International Journal for Modern Trends in Science and Technology, 9(02): 111-115, 2023

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0902021
Available online at: http://www.ijmtst.com/vol9issue02.html

Detection of Lung cancer from CT image using SVM classification and compare the survival rate of patients using 3D Convolutional neural network (3D CNN) on lung nodules data set

Dr.D.Suneetha[1] | N.Sai Rushitha [2] | K.Tulasi [2] | K .Gowtham [2] | K.Naga Siva Leela [2]

[1]Professor & HOD, Department of CSE, NRI Institute of Technology, India
[2]B.Tech Student, Department of CSE, NRI Institute of Technology, India

To Cite this Article
Dr.D.Suneetha, N.Sai Rushitha, K.Tulasi, K .Gowtham, K.Naga Siva Leela. Research on Detection of Lung cancer from CT image using SVM classification and compare the survival rate of patients using 3D Convolutional neural network (3D CNN) on lung nodules data set. International Journal for Modern Trends in Science and Technology 2023, 9(02), pp. 111-115. https://doi.org/10.46501/IJMTST0902021

Article Info
Received: 04 January 2023; Accepted: 01 February 2023; Published: 04 February 2023.

ABSTRACT

We are fully committed to automating everything at this point in the computer era, and the medical business is no exception, using image processing and data analytics to streamline its operations. Early identification is the most effective method for reducing cancer-related mortality. Pre-processing the medical picture or CT scan image. CLAHE Equalization is a method used to improve an image's contrast. Then, the random walk segmentation approach is used to divide it into smaller pieces. After the region of interest (ROI) in an image has been identified, segmentation will proceed to the next step, which is border correction. Finally, the final step involves slicing up the gradual pixel shift. The bulk of work consists of classifying samples as malignant or benign using a previously trained model. The aforementioned procedures all adhere to the standard approach to data analytics and picture processing. Tomorrow, the cutting-edge XGboost algorithm, which requires less data to achieve the same level of accuracy, will be implemented to improve this precision.
KEY WORDS: Lung cancer; computed tomography; deep learning; convolutional neural networks; segmentation

1. INTRODUCTION

The progression of lung cancer in both of these persons has emerged as a leading contender for the position of most commonly known cause of illness. Each year, untold numbers of people pass away as a direct result of lung cancer. The disease progresses through discrete phases, during which it starts in a localized area of tissue and then, via a process known as metastasis, spreads to all of the different parts of the lungs. The unchecked growth of unwanted cells in the lungs is what causes this condition. It is estimated that 12,203 persons had lung illness in 2016, 7130 men and 5073 women; the number of people who passed away from lung cancer in 2016 was 8839. The handling of biomedical images is the most modern and rapidly developing instrument in the field of medical research, and it is used for the early detection of cancers. In the realm of restoration, biomedical image processing algorithms may be applied to analyze diseases at an earlier stage. It makes use of many types of medical imaging, such as MRIs, X-rays, and computed tomography scans. The primary objective of image processing in the area of restorative medicine is to identify malignant development at an early stage, hence increasing the percentage of patients who survive their condition. When it comes to malignancies of the brain, lungs, and breasts, the element of time is of the utmost importance. Image processing may detect cancerous growths even in their earliest stages, which paves the way for earlier diagnosis and treatment. The process of producing the picture consists of four crucial stages: the pre-handling stage, the division stage, the stage that includes extraction, and the grouping step. This paper presents image preparation procedures in which the CT examination image is utilized as information image, is handled, and beginning period lung disease is distinguished using an SVM (support vector machine) calculation as a classifier in the grouping stage to improve accuracy, affectability, and explicitness. The picture is pre-handled and then separated into its component parts. After that, features are eliminated from the sectioned picture, and then the image is designated as either normal or destructive. The use of personal computer computations to the process of picture preparation on digital photographs is what is meant by "advanced image handling." Computerized image management, which is a branch of advanced flag preparation, provides various advantages over basic image preparation and is becoming more popular. [19] [2] It enables a much more extensive scope of calculations to be connected to the information data. The purpose of advanced image handling is to improve the image information (Features) by stifling undesirable mutilations as well as upgrade of some important image features in order that our AIComputer Vision models can benefit from this improved information to work with. An underlying arrangement of estimated information is the starting point for feature extraction, which then assembles determined qualities (features) that are proposed to be useful and non-excessive. This process encourages the resulting learning and speculation steps, and it can also sometimes prompt better human elucidations. [12] The method of feature extraction is a dimensionality reduction technique. In this process, an initial arrangement of crude components is reduced to more manageable gathers (Features) for handling, while still accurately and completely expressing the initial data collection. When the data for a calculation is too extensive to be handled in any way and it is suspected to be repetitive (for instance, a similar measurement in both feet and meters, or the redundancy of images introduced as pixels), then it is possible to change the data into a reduced arrangement of Features. This is done to make the data more manageable (additionally named a component vector). Include choice refers to the process of deciding on a subset of the underlying features. It is expected that the selected Features will comprise the relevant facts from the information, with the aim that the best task may be accomplished by employing this condensed representation rather than the whole introduction material. The number of assets that are needed to display a significant arrangement of information may be reduced as part of the feature extraction process. When doing analysis on complicated data, one of the most significant challenges comes from the large number of considerations that must be taken into account. Examination with numerous often calls for a significant amount of memory and calculating management. Additionally, it has the potential to cause an arrangement calculation to get overfit to preparation tests and add ineffectively to fresh instances. [08] The process of constructing combinations of the components to circumvent these problems while still accurately displaying the information is referred to as feature extraction. Feature extraction is a broad name for this process. There is a widespread belief among AI professionals that effectively streamlining the extraction of components is the key to productive model creation.

2. LITERATURE SURVEY

2.1 An image retrieval system that is based on the extraction of features and feedback on relevance

D. Harini, &D. Bhaskari Many academics have felt the need to create efficient techniques of retrieval based on the content of enormous multimedia databases, which has been prompted by the growth of information highways and the availability of these databases. The standard method for searching the available vast collections of multimedia data was to utilize keyword indexing or just browse through the collections, with the user's primary focus being on the greatest retrieval of material that is comparable to what they are looking for. However, content-based searching and retrieval became possible with the introduction of digital picture databases. A significant amount of investigation has been put into the process of obtaining the information based on visual characteristics such as color, texture, and shape. In this research, an effort is made to create a technique for an effective image retrieval system by extracting low-level and high-level characteristics from pictures using relevant feedback. The results of this endeavor are shown. An strategy with two phases has been used as a means of cutting down on the computing complexity and increasing the level of efficiency. The first step involves performing color segmentation and GLCM of second order statistics for the texture. The feedback that was acquired from phase1 is used in the second phase, which comprises the use of wavelets in conjunction with PCA for a more refined search and the subsequent retrieval of pictures that are comparable.

2.2 Detection of tumors in environments that are not stationary

R.N. Strickland authors provide two detectors that they employ to find simulated cancers of a fixed size in clinical gamma-ray pictures. The size of the simulated tumors is kept constant. The first method was conceived after it was discovered that even small tumors have a distinguishable signature in curvature feature space. The term "curvature" refers to the local curvature of the image data when it is viewed as a relief map. This discovery led to the development of the first method. Using a windowed statistic, the values of the computed curvature are transferred to a significance space that has been normalized. A positive detection is achieved by setting the threshold for the ensuing test statistic at a predetermined level of statistical significance. There is a significant reduction in the amount of nonuniform anatomic background activity. The second detector is an adaptive prewhitening matched filter. This filter makes use of a kind of preprocessing that is known as statistical scaling in order to prewhiten the background in an adaptable manner. Simulations of tumors with a Gaussian shape are used in the tests, and these simulations are placed on a total of twelve clinical gamma ray pictures. The curvature detector outperforms the matching filter in terms of the ratio of true positives to false positives when the tumors that need to be identified are relatively tiny (less than 3 pixels in diameter). When the local signal-to-noise ratio of the tumor-background is /splges/2, a mean true positive rate of 95% can be reached with just one false positive per picture. When used to tumors of greater sizes, a new kind of matching filter, namely the statistical correlation function introduced by Pratt, demonstrates the highest level of performance (1991).

2.3 A cutting-edge computer-assisted lung nodule identification method for computed tomography scans

M. Tan, R. Deklerck, et al.,The purpose of this work is to offer a comprehensive computer-aided detection (CAD) method for identifying lung nodules in computed tomography images. A challenging issue involving medical image analysis is used as the testing ground for a novel approach of mixed feature selection and classification for the very first time. The computer-aided design (CAD) system was educated and validated using pictures from the Lung Image Database Consortium (LIDC), which may be seen on the website of the National Cancer Institute. The detection step of the system is comprised of a nodule segmentation approach that is based on nodule and vascular enhancement filters in addition to a computed divergence feature that is used to pinpoint the centers of the nodule clusters. In the subsequent classification stage, invariant features that are defined on a gauge coordinates system are used to differentiate between actual nodules and certain types of blood vessels that are prone to easily generating false positive detections. This is done so that accurate nodule detections can be made. The performance of the innovative feature-selective classifier that is based on evolutionary algorithms and artificial neural networks (ANNs) is compared with that of two other existing classifiers, namely support vector machines (SVMs) and fixed-topology neural networks. In order to train the CAD system, we employed a collection of 235 examples that were chosen at random from the LIDC database. 125 separate instances from the LIDC database were used to test the system before it was approved for use. If the number of internal ANN nodes is carefully selected, the fixed-topology ANN classifier has an overall performance that is somewhat superior than that of the other classifiers. Because the new feature-selective classifier does not need users to make informed predictions about the number of internal ANN nodes, this classifier continues to be attractive owing to its flexibility and adaptability to the complexity of the classification issue that has to be handled. For nodules with a diameter that is more than or equal to 3 millimeters, our fixed-topology ANN classifier with 11 hidden nodes achieves a detection sensitivity of 87.5% with an average of four false positives per scan. This is the case even as the number of hidden nodes is increased. The examination of the items that gave a false positive result revealed that a sizeable number (18%) of them were tiny nodules with a diameter of less than 3 millimeters. A comprehensive computer-aided design (CAD) system that incorporates innovative characteristics is provided, and the performance of the system with three distinct classifiers is compared and examined. When compared to the performance of other approaches published in the literature, our CAD system that is equipped with any of the three classifiers performs admirably in terms of its overall performance.

2.4 Detection of nodules in lung CT images using the learning of orientation-independent contextual factors

J. Bai, X. Huang, et alThis study improves the diagnosis of pulmonary nodules in low dose computed tomography (LDCT) chest imaging by combining data-driven local contextual feature learning with model-based local shape analysis. We are able to limit the variability in appearance that is caused by orientation by using an intensity-weighted principal component analysis (PCA) technique to determine the local orientation at each candidate site. In order to represent the local environment around a nodule candidate, random comparison primitives that are specified inside a local coordinate system are used. In order to learn and classify nodule candidates using discriminative orientation-invariant contextual cues, a random forest is trained to learn and combine a subset of these primitives. The benefits of merging data-driven machine learning with geometric modeling were validated using 99 CT images from the Lung Image Database Consortium (LIDC), which are accessible to the public. The suggested strategy decreases the number of false positives produced by the model-based method used as the baseline by more than 80% and does so in a consistent manner over a broad range of sensitivity levels (70%-90%).

3.PROPOSED SYSTEM

The many steps that make up the image processing process are approached by the suggested model using a variety of different methods. In the model that has been suggested, the first step is to pre-process the CT scan picture, and then the ROI (region of interest) will be isolated in order to get it ready for segmentation. [17] During the segmentation process, a Discrete Wavelet Transform (DWT) is used, and features including correlation, entropy, variance, contrast, dissimilarity, and energy are retrieved from the data by making use of a GLCM (Gray level co-occurrence matrix). Following the step of feature extraction, the process of classification is carried out using a support vector machine (SVM) in order to differentiate between malignant and non-cancerous nodules.

4. RESULTS

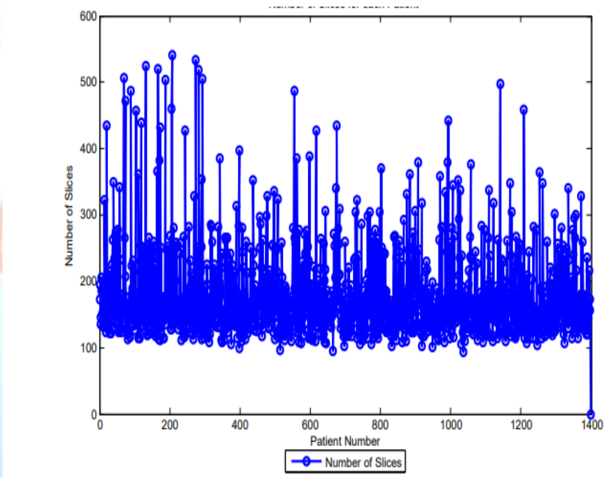

Figure 1: Number of slices per patient in data science bowl dataset

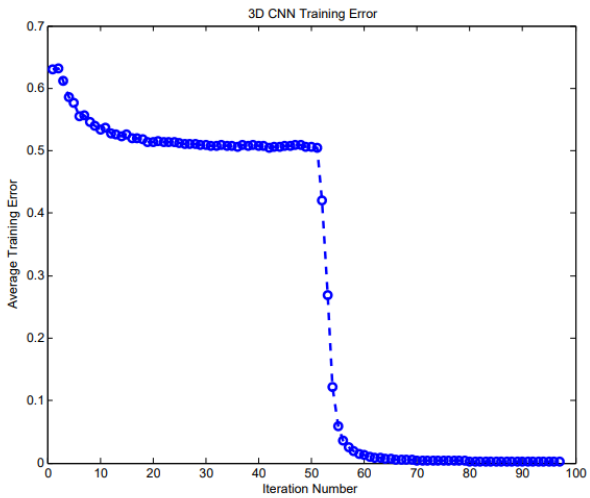

Figure 2: Average 3D CNN Training Error

5. CONCLUSION

In the principal period of the venture the Region of Interest in a picture is distinguished. The Identified district is situated in an item. The highlights in the picture are distinguished by utilizing some picture handling system. In second period of the task the component removed information is then used to arrange the picture is destructive or not utilizing a portion of the SVM – bolster vector machine grouping. At that point some boosting calculation is utilized to expand the exactness of the instrument.

In existing paper, a picture handling procedure has been utilized to recognize beginning time lung malignant growth in CT examine pictures. The CT filter picture is pre-prepared pursued by division of the ROI of the lung. Discrete waveform Transform is connected for picture pressure and highlights are extricated utilizing a GLCM. The outcomes are encouraged into a SVM classifier to decide whether the lung picture is carcinogenic or not. The SVM classifier is assessed dependent on a LIDC dataset. In future the advanced level of algorithm is used to increase the level of prediction while we are in process to include the Extreme gradient boosting Algorithm to use the data set more effectively.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

1. M. Debois, “TxNxM1 : the anatomy and clinics of metastatic cancer.Boston,” Kluwer Academic Publishers, 2006.

2. J.M. Fitzpatrick, & M. Sonka, “Medical imaging 2003: image processing. USA: SPIE Society of PhotoOptical Instrumentation Engineers,” 2003.

3.C. M. Haskell, & J.S. Berek, “ Cancer treatment. Philadelphia: W.B. Saunders, 2001

4. B. Srinivasa Rao D. Vijaya Kumar and K. Kiran Kumar, “Power quality improvement using Cuckoo search based multilevel facts controller”, Journal of Engg. Research Vol.10 No. (4A) pp. 252-261, DOI: 10.36909/jer.10895.

5.Sathish Kumar R, R. Logeswari, N. Anitha Devi, S. DivyaBharathy Efficient Clustering using ECATCH Algorithm to Extend Network Lifetime in Wireless Sensor Networks. International Journal of Engineering Trends and Technology. 45. 476-481. 10.14445/22315381/IJETT-V45P290 – March 2017

6. C. Charalambous, Conjugate gradient algorithm for efficient training of artificial neural networks, IEEE Proceedings 139 (3) (1992) 301–310.

7. B. H. Boyle, “Support vector machines : data analysis, machine learning, and applications,” Retrievedhttp://public.eblib.com/choice/publicfullrecord.aspx?p=3021500, 2011.

8. A.A. Abdulla, S.M. Shaharum, Lung cancer cell classification method using artificial neural network, Information Engineering Letters 2 (March) (2012) 50–58.

9. Sathish Kumar R , Nive tha M, Madhu mita G &Santhoshy P. Image Enhancement using NHSI Model Employed in Color Retinal Images. International Journal of Engineering Trends and Technology. 58. 14-19. 10.14445/22315381/IJETT-V58P203- April 2018.

10. D. Harini, & D. Bhaskari, “Image retrieval system based on feature extraction and relevance feedback,” ACM, 2 Penn Plaza, Suite 701, New York,NY 10121-0701, 2012.