have no competing interests.

Consent for publication Not applicable.

Ethics approval and consent to participate Not applicable.

References

1. Ada RK (2013) Early detection and prediction of lung cancer survival using neural network classifier. *Int J Appl Innov Eng Manag* 2(6):375–383
2. Al-ani SM, Abbod M (2015) Prediction of cancer behavior based on artificial intelligence. In: *Proceedings of the international conference on circuits, systems, signal processing, communications and computers*, pp 15–19
3. Ali I, Wani WA, Saleem K (2011) Cancer Scenario in India with future perspectives. *Cancer Therapy* 8:56–70
4. Asria H, Mousannif H, Al Moatassime H, Noeld T (2016) Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Proc Comput Sci* 83:1064–1069
5. Badawy SM, Hefnawy AA, Zidan HE, GadAllah MT (2017) Breast cancer detection with mammogram segmentation: a qualitative study. *Int J Adv Comput Sci Appl* 8(10):117–120
6. Bellaachia A, Guven E (2006) Predicting breast cancer survivability using data mining techniques, Ninth workshop on mining scientific and engineering datasets in conjunction with the sixth SIAM international conference on data mining, 1–4
7. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, Allison T, Arnaout O, Abbosh C, Dunn IF, Mak RH, Tamimi RM, Tempny CM, Swanton C, Hoffmann U, Schwartz LH, Gillies RJ, Huang RY, Aerts HWL (2019) Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin* 2019(69):127–157
8. Chaudhary K, Poirion OB, Lu L, Garmire LX (2017) deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clin Cancer Res* 24(6):1248–1259
9. Chen Y, Yi Q, Mao Y (2008) Cluster of liver cancer and immigration: a geographic analysis of incidence data for Ontario 1998–2002. *Int J Health Geograph* 7:28. <https://doi.org/10.1186/1476-072X-7-28>
10. Cheng HD, Cai X, Chen X, Hu L, Lou X (2003) Computer-aided detection and classification of microcalcifications in mammograms: a survey. *Pattern Recognit* 36:2967–2991
11. Coccia M (2019) Artificial intelligence technology in oncology: a new technological paradigm. *ArXiv*, abs/1905.06871
12. Das A, Das P, Panda SS, Sabut S (2019) Detection of liver cancer using modified fuzzy clustering and decision tree classifier in CT images. *Pattern Recognit Image Anal* 29(2):201–211
13. Das A, Rajendra Acharya U, Panda SS, Sabut S (2019) Deep learning based liver cancer detection using watershed transform and Gaussian mixture model techniques. *Cognit Syst Res* 54:165–175
14. Delen D, Walker G, Kadam A (2005) Predicting breast cancer survivability: a comparison of three data mining methods. *Artif Intell Med* 34(2):113–127
15. Deserno TM (2010) Fundamentals of biomedical image processing. In: Deserno T (eds) *Biomedical image processing. biological and medical physics, biomedical engineering*. Springer, Berlin, pp 1–51

16. Devi P, Dabas P (2015) Liver tumor detection using artificial neural networks for medical images. *Int J Innov Res Sci Technol* 2(3):34–38
17. Dwivedi SA, Borse RP, Yametkar AM (2014) Lung cancer detection and classification by using machine learning and multinomial Bayesian. *IOSR J Electron Commun Eng (IOSR-JECE)* 9(1):69–75
18. Fernandez E, González JR, Borràs JM, Moreno V, Sánchez V, Perisa M (2001) Recent decline in cancer mortality in regression analysis. *Eur J Cancer* 37:2222–2228
19. Gopal RR, Vanitha E (2018) Liver tumor detection using neural network classifier. *Int J Pure Appl Math* 119(12):553–563
20. Gupta B, Tiwari S (2014) Lung cancer detection using curvelet transform and neural network. *Int J Comput Appl* 86(1):15–17
21. Gupta R, Ahmed KKM, Gupta BM, Bansal M (2016) Lung cancer in India: a scientometric study of publications during 2005–14. *Int J Med Public Health* 6(4):200–208
22. Guzmán-Cabrera R, Guzmán-Sepúlveda JR, Torres-Cisneros M, May-Arriola DA, Ruiz-Pinales J, Ibarra-Manzano OG, Aviña-Cervantes G, Parada AG (2013) Digital image processing technique for breast cancer detection. *Int J Thermophys* 34:1519–1531
23. Hashemi A, Pilevar AH, Rafah R (2013) Mass detection in lung CT images using region growing segmentation and decision making based on fuzzy inference system and artificial neural network. *Int J Image Gr Signal Process* 6:16–24
24. Houssami N, Lee CI, Buist DSM, Tao D (2017) Artificial intelligence for breast cancer screening: opportunity or hype? *Breast* 36:31–33
25. Hussain MA, Ansari TM, Gawas PS, Chowdhury NN (2015) Lung cancer detection using artificial neural network and fuzzy clustering. *Int J Adv Res Comput Commun Eng* 4(3):360–363
26. Jiang H, Zheng R, Yi D, Zhao D (2013) A novel multiinstance learning approach for liver cancer recognition on abdominal CT images based on CPSO-SVM and IO. In: *Computational and mathematical methods in medicine*, pp 1–10
27. Kumari M, Singh V (2018) Breast cancer prediction system. In: *International conference on computational intelligence and data science*. *Procedia Computer Science*, vol. 132, pp 371–376
28. Limkin EJ, Sun R, Dercle L, Zacharakis EI, Robert C, Reuzé S, Schernberg A, Paragios N, Deutsch E, Ferte C (2017) Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol* 28:1191–1206
29. Marcano-Cedeño A, Quintanilla-Domínguez J, Andina D (2011) WBCD breast cancer database classification applying artificial metaplasticity neural network. *Expert Syst Appl* 38(8):9573–9579
30. Mary A, Majeed AKM (2015) Artificial intelligence and medical science: a survey. *Int J Adv Res Eng Manag* 1(1):66–72
31. Massoptier L, Casciaro S (2008) A new fully automatic and robust algorithm for fast segmentation of liver tissue and tumors from CT scans. *Eur Radiol* 8:1658–1665
32. Menéndez LA, de Cos Juez FJ, Lasheras FS, Riesgo JAA (2010) Artificial neural networks applied to cancer detection in a breast screening programme. *Math Comput Model* 52:983–991
33. Mustafa M, JamalulAzizi AR, Izzam EL, Nazirah A, Sharifa AM, Abbas SA (2016) Lung cancer: risk factors, management, and prognosis. *J Dent Med Sci* 15(10):94–101
34. Nancy PK (2015) Identifying lung cancer in its early stage using neural network and GA Algorithm. *Int J Adv Res Comput Eng Technol* 4(2):341–344
35. Naresh P, Shettar R (2014) Early detection of lung cancer using neural network techniques. *Int J Res Appl* 4(8):78–83
36. Pathan A, Saptalkar BK (2012) Detection and classification of lung cancer using artificial neural network. *Int J Adv Comput Eng Commun Technol* 1(1):62–67
37. Praveen SS, Kavitha C (2013) Detection of lung cancer nodules using automatic region growing method. In: *Fourth international conference on computing, communications and networking technologies (ICCCNT)*, Tiruchengode, pp 1–6
38. Rajagopal R, Subbiah P (2014) Computer aided detection of liver tumor using SVM classifier. *Int J Adv Res Electr Electron Instrum Eng* 3(6):10170–10177
39. Rodriguez C, Plasencia A, Schroeder DG (1995) Predictive factors of enrollment and adherence in a breast cancer screening program in Barcelona (Spain). *Soc Sci Med* 40(8):1155–1160
40. Sharma A, Kaur P (2013) Optimized liver tumor detection and segmentation using neural network. *Int J Recent Technol Eng* 2(5):7–10
41. Singh AK, Gupta B (2015) A novel approach for breast cancer detection and segmentation in a mammogram. In: *Eleventh international multi-conference on information processing-2015*. *Procedia Computer Science*, vol 54, 676–682
42. Sivakumar S, Chandrasekar C (2013) Lung nodule detection using fuzzy clustering and support vector machines. *Int J Eng Technol* 5(1):179–185
43. Taher F, Werghi N, Al-Ahmad H (2012) Bayesian classification and artificial neural network methods for lung cancer early diagnosis. In: *19th IEEE international conference on electronics, circuits, and systems (ICECS 2012)*, pp 773–776
44. Tariq N (2017) Breast cancer detection using artificial neural networks. *J Mol Biomark J Diagn* 9(1):1–6
45. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A (2015) Global cancer statistics, 2012. *CA Cancer J Clin* 65:87–108
46. Ulagamuthalvi V, Sridharan D (2012) Automatic identification of ultrasound liver cancer tumor using support vector machine. In: *International conference on emerging trends in computer and electronics engineering*, pp 41–43
47. Wang X, Garibaldi JM (2005) Simulated annealing fuzzy clustering in cancer diagnosis. *Informatica* 29:61–70
48. Watanabe AT, Lim V, Vu HX, Chim R, Weise E, Liu J, Bradley WG, Comstock CE (2019) Improved cancer detection using artificial intelligence: a retrospective evaluation of missed cancers on mammography. *J Digit Imaging* 32(4):625–637
49. Zhou Z-H, Jiang Y, Yang Y-B, Chen S-F (2002) Lung cancer cell identification based on artificial neural network ensembles. *Artif Intell Med* 24(1):25–36

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.