International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October2023*

Efficient Heterogeneous Medical Image Data Handling for Lung Cancer Initial Screening using Deep Learning Techniques

Saroja. M[1], Victor. S.P[2],
[1]Research Scholar, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India
[1]Assistant Professor, Department of Computer Science,
Nadar Mahajana Sangam S.Vellaichamy Nadar College (Autonomous),
Affiliated to Madurai Kamaraj University, Madurai, Tamil Nadu, India
Email: msarojamca@gmail.com
[2]Associate Professor/Computer Science, St. Xavier's College (Autonomous),
Affiliated to Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India
Email: drspvictor@gmail.com

Abstract: Medical image processing in lung cancer is a complex task due to its immense impact of medical image data components associated with essential bond of lung data characteristics so that nothing will be eliminated without proper care. The generation, storage and analysis of lung cancer data from heterogeneous resources will be properly handled by the efficient approach in order to perform the optimal medical services for better cure. Deep learning plays the vital role in simulating multi layered brain network of experts with the usage of machine learning for handling this complex lung cancer image data in an efficient way. This paper presents the efficient heterogeneous medical image data handling for lung cancer initial screening using deep learning techniques. This paper concentrates on the efficient selection of deep learning approach for taken care of the complex dimensional multi resource lung cancer data units towards the best tuning in the medical field service. The future extension of this paper focuses on a real time lung cancer data interpreter model to access the lung cancer data directly using deep learning approaches.

Keywords—*Deep learning, Lung cancer, Resource, Medical service, Performance*

I. INTRODUCTION

a. Deep learning is a sub set of machine learning which is derived from artificial neural networks. It is used for learning complex patterns and relationships within data [1]. The key features are as follows:

- ✓ Deep Learning is a subfield of Machine Learning that involves the use of neural networks to model and solve complex problems [2,3].
- ✓ The key characteristic of Deep Learning is the use of deep neural networks, which have multiple layers of interconnected nodes [6,4].
- ✓ Training deep neural networks typically requires a large amount of data and computational resources. However, the availability of cloud computing and the development of specialized hardware, such as Graphics Processing Units (GPUs), has made it easier to train deep neural networks [5,7].

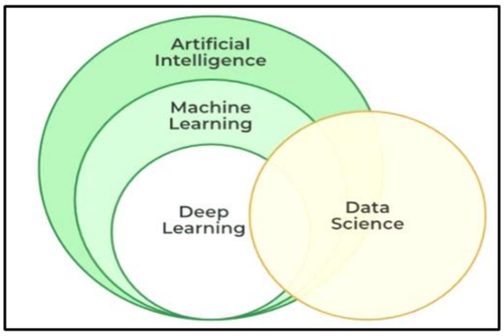


Fig-1: Deep Learning sub domain

II. LITERATURE REVIEW

❖ **David Baua , Jun-Yan Zhua , Hendrik Strobel & Agata Lapedriza, "Understanding the Role of Individual Units in a Deep Neural Network", Journal of Towards Data Science vol-21 issue-9 (2020)** This paper deals with the process of accesing the deep learning procedures with its components and actions.

❖ **Geetha gowri," Machine Learning", 9 IJRAR June 2019, Volume 6, Issue 2**

In this paper, the application of machine learning approach for the gadgets based on the user experiences.

❖ **Mudit varma ," Artificial intelligence and its scope in different areas with special reference to the field of education", International Journal of Advanced Educational Research ISSN: 2455-6157 Volume 3; Issue 1; January 2018; Page No. 05-10**

This paper deals with the different data analysis structure and functions followed by the artificial intelligence approaches and applications for medical diagnosis.

III. METHODOLOGY

The proposed methodology consists of 3 stages, they are

1. Lung cancer data handling methodology.

2. Proposed methodology design.

3. Deep learning approach for the proposed methodology.

Stage-1. Lung cancer data handling methodology

The proposed lung cancer data handling methodology consists of 5 phases. Each phase acts as an individual dimension for Lung cancer data handling the data in an efficient way.

a. Lung cancer inspection data collection types.

b. Lung cancer datasets.

c. Lung cancer image data formats.

d. Lung cancer medical data quality.

e. Data originality/authenticity.

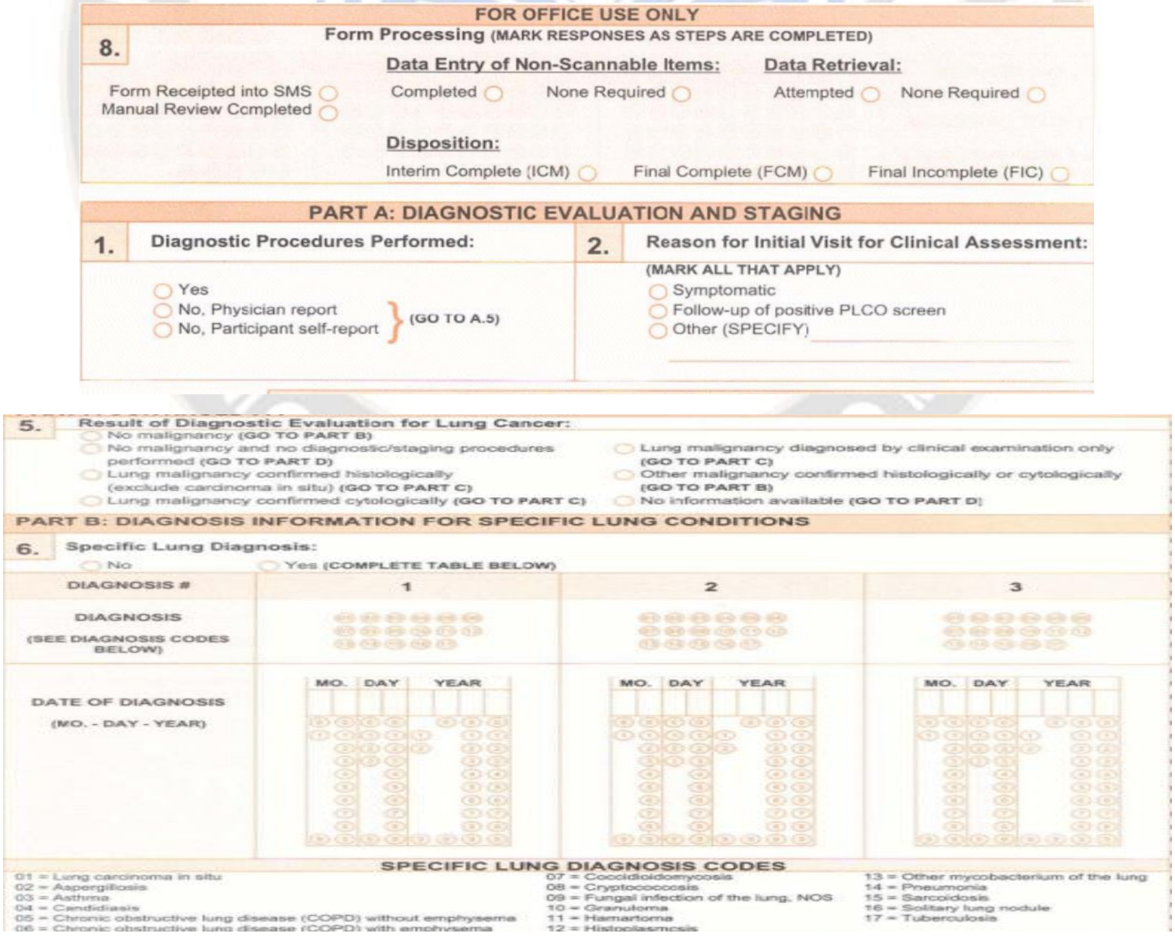
a. Medical inspection data collection types:

The data type collections associated with lung cancer follows certain medical procedural outputs they are as follows:

i. Physical check-up-procedure-Obtained through questioners and tests.

It includes the physical examination and questioner part from the medical practitioner/expert/Nurse from the patient as his main source of data.

The components include the following structure as in fig-2.


FOR OFFICE USE ONLY

8. Form Processing (MARK RESPONSES AS STEPS ARE COMPLETED)

Form Receipted into SMS ◯
Manual Review Ccmpleted ◯

Data Entry of Non-Scannable Items: Completed ◯ None Required ◯

Data Retrieval: Attempted ◯ None Required ◯

Disposition: Interim Complete (ICM) ◯ Final Complete (FCM) ◯ Final Incomplete (FIC) ◯

PART A: DIAGNOSTIC EVALUATION AND STAGING

1. Diagnostic Procedures Performed:
◯ Yes
◯ No, Physician report
◯ No, Participant self-report } (GO TO A.5)

2. Reason for Initial Visit for Clinical Assessment: (MARK ALL THAT APPLY)
◯ Symptomatic
◯ Follow-up of positive PLCO screen
◯ Other (SPECIFY)

5. Result of Diagnostic Evaluation for Lung Cancer:
◯ No malignancy (GO TO PART B)
◯ No malignancy and no diagnostic/staging procedures performed (GO TO PART D)
◯ Lung malignancy confirmed histologically (exclude carcinoma in situ) (GO TO PART C)
◯ Lung malignancy confirmed cytologically (GO TO PART C)
◯ Lung malignancy diagnosed by clinical examination only (GO TO PART C)
◯ Other malignancy confirmed histologically or cytologically (GO TO PART B)
◯ No information available (GO TO PART D)

PART B: DIAGNOSIS INFORMATION FOR SPECIFIC LUNG CONDITIONS

6. Specific Lung Diagnosis:
◯ No ◯ Yes (COMPLETE TABLE BELOW)

DIAGNOSIS #	1	2	3
DIAGNOSIS (SEE DIAGNOSIS CODES BELOW)			
DATE OF DIAGNOSIS (MO. - DAY - YEAR)	MO. DAY YEAR	MO. DAY YEAR	MO. DAY YEAR

SPECIFIC LUNG DIAGNOSIS CODES

01 = Lung carcinoma in situ
02 = Aspergillosis
03 = Asthma
04 = Candidiasis
05 = Chronic obstructive lung disease (COPD) without emphysema
06 = Chronic obstructive lung disease (COPD) with emphysema
07 = Coccidioidomycosis
08 = Cryptococcosis
09 = Fungal infection of the lung, NOS
10 = Granuloma
11 = Hamartoma
12 = Histoplasmosis
13 = Other mycobacterium of the lung
14 = Pneumonia
15 = Sarcoidosis
16 = Solitary lung nodule
17 = Tuberculosis

Fig-2: Sample Physical test evaluation report

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October2023*

ii. Chest-X-ray inspection-Obtained through X-ray machine in Labs.

The labs associated with the hospital either internal or external facility is used to perform the chest-X-ray inspection. The corresponding expert/technician completes the X-ray taking process and forwards the report to the hospital for further processing. The sample chest X-ray report as in fig-3.

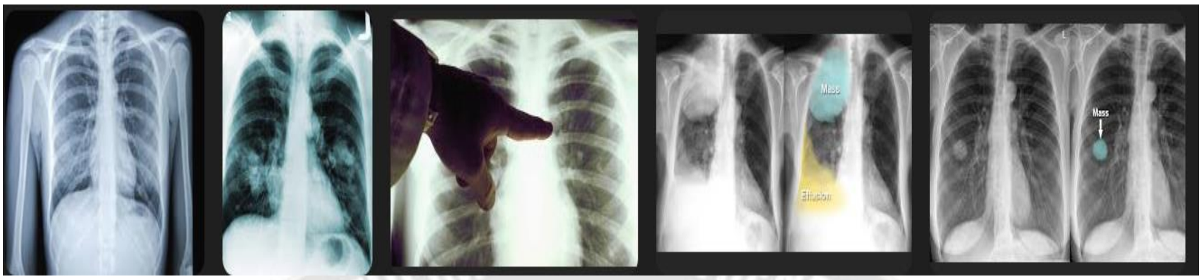


Fig-3: Sample Chest X-ray report

iii. Scan results-Obtained through Scan devices in Labs.

Normally CT-scan inspection is suggested for the final confirmation of lung cancer stages starting from initial stage-0 to stage-4 of critical condition. Moreover CT-Scan radiation itself may acts as a catalyst for disease improvisation due to frequent exploitation of scan radiation. The sample CT-scan images are as follows as in fig-4:

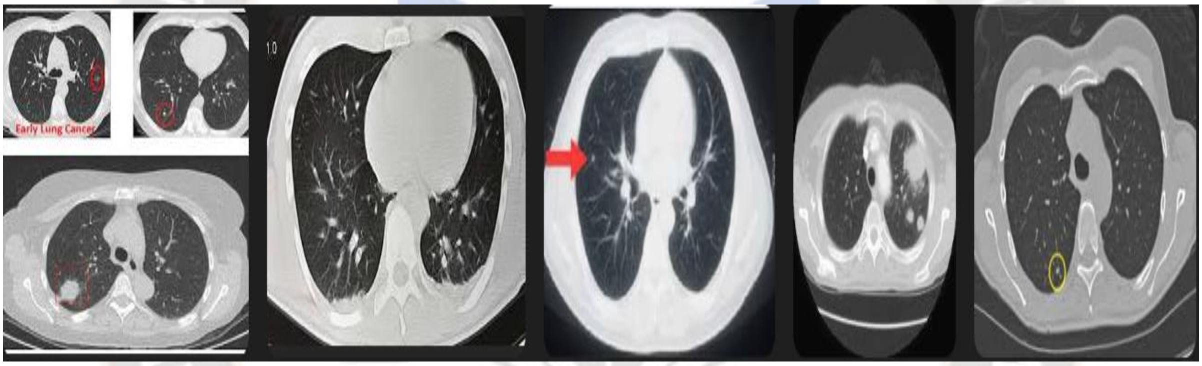


Fig-4: Sample CT-Scan report

b. Lung cancer medical data sets:

There are 3 types of datasets they are structured, semi structured and unstructured lung cancer datasets in which structured data sets are complete and easily handled directly by the medical data handling approaches. Semi structured is not in relational database format whereas unstructured data are huge complex video or audio collections on lung cancer information's. The process of handling datasets are as follows in the table-1:

Table-1: Data set table

Data Set	**Format**	**Access method**	**Handling Tools**
Structured	Relational Tables	Queries	SQL
Semi structured	Collections	Key-value pair	JavaScriptObjectNotation
Unstructured	Audio, Video	Schema-on-read	Data Graph converter

c. Lung cancer image data formats:

The lung cancer CT scan images are in JPEG format, but for the international standard towards unique access the image Digital Imaging and Communications in Medicine format called as DICOM format is used in the current practice for Lung cancer analysis. The sample DICOM image file format is as in Fig-5.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October2023*

Fig-5: DICOM image format file

d. Lung cancer data quality:

The lung cancer data quality represents the clarity and standard for the experts to understand, believe and respond towards the current patient conditions. Since lung cancer medical data are highly sensitive the process of achieving the quality data is very essential. The fuzzy membership assignment produces the lung cancer data quality level as in table-2.

Table-2: Data quality table

Sl.No	**Quality factor**	**Fuzzy membership value-Fdq$_i$**
1	Accessibility	0.1
2	Accuracy	0.1
3	Completeness	0.1
4	Consistency	0.1
5	Integrity	0.1
6	Reasonability	0.1
7	Timeliness	0.1
8	Uniqueness	0.1
9	Validity	0.1
10	Fitness for purpose	0.1

e. Lung cancer medical data Authenticity.

The authenticity of a medical data in lung cancer screening is very important and sensitive due to its impact on the human life. It must be properly handled in order to deliver the prompt information to the patient. The authenticity plays the vital role for acknowledging the information for further processing. The following table-3 represents the assessment computation for lung cancer image data authenticity.

Table-3: Data authenticity table

Sl.No	**Authentic Component**	**Authenticity value Avi**
1	Image data origin	0.3
2	Image data integrity	0.3
3	Image data reliability	0.2
4	Image data robustness	0.2

Stage-2. Proposed methodology design for Lung cancer Screening:

The proposed methodology lung cancer screening is as follows in fig-6.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October2023*

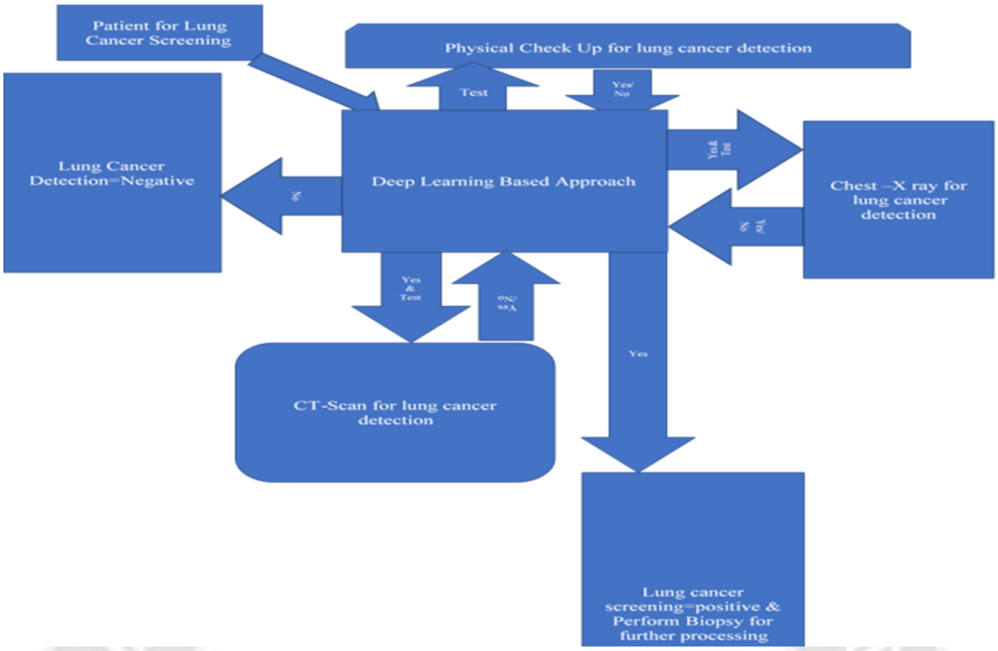


Fig-6: Proposed methodology Lung cancer screening design using deep learning approach.

Stage-3. Deep learning approach for the proposed methodology:

Deep learning approaches entirely depend on convolution neural network, Recursive neural network and generative adversarial neural networks. In this proposed method we used the combination of CNN and RNN for the efficient heterogeneous medical image data handling for lung cancer initial screening using deep learning techniques. Since machine learning executes individual tasks separately whereas deep learning perform the combined dependent execution of the proposed methodology components in an efficient manner as in fig-7.

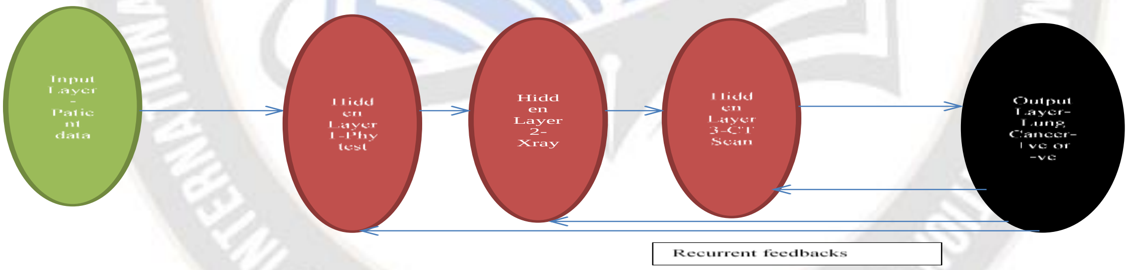


Fig-7: Deep learning approach implementation graph

IV. IMPLEMENTATION

The implementation part of this research module concentrates on the individual component implementation towards the proposed methodology with proper care.

a. Lung cancer medical inspection data collection types:

i. Physical check-up procedure.

The physical check-up includes the following components with image references if any for clarified verification as in below structure-4.

Table-4: Physical test structure

No	Component	Result State	Patient feeling	Expert observation
			value-Max=Pmx	value-Max=Emx
1	Smoking state-lips image if any	Yes/No	0.3	0.3
2	Workplace	Yes/No	0.1	0.1
3	Shortness of breath	Yes/No	0.1	0.1
4	Chest pain	Yes/No	0.1	0.1
5	Cough	Yes/No	0.1	0.1
6	Fatigue	Yes/No	0.1	0.1

7	Wheezing	Yes/No	0.1	0.1
8	Swelling in feet-image	Yes/No	0.025	0.025
9	Swelling in arms-image	Yes/No	0.025	0.025
10	Heredity	Yes/No	0.5	0.5

Compute $Phy(X)= \sum_{i=1}^{10} (Pmxi+Emxi)/2$ such that Phy(X)>0.5 represents the physical test initially suspected the case as an initial doubtful subject for lung cancer.

ii. Chest-X-ray

Chest-X-ray is used to diagnose the lung cancer for the second level of conformity. The mathematical computation for the chest –X-ray is as follows:

ChstX(Z)=Yi+Zi where 0<=Yi<=0.5 and 0<=zi<=0.5

Such that Yi=Medical Experts review of X-ray image on

Yi = 0 if visible tumours/nodules/masses=NO

0.5 if visible tumours/nodules/masses=YES

Zi= 0 if fluid build-up in lungs =NO.

0.5 if fluid build-up in lungs=YES.

Compute ChstX (Z) such that ChstX (Z) >= 0.5 represents the Chest-X-ray suspected the

i=1

case as a next level subject for lung cancer.

iii. Low Dose Computer Tomography

The LD-CT scan is recommended rare for the final screening confirmation. It may provide results

If (lung nodules present=false) then

Goto x:

Else if (tumour size > normal) then

Goto y:

Else if (lymph node involvement =false) then

Goto x:

y: Perform Biopsy; exit;

x: False positive perform Magnetic Resonance Imaging scan for confirmation

End if

b. Lung cancer medical data sets:

The process of handling datasets are as follows in the below architecture:

Table-5: Data set architecture

Data Set	**Format**	**Access method**	**Handling Tools**
Structured	Relational Tables	Queries	SQL
Semi structured	Collections	Key-value pair	JavaScriptObjectNotation
Unstructured	Audio, Video	Schema-on-read	Data Graph converter

c. Lung cancer image data formats:

The free online converters or the familiar aspose, microdicom portals are used to convert JPEG to DICOM formats.

d. Lung cancer data quality:

The quality factor $Qf (Patient\text{-}j) = \sum_{i=1}^{10} Fdq_i$

If Qf(Patient-j) < 0.5 then

Data quality=Poor

Else if 0.5< = Qf(Patient-j) < =0.6 then

Data quality=Average

Else if 0.7< = Qf(Patient-j) < =0.9 then

Data quality=Good

Else if Qf (Patient-j) > 0.9 then

Data quality=perfect

e. Lung cancer medical data Authenticity.

LCIDAi=Av (Image data origin) +Av (Image data integrity) +Av (Image data reliability) +Av (Image data robustness)

If LCIDAi (Patient-j) > 0.8 then

***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October2023*

Authenticity = Genuine

else

Authenticity= Corrupted.

f. Lung cancer Data handling tools:

Lung cancer data handling tools are available in the open market with valuable performance in storage and accessing features. They are as follows,

1. HeathPlix
2. GNU Health
3. OpenMRS
4. FreeMED
5. OpenEMR etc.

The following real time sample demo implementation image using Open MRS [8] which is one among the very useful and flexible tool that shows its performance effectiveness in storage and accessing nature with dummy name and ID as in fig-8.

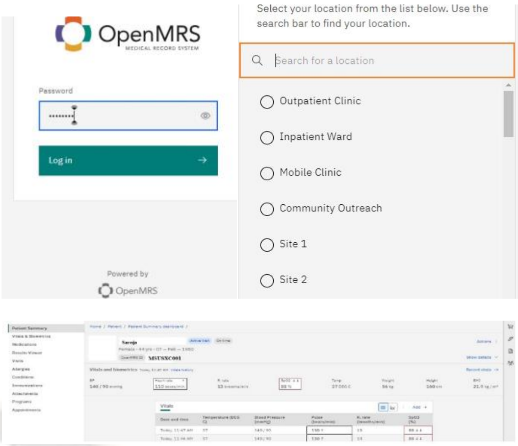


Fig-8: Data handling implementation tool for Lung cancer screening

V. RESULTS AND DISCUSSION

The proposed methodology uses the standard data sets from NLST Lung cancer data dictionary and also from the collection of real time images consisting of 3150 images. The experimental results are tabulated with highly suspected 1520 patient in the below summary structure.

Table-6: Result computation summary structure

No	**Patient count**	**Physical test Success Count**	**Chest X-ray Success Count**	**CT-Scan Success Count**	**Success Count**	**Ambiguity Count**
1	343	87	118	134	339	4
2	412	61	135	211	407	5
3	229	61	62	103	226	3
4	536	78	193	260	531	5

The following graph as in fig-9 shows the proposed methodology execution performance based on the table-6.

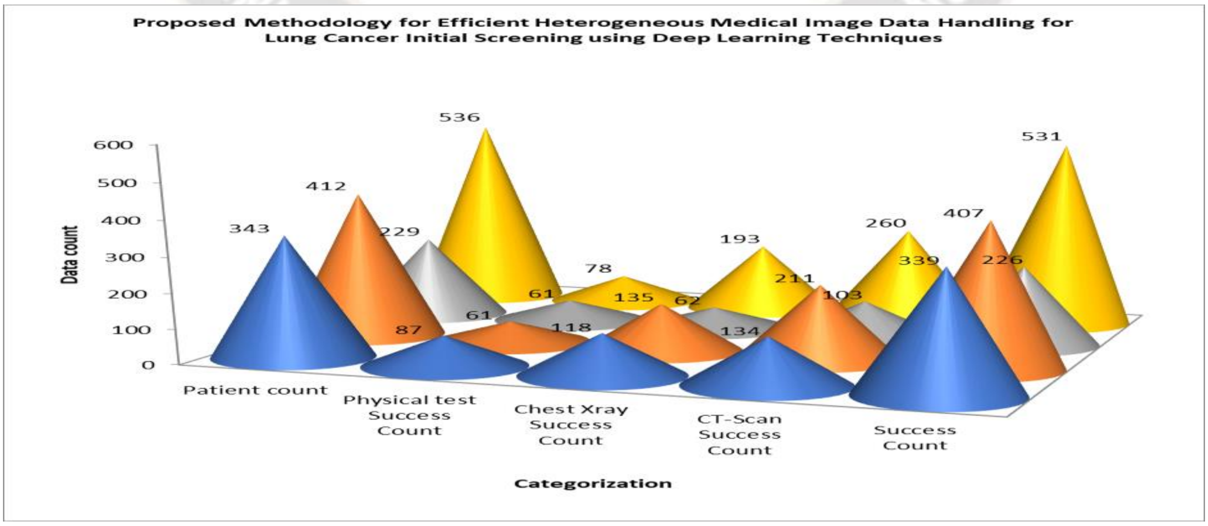


Fig-9: Proposed methodology for lung cancer screening success rate graph

The performance of the proposed methodology for efficient heterogeneous medical image data handling for lung cancer initial screening using deep learning techniques shows that 1502 out of 1520 lung cancer image data produces success in the screening for lung cancer positive which achieves 99% of success.

While comparing the proposed methodology results with existing image processing approaches which produces only 865 out of 1520 lung cancer image screening tests of success which achieves only 57% of success rate.

The parametric comparison between existing and proposed methods with precision,accuracy etc are represented in the below format ,

Table-7: Proposed methodology Parametric comparisons

No	Approach	Accuracy	Precision	Recall	F1 score value
1	Image processing method	69%	0.68	0.72	0.77
2	Proposed efficient heterogeneous medical image data handling for lung cancer initial screening using deep learning techniques	98%	0.99	0.98	0.99

The following fig-10 shows the performance comparison between the proposed and existing methodologies.

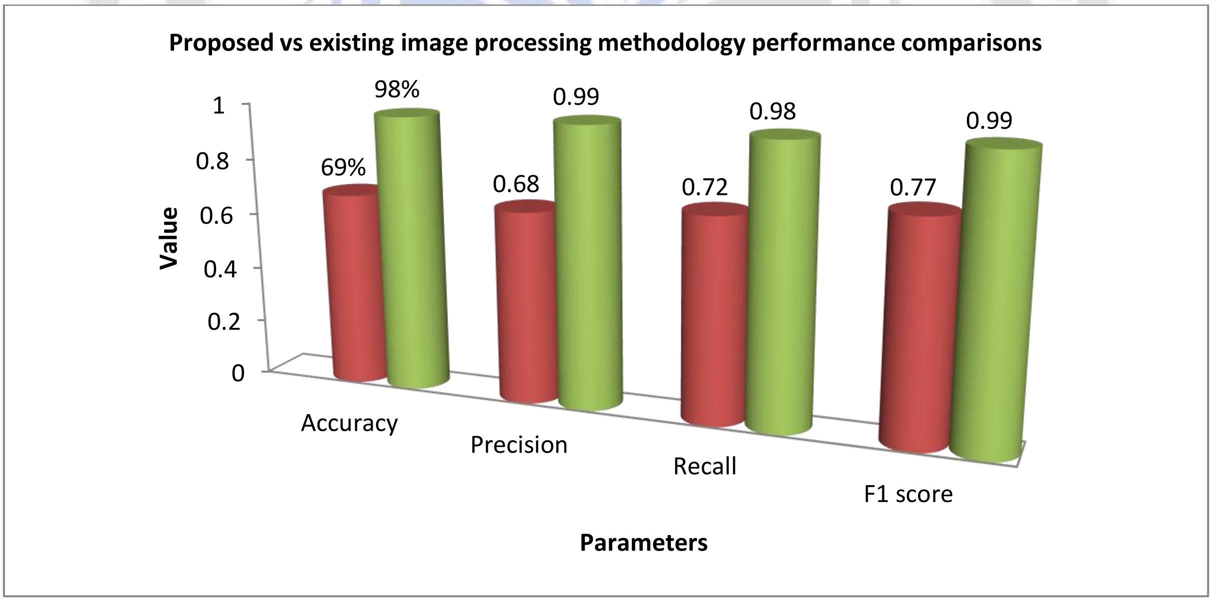


Fig-10: Proposed vs. existing image processing methodology performance comparisons

VI. CONCLUSION

Deep learning is now a needed component in the lung cancer screening environment. The steps for focusing on different data access control methods for lung cancer patient improvement system is an essential criterion for the development of the medical field. This research module set its goal for the efficient heterogeneous medical image data handling for lung cancer initial screening using deep learning techniques. The first part of this research task is to focuses on the physical test for the lung cancer screening with text, image and videos with dicom format conversions. Then this research article performs the Chest-X-ray Scan result analysis using the computation modelling from deep learning techniques. The final module focuses on the CT-Scan analysis verification for improved lung cancer screening procedure. The experimental results after several observations of this proposed research methodology yields the correct results for 1502 out of 1520 lung cancer image data component analysis for the patients. This research will be further enriched with artificial intelligence based automation methodology for optimized responses in this lung cancer screening domain.

REFERENCES

[1] P. Dande and P. Samant, “Acquaintance to Artificial Neural Networks and use of artificial intelligence as a diagnostic tool for tuberculosis: A review,” Tuberculosis, vol. 108, pp. 1–9, 2018, doi: https://doi.org/10.1016/j.tube.2017.09.006.

[2] Y. Liu, R. Sun, and S. Jin, “A survey on data-driven process monitoring and diagnostic methods for variation reduction in multi-station assembly systems,” Assem. Autom., vol. 39, no. 4, pp. 727–739, Jan. 2019, doi: 10.1108/AA-10-2018-0174.

[3] S.-L. Wamba-Taguimdje, S. Fosso Wamba, J. R. Kala Kamdjoug, and C. E. Tchatchouang Wanko, “Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects,” Bus. Process Manag. J., vol. 26, no. 7, pp. 1893–1924, Jan. 2020, doi: 10.1108/BPMJ-10-2019-0411.

[4] C. Kahraman, İ. Kaya, and E. Çevikcan, “Intelligence decision systems in enterprise information management,” J. Enterp. Inf. Manag., vol. 24, no. 4, pp. 360–379, Jan. 2011, doi: 10.1108/17410391111148594.

[5] F. F. Kharbat, A. Alshawabkeh, and M. L. Woolsey, “Identifying gaps in using artificial intelligence to support students with intellectual disabilities from education and health perspectives,” Aslib J. Inf. Manag., vol. 73, no. 1, pp. 101–128, Jan. 2021, doi: 10.1108/AJIM-02-2020-0054.

[6] M. Jain, A. Goel, S. Sinha, and S. Dhir, “Employability implications of artificial intelligence in healthcare ecosystem: responding with readiness,” foresight, vol. 23, no. 1, pp. 73–94, Jan. 2021, doi: 10.1108/FS-04-2020-0038.

[7] M. Newman, ―Finding community structure in networks using the eigenvectors of matrices,‖ Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), vol. 74,no. 3, 2006. [Online]. Available: http://dx.doi.org/10.1103/PhysRevE.74.036104.

[8] https://openmrs.org/