Effective lung cancer detection using deep learning network

Vidyul Narayanan[1], Nithya P.[2], Sathya M.[3]
[1]Department of Computer Science, Pondicherry University, India
[2] Department of Computer Science, Pondicherry University, India
[3]Department of Computer Science, Pondicherry University, India
Emails: vidyulnarayananps1@gmail.com; nithyapoupathy@gmail.com; satsubithra@gmail.com

Abstract

The use of a computer-assisted diagnosis system was crucial to the results of the clinical study conducted to determine the nature of the human illness. When compared to other disorders, lung cancer requires extra caution during the examination process. This is because the mortality rate from lung cancer is higher because it affects both men and women. Poor image resolution has hampered previous lung cancer detection technologies, preventing them from achieving the requisite degree of dependability. Therefore, in this study, we provide a unique approach to lung cancer prognosis that makes use of improved machine learning and processing of images. Images of lung disease from CT scan databases created using quasi cells are used for diagnosis. Multilayer illumination was used to analyse the generated images, which improved the precision of the lungs' depiction by probing each and every one of their pixels while simultaneously decreasing the amount of background noise. Lung CT images are pre-processed to remove noise, and then a more advanced deep learning network is used to isolate the affected region. The territory is partitioned into subnetworks according to the number of existing networks, from which different features are subsequently extracted. Next, an ensemble classifier should be used to correctly diagnose lung diseases. Using MATLAB simulation, the authors examine how the provided technique improves the rate at which lung cancer could be diagnosed.

Keywords: Deep learning network; CT imaging; Multilayer illumination; quasi cells; MATLAB.

1. Introduction

False mutations, made possible by scientific progress, currently affect a great deal of the population and have a profound effect on people's daily lives. The DNA's structure and function have been drastically changed as a result of the faulty mutation. The newly generated DNA cell wrongly alters the prior DNA cell, leading to abnormal DNA cell progression. The population's actions, such as breathing air, consuming alcohol, getting surrounded by hazardous gases, and so on, contribute to the atypical mutation [2]. Tumours, which can form anywhere in an individual's body, particularly the upper body, the epidermis breasts, and the head, are mostly caused by the abnormal cell (DNA) mutations [3]. Cancerous tumours are often lethal. Lung cancer [4] ranks among the most vulnerable forms of the disease because of the many extraneous factors that might affect the cardiovascular system. The 2005 study estimates a death toll of 195,223; in 2018, the total has increased by 25%. There are 248,590 people in the USA who are affected by lung cancer, as reported by a study conducted by the North-eastern American organisation of localised cancer centres and published in 2018 [5].

Another 299,890 people in the whole of the United States of America were confirmed to have lung cancer so far this year, revealed to a poll by the American Cancer Society [6]. It includes 143,690 females and 19,450 males in this total. The study found that lung cancer was the direct cause of death for 164,703 people. The study results provide conclusive evidence that the percentage of people diagnosed with lung tumours has increased during the last five years. Among the most common diseases for which an early diagnostic [7] is sought in the field of medicine is cancer of the lungs, the results of the study show. Lung cancer is typically

Doi : https://doi.org/10.54216/JCHCI.050202
Received: December 18, 2022 Revised: March 18, 2023 Accepted: May 11, 2023

diagnosed by a doctor by looking for symptoms and signs of the disease [8]. Problems with respiration, feeling short of breath, exhaustion, reduced calorie intake, memory loss, femur breakage, pain in the muscles, headaches, neurological problem, bleeding, itchiness of the skin, voice change, and persistent sputum colour change are all symptoms associated with the blood discomfort of the the skies. Numerous screening procedures [9] have been routinely used for evaluation after these advances made an impact on a person, including genetic screening, bronchoscopy, reflex sprinting tests, fluid samples, colonoscopy, and medical examinations.

2. Literature Survey

The National Medical and Welfare Directorate provides the fundamentals needed for precise forecasting of lung disease and its stages. The diagnostic tools provide the basis for these suggestions. It is difficult to sustain the prediction performance over time, although the aforementioned testing procedures are excellent at analysing pulmonary tissue and identifying variations in cell activity that are employed for forecasting long illness. Computerised tomography [10] serves as one of the effective testing technologies that transfers X-rays onto the human anatomy to identify variations and alterations. This screening is a single of several used in the first round of elimination. In comparison to PET and MRI's lengthy screening procedures, an X-ray exam takes only twenty minutes and provides an accurate assessment of the organ's functioning and a detailed breakdown of the affected areas. An automated approach for diagnosing lung cancer employing CT scans is currently under development [11]. Methods such as noise reduction, region segmentation, feature extraction and categorization for cancer types are used in this method [12]. The area segmentation and emphasising choice criteria serves multiple purposes, as evidenced by the procedures described in [11]. This is because the contrast between healthy cells and malignant ones is best defined by the area that was correctly anticipated to be impacted. In addition, the simplified system results from the subdivided region making it simpler to separate the important cancer features. Then, feature selecting [13] cuts down on computing time for cancer forecasting, hence reducing the impact of data excessive fitting. Different approaches to segmentation [14] are used to locate positive regions in the accumulated x-ray images. The k-means clustering method, the decentralised cluster assessment, the edge-to-edge identification, the sobel recognition, the fuzzy c-means classification, the fuzzy k-means classification, the self-organized map, and the multi-layered perceptron system are all examples of such methods. After that, a number of features are extracted from the generated region, and the best features are selected using various approaches for feature extraction [15]. Wrapper approaches, ant colony strategies, algorithms for genetics, sparkle computations, and microbial optimisation methods are just a few examples of the methods used to pick useful attributes from a pool of possibilities. Several sophisticated classifiers, such as the K-nearest neighbours' method, a support vector machine, and others, are then used to categorise these distinct lung cancer characteristics.

Traditional automated methods may reliably predict instances of lung cancer; nevertheless, they have difficulty with identification rate [16] and require more time to evaluate large datasets. Also, the system can't deal with low-quality CT images, which raises the possibility of erroneous lung characteristics and, hence, the number of misclassifications [17]. The many writers then shared their individual takes on lung cancer detection, explaining how their work had helped shape a sophisticated prognostic model for the disease. According to [18], several different optimisation techniques are used in the procedure of using CT scans to test for lung cancer. Median clustering, mean clustering, particle swarm optimisation, and convergent particle algorithms are used throughout the analytic process to delve into the cancer shown in the lung CT scan. Noise in an obtained CT picture is first removed using adaptive segmentation, after which the image is improved using histogram analysis. Then, after retrieving the various properties, a similar technique is utilised to find the affected regions. Therefore, as much as 96.96% of cases of lung illness can be correctly identified using the author-proposed method. The convolutional approach aids in the procedure of diagnosing cancer of the lungs from a CT scan, as stated by [19]. The CT images are initially retrieved by the stack encoder from the LIDC. IDRI database. Different features are taught to the deep neural network so that the network can extract them automatically. By efficiently utilising many layers, it is possible to reach an identification rate of as high as 85.9% of abnormal traits. Despite the use of various approaches, misinterpretation and the handling of huge data dimensionality remain formidable obstacles.

Doi : https://doi.org/10.54216/JCHCI.050202
Received: December 18, 2022 Revised: March 18, 2023 Accepted: May 11, 2023

3. Methodology

In order to begin the task, it is necessary to eliminate noise [20] within an acquired CT lung image since this optical measuring approach includes numerous unwanted data components, electromagnetic processing elements, and particular characteristics which predominate the resulting X-ray image. The strategy for predicting lung cancer is hampered further by an abundance of superfluous information. Then, using the aid of a layered iridescent highlight's method, the identifiable noise is eliminated. This technique does a thorough examination of each and every one of the collected X-ray image's pixels. The given approach does a good job of analysing each image and its pixel density to enhance the quality of photographs as a whole. At this point in the process, the method calculates the mean value [21] of a pixel within the image to increase the picture's luminance. After the brightness of the picture was lowered to that of its neighbouring pixels, the mean amount of those pixels is used as the replacement for the pixel's original quantity. The image is segmented into several smaller sub pictures throughout this step of the process. The overall image is broken down into its component parts, and each of those parts is analysed separately.

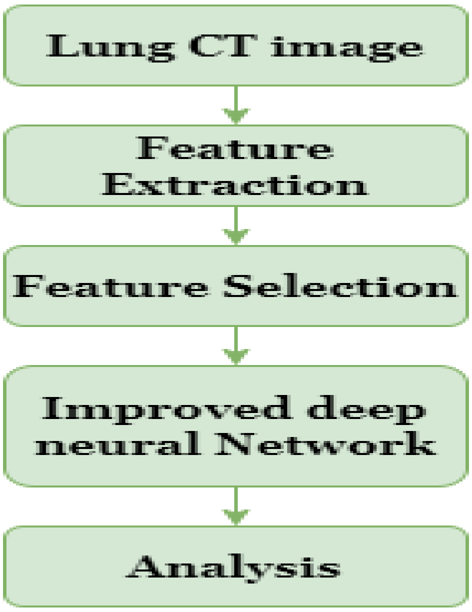

Figure 1: Workflow diagram

Taking into consideration the initial lung CT image, which is divided into two distinct sub-images based on the image intensity value xxx, the dissected pictures are designated as the foreground image and the reference image, and they are indicated by the symbols J_f and J_c respectively. The image is displayed in the following manner after it has been initialised:

Considering the first lung CT image, that is split into two parts according to the image intensity score xxx, the dissected images are labelled as the primary image along with the background image, and they are represented by the letters μ_J and μ_I, respectively.

$$P = P_c \cup P_f \quad \{1\}$$

$$P_c(a,b) = \{P((a,b) \mid P(a,b)) < \mu_J, \forall P(a,b) \in P\} \qquad \{2\}$$

$$P_g(a,b) = \{P((a,b) \mid P(a,b)) \geq \mu_J, \forall P(a,b) \in P\} \qquad \{3\}$$

Doi : https://doi.org/10.54216/JCHCI.050202
Received: December 18, 2022 Revised: March 18, 2023 Accepted: May 11, 2023

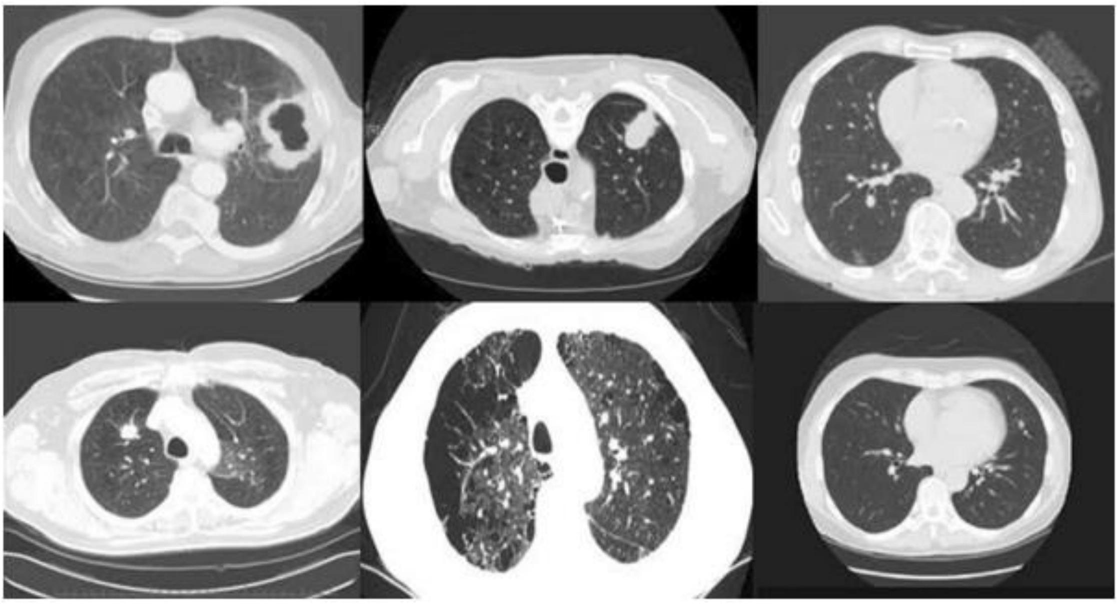

Figure 2: Sample CT image from database.

3.1 Feature extractor for lung cancer detection

The derivation of the relevant lung features is the final step that needs to be completed for this assignment [22]. The findings were achieved through the utilisation of a number of distinct features, the likes of which include variance, 3rd instant skewing, productivity, random variability, and 4th kurtosis, amongst others. These findings were acquired all the way through the separating procedure.

These characteristics originate from the whole image which was gathered, and they are collected within a recognised region in which each individual holds multiple perceptual images. The complete image was gathered by using an optical character recognition system. The findings of the study lead to an increase in the dimensions of the image. Therefore, information's complication needs to be reduced in order to offer for an enhancement in the lung cancer diagnosis process as a whole. In order to accomplish this objective, a method known as feature selection is applied to the collection of attributes with the goal to get the pulmonary features which have been improved. This is done in order to gain the optimal features.

3.2 Feature selection for lung cancer detection

The following phase involves selecting features from the lung [23] using a hybrid recommended spiral optimisation based generalised rule generation approach. The new method uses few if any control factors, produces fast results from selection, employs an area-specific search, has a simple framework, requires minimal tweaking, and singles out developments that appear easy to spot, understand, and define with minimal effort. For these reasons, this research takes advantage of the generalised rough set approach, which is based on a combination of recommended spiral optimisation, to select the best attributes from among the preselected qualities. To deal with the most effective problem when selecting features, the spiral optimisation approach [24] works as described by spiral occurrences. These speeds up the algorithm's ability to locate optimal solutions. The method accomplishes its goals by employing a well-optimized setup in the n-dimensional recurrent model, which includes the aforementioned diverging and periodic descending orientation. Optimisation features are predicted to correspond with the inquiry (global method) and exploitation with the aid of the parameters. In place of a dedicated gradient goal, this method utilises many spiral locations [25], making it simpler to select the optimal location. This method is employed in the optimisation procedure selection. When all possible permutations of a feature have been considered, the best option can be selected using the specified points.

Doi : https://doi.org/10.54216/JCHCI.050202
Received: December 18, 2022 Revised: March 18, 2023 Accepted: May 11, 2023

3.3 Classification of lung cancer using Improved Deep Neural Network (IDNN)

After the background noise in a CT scan of the lungs has been removed, the remaining step is to locate the affected region. Segmentation is the more common term for this procedure. The improved deep artificial neural network (IDNN) used in this study aids in the segmentation known to as [25] because it uses several layers when examining the image. Furthermore, the system contains a massive amount of information amassed through previous research efforts. This data was collected to aid in locating the affected area with the minimum amount of time spent processing it. In order to execute semantically comparable categorization on the lung CT image, the neural net must go through a series of steps in which it effectively analyses each pixel and predicts entire inputs appearing in the image. Pixel classification is one example of these methods. After the image's pixels have been labelled, a localization process is run so that additional data may be anticipated. This is done because the data can be used to identify normal and abnormal pixel characteristics. Last but not least, we have semantic delineation, which verifies that each pixel is properly labelled as either normal or abnormal. To create a segmented image using a deep neural network [27], several steps must be taken, including stacking the pre-trained network before attempting segmentation, selecting an input lung cancer image containing a clearly defined lung collection of data, wrapping pixel-labelled (usual and unusual) images, describing groups, and plotting the categorised images using constructed system parameters. Several of the procedures are listed here.

In this part of the segmentation phase, many hidden layers are used for analysing the input lungs CT images, and the architecture of the neural network is first constructed based on the methodologies that were covered so far. To further enhance the segmentation's precision, weight values are then calculated for each node. Lung image segmentation is shown to make use of the VGG deep learning complete network, as shown in the evaluation [28].

$$H(b) = \log(1 + e^{b}) \qquad \{4\}$$

$$K(\theta) = E_{\theta}[g(y)] = \int g(b)\pi(b \mid \theta)\mathrm{d}b \qquad \{5\}$$

4. Experimentation and result discussion

Here, we evaluate the efficacy of a cancer detection method that utilises a network of deep learning algorithms and an ensemble of classifiers. As stated before, the system is now utilising the aforementioned cancer picture data set throughout the deployment phase. We use 3600 of the images for training features and the remaining 2543 for testing. The data is used for this purpose. To analyse the segmented characteristics, MATLAB is employed. The system that was developed successfully segmented the affected region using multiple layers of neural networks. Since the segmented area is used for obtaining the important properties, its dependability must be evaluated.

Table 1: Performance of the classification algorithms (without pre-processing techniques)

Model	Accuracy	Sensitivity	Specificity	Precision	Recall
J48 classifier	85.32	87.23	88.21	88.21	89.18
NB classifier	86.12	88.21	91.19	90.24	90.79
IBK classifier	88.93	89.17	89.23	89.91	90.71
RF classifier	90.14	91.32	91.13	91.23	92.31
IDNN	**94.89**	94.72	94.42	93.72	95.71

Doi : https://doi.org/10.54216/JCHCI.050202
Received: December 18, 2022 Revised: March 18, 2023 Accepted: May 11, 2023

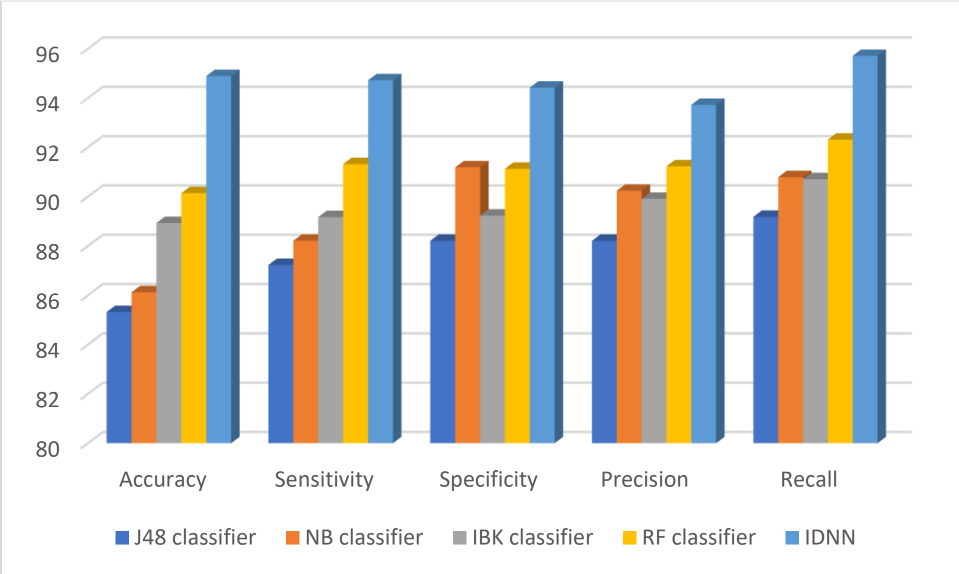

Figure 3: Comparison of performance of IDNN with existing approaches

Table 2: Performance of the classification algorithms (with pre-processing techniques)

Model	Accuracy	Sensitivity	Specificity	Precision	Recall
J48 classifier	93.31	94.23	94.82	92.95	94.17
NB classifier	96.12	93.21	95.64	92.95	94.68
IBK classifier	97.89	94.15	95.25	95.91	95.71
RF classifier	98.76	95.34	95.85	95.29	95.32
IDNN	**99.15**	95.78	95.73	95.79	95.27

Doi : https://doi.org/10.54216/JCHCI.050202
Received: December 18, 2022 Revised: March 18, 2023 Accepted: May 11, 2023

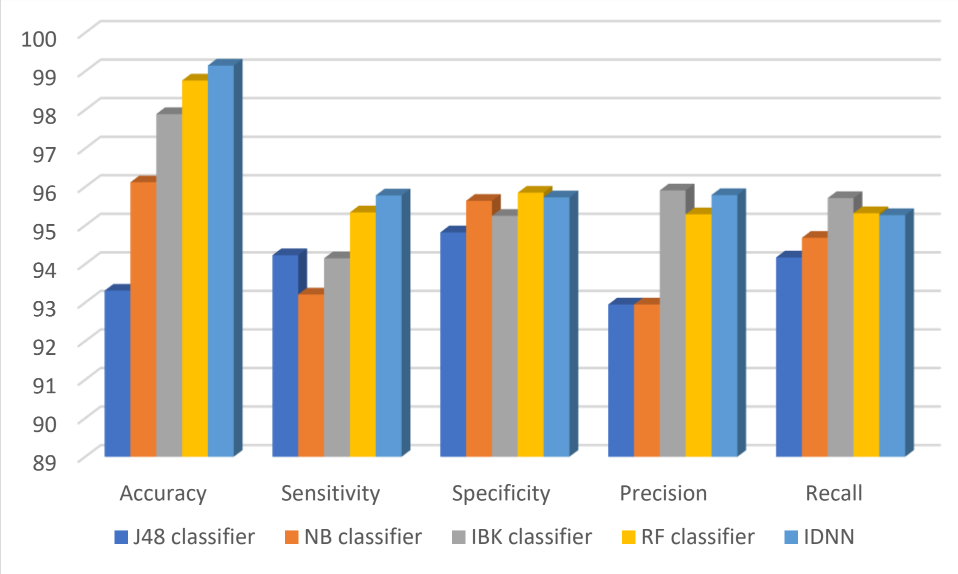

Figure 4: Comparison of performance of IDNN with existing approaches

5. Conclusion

As a result, this study investigates lung cancer by employing an enhanced deep neural network technique as well as an ensemble classification. The tumour image is obtained from the tumor image archives (CIA) collection, and the method then divides the images into two distinct categories: testing (2693) and training (2639). After that, the image quality of the gathered images is analysed in order to enhance the light levels and get rid of the noise that is contained in the CT lung image. Afterwards when, numerous layers of a network are applied to every pixel in order to analyse them and extract the damaged regions from the lung image. The segmented area is efficiently evaluated, and many characteristics are gathered, all of which are large in dimensions, which results in the identification of cancer taking significantly more time. Hence, the complexity of the network can be decreased by utilising spiral parameters and an approximating approach that picks optimum characteristics in such an effective manner. The characteristics are improved with the assistance of an ensemble classifier, which classifies the aberrant cancer features in an efficient manner. The system's efficacy is assessed based on the outcomes of the experiments, and the system detects cancer with the highest possible degree of precision.

Reference

[1] Hong Y, Hong SH, Oh YM et al (2018) Identification of lung cancer specific differentially methylated regions using genomewide DNA methylation study. Mol Cell Toxicol 14:315.

[2] Nair SS et al (2011) Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpGsequence coverage bias. Epigenetics 6:34–44.

[3] Gaudet F et al (2003) Induction of tumors in mice by genomic hypomethylation. Science 300:489–492.

[4] Tang M, Xu W, Wang Q, Xiao W, Xu R (2009) Potential of DNMT and its epigenetic regulation for lung cancer therapy. Curr Genomics 10:336–352

[5] Liu Z, Wang J, Yuan Z, Zhang B, Gong L, Zhao L, Wang P (2018) Preliminary results about application of intensity-modulated radiotherapy to reduce prophylactic radiation dose in limited-stage small cell lung cancer. J Cancer 9(15):2625–2630.

[6] Balmelli C, Railic N, Siano M, Feuerlein K, Cathomas R, Cristina V, Gu¨thner C, Zimmermann S, Weidner S, Pless M, Stenner F, Rothschild SI (2018) ‘‘Lenvatinib in advanced radioiodine-refractory thyroid cancer: a retrospective analysis of the swiss lenvatinib named patient program. J Cancer 9(2):250–255.

Doi : https://doi.org/10.54216/JCHCI.050202
Received: December 18, 2022 Revised: March 18, 2023 Accepted: May 11, 2023

[7] Manser R, Lethaby A, Irving LB, Stone C, Byrnes G, Abramson MJ, Campbell D (2013) Screening for lung cancer. Cochrane Database of System Rev 6(6):CD001991.
[8] Brock MV et al (2008) DNA methylation markers and early recurrence in stage I lung cancer. N Engl J Med 358:1118–1128
[9] Wang CC et al (2015) HOXA5 inhibits metastasis via regulating cytoskeletal remodelling and associates with prolonged survival in non-small-cell lung carcinoma.
[10] Hulbert A, Jusue-Torres, I, Stark A, Chen C, Rodgers K, Lee B, Belinsky SA (2017) Early detection of lung cancer using DNA promoter hypermethylation in plasma and sputum. Clin Cancer Res 23(8):1998–2005.
[11] Nilaiswariya, R., J. Manikandan, and P. Hemalatha. "Improving Scalability And Security Medical Dataset Using Recurrent Neural Network And Blockchain Technology." In 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), pp. 1-6. IEEE, 2021.
[12] Sriram, S., J. Manikandan, P. Hemalatha, and G. Leema Roselin. "A Chatbot Mobile Quarantine App for Stress Relief." In 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), pp. 1-5. IEEE, 2021.
[13] Kiruba, K., P. Hemalatha, J. Manikandan, M. Madhin, and Raj S. Mohan. "Revolutionizing Secure Commercialization In Agriculture Using Blockchain Technology." In 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), pp. 1-6. IEEE, 2021.
[14] Hemalatha, P., J. Manikandan, G. Leemarosilin, and P. Kanimozhi. "SEA FOOD PROCESSING USING INTERNET OF THINGS AND CLOUD TECHNOLOGIES." PalArch's Journal of Archaeology of Egypt/Egyptology 17, no. 9 (2020): 5877-5885.
[15] Manikandan, J., Brahmadesam Viswanathan Krishna, and K. Surendar. "Sign Language Recognition using Machine Learning." In 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), pp. 1-5. IEEE, 2022.
[16] Sridhar KP, Baskar S, Shakeel PM, Dhulipala VS (2018) Developing brain abnormality recognize system using multi-objective pattern producing neural network. J Ambient Intell Humaniz Comput.
[17] Shakeel PM, Tobely TEE, Al-Feel H, Manogaran G, Baskar S (2019) Neural network based brain tumor detection using wireless infrared imaging sensor. IEEE Access 7:5577–5588
[18] Senthil Kumar K, Venkatalakshmi K, Karthikeyan K (2019) Lung cancer detection using image segmentation by means of various evolutionary algorithms. Comput Math Methods Med 2019:4909846.
[19] Song QZ, Zhao L, Luo XK, Dou XC (2017) Using deep learning for classification of lung nodules on computed tomography images. J Healthc Eng 7:8314740.
[20] Venkatalakshmi K, Mercyshalinie S (2005) Classification of multispectral images using support vector machines based on PSO and k-means clustering. In: Proceedings of IEEE international conference on intelligent sensing and information processing, pp 127–133, Bangalore, India, Dec 2005
[21] Zhang X, Wang S (2012) Efficient data hiding with histogram preserving property. Telecommun Syst 49:179–185
[22] Farabet C, Couprie C, Najman L, LeCun Y (2013) Learning hierarchical features for scene labeling. IEEE Trans Pattern Anal Mach Intell 35(8):1915–1929
[23] Golmohammadi D, Creese RC, Valian H, Kolassa J (2009) Supplier selection based on a neural network model using genetic algorithm. IEEE Trans Neural Netw 20(9):1504–1519
[24] Tsai CW, Huang BC, Chiang MC (2014) A novel spiral optimization for clustering. In: Park J, Adeli H, Park N, Woungang I (eds) Mobile, ubiquitous, and intelligent computing. Lecture notes in electrical engineering, vol 274. Springer, Berlin

Doi : https://doi.org/10.54216/JCHCI.050202
Received: December 18, 2022 Revised: March 18, 2023 Accepted: May 11, 2023

[25] Tamura K, Yasuda K (2011) Spiral multipoint search for global optimization. In: International conference on machine learning and applications, vol 1, pp 470–475

[26] Raoof K, Kamoona K, Budayan C (2019) Implementation of genetic algorithm integrated with the deep neural network for estimating at completion simulation. Adv Civ Eng. https://doi.org/10.1155/2019/7081073

[27] Kong Z, Li T, Luo J, Xu S (2019) Automatic tissue image segmentation based on image processing and deep learning. J Healthc Eng. https://doi.org/10.1155/2019/2912458

[28] Havaei M, Davy A, Warde-Farley D et al (2017) Brain tumor segmentation with deep neural networks. Med Image Anal 35:18–31.

[29] K. Nanagasabapathy; G. Harish; O. I. Allen Sebastian; N. Sowrabh Chandra; V. D. Ambeth Kumar," Validation system using smartphone luminescence", IEEE International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Pages: 235 – 239, 2017. DOI: 10.1109/ICICICT1.2017.8342566

[30] V.D.Ambeth Kumar et.al. , "A Survey on Face Recognition in Video Surveillance," Lecturer Notes on Computational and Mechanism, Vol. 30, pp: 699-708, 2019

[31] V. D. Ambeth Kumaret.al , " Cloud enabled media streaming using Amazon Web Services", IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials, Pages: 195 – 198, 2-4 Aug. 2017, India (DOI: 10.1109/icstm.2017.8089150)

Doi : https://doi.org/10.54216/JCHCI.050202
Received: December 18, 2022 Revised: March 18, 2023 Accepted: May 11, 2023