INVENTION DISCLOSURE FORM FOR PATENTS

Applicant Name-Marwadi University

1. **Particulars of Inventors**

Mr./Ms/ Dr.	Name (Full)	Department	Designation	Mobile No.	Email	Postal Address
Mr.	Shailendrasinh Chauhan	Computer Engineering	Assistant Professor	84601 14578	Shailendrasinh .chauhan@mar wadieducation. edu.in	Department of Computer Engineering, Marwadi University, Rajkot - 360003
Mr.	Shubhamkuma r Pandit	Computer Engineering	Student	9978342526	Panditshubha m283@gmail. com	Rajkot, Gujarat
Ms.	Neeharika Joshi	Computer Engineering	Student	8780623173	Joshineeharika 981@gmail.co m	Rajkot, Gujarat

2. **Provide title of the invention:**
   Lung Cancer Prediction Using SHAP and DL Techniques

3. **In 100 words or less, please provide an abstract or summary of the invention:**

   The invention pertains to an AI-based system of early lung cancer detection combining clinical information with CT-scan data. It makes use of machine learning algorithms, such as Random Forest, Logistic Regression, and DenseNet121, and complemented with explainable AI methods, like SHAP and Grad-CAM. The system classifies cases as benign, malignant, or normal and gives interpretable feature-level information to gain clinical trust. The invention, through the combination of multimodal data and transparent predictions, can successfully detect lung cancer accurately, reliably and in time, thereby enhancing the diagnostic efficiency and in decision-making related to healthcare.

4. **Detail description of the invention:(Answer to all below are required in detail)**

   a. **Problem the invention is solving**
   Lung cancer can be detected at advanced stages since no proper and readable forms of early detection can be taken, and thus treatment may be delayed leading to death. The current models of AI lack transparency (black-box) and cannot incorporate multimodal data (clinical records and CT scans), which restricts their use in the clinical environment. The solution to this would be this invention that offers a very precise, explainable, and multimodal diagnostic system that enables

early diagnosis, less misdiagnosis, and builds trust among the healthcare professionals in their ability to provide medical care effectively and in a timely manner.

b. General Utility/application of the invention

In clinical diagnostics and healthcare systems, the invention can be used in support of an early and accurate diagnosis of lung cancer. It helps physicians by integrating both CT scan analysis and patient clinical data giving interpretable AI-driven forecasts to make timely decisions. In addition to hospitals and diagnostic centers, the invention can be incorporated in telemedicine systems, health monitoring, and extensive screening programs to minimize diagnostic errors and enhance survival rates and resource consumption in the general health.

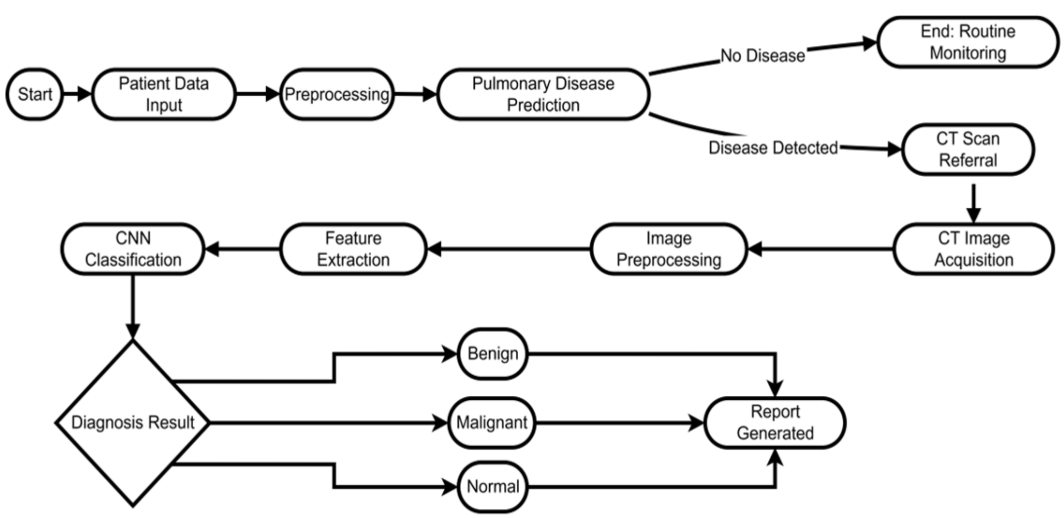
c. Advantages of the invention disclosing about the increased efficiency/efficacy

The invention has a notable effect on the efficiency and effectiveness of lung cancer diagnosis as the multimodal data (clinical features and CT scans) are integrated in one predictive system. The SHAP and Grad-CAM-optimized machine learning models can be trusted and relied upon by clinicians to make predictions, which are both accurate and understandable. This minimizes problematic errors in diagnosis, increases early diagnosis, and enhances faster clinical decision-making. Relative to traditional techniques, the system offers greater predictive capability, clear line of thought and a smooth fit with digital healthcare workflows, thus improving both the reliability of the diagnosis and patient outcomes.

d. Best way of using the invention as well as possible variants

It is most useful as a clinical decision-support system embedded into hospital diagnosis systems with the clinical records and CT scans of the patients processed to provide precise and explainable predictions in real time. The use of a web or cloud-based deployment guarantees both urban and remote healthcare centers accessibility. It can be variated in possible ways: it may be integrated with telemedicine applications (remote-screening), adapted to mobile health devices (preliminary risk assessment), and extended to other types of cancers or pulmonary illnesses (retraining the system with corresponding datasets). Combination of several AI algorithms into hybrid models can further be used to increase predictive accuracy and robustness.

e. Working of invention along with Drawing, schematics and flow diagrams if required with complete explanations



The process of the invention starts with the entry of patient clinical information including age, smoking history, symptoms, and oxygen saturation. This information is pre-processed to achieve consistency and is analysed with machine learning algorithms to predict pulmonary diseases. In case no disease is identified, then the patient is recommended routine monitoring. The system orders the patient to receive CT scan imaging in cases where the disease indicators are available. The obtained CT images are pre-processed and significant features are identified before being given to a Convolutional Neural Network (CNN) to be classified. CNN model subdivides the scans into benign, malignant and normal condition. These findings, along with clinical data analysis, make up a complete diagnostic final product. Lastly, the system produces a structured report, which contains not only the results of the prediction but also demystified information, including the contributions of the features and the lung regions highlighted, and hence helps healthcare professionals in taking accurate and reliable clinical decisions.

5. **Have you conducted Primary Patent Search? Yes / No (if yes, attach the patent search results)**
   No

6. **Existing state-of-the-art and prior arts: (Brief background of the existing knowledge/product/process in the market)**
   Existing methods of lung cancer prediction mainly use machine learning and deep learning models on clinical data or medical imaging. Some have used the algorithms of Random Forest, Support Vector Machines, and Neural Networks on patient records, and deep CNN structures (e.g., ResNet, DenseNet) with CT scans. Despite their high accuracy, these methods have important limitations: they are mostly black-box models, which cannot be interpreted, most are trained on small or unbalanced data, and few incorporate multimodal data that combines clinical features and imaging. Additionally, current business systems have tended to give image-level analysis with no commentary at the feature level, decreasing physician confidence and uptake. Past literature has also identified the problems of overfitting, low generalizability and low sensitivity to detecting disease in its initial stages. Therefore, an elaborate, explainable, and multimodal diagnostic framework, which enhances accuracy and transparency to health care professionals, is still required even in light of the great progress.

7. **List out the known ways about how others have tried to solve the same or similar problems? Indicate the disadvantages of these approaches. In addition, please identify any prior art documentation or other material that explains or provides examples of such prior art efforts.**

S. No.	Existing state of art	Drawbacks in existing state of art	Overcome (how your invention is overcoming the drawback)
1.	S. Murthy Nimmagadda et al., *"Lung Cancer Prediction and Classification Using Machine Learning Algorithms"*, ICOECA 2024	Relies only on clinical datasets; limited dataset size; models act as black-box with no explainability.	Our invention integrates both clinical and CT scan image data and applies SHAP/Grad-CAM for explainability, ensuring transparency and trust.
2.	Mohammad Shafiquzzaman Bhuiyan et al., *"Advancements in Early Detection of Lung Cancer in Public Health"*, JCSTS 2024	Focused on accuracy but lacked integration of multimodal (text + image) data; no real-time deployment framework.	Our system combines multimodal data and provides a web-based deployment using FastAPI, enabling practical clinical usage.

3.	R. K. Pathan et al., *"The efficacy of machine learning models in lung cancer risk prediction with explainability"*, PLoS One 2024	Provided explainability but limited to clinical data; no image-level transparency; dataset imbalance affected performance.	Our invention integrates image explainability (Grad-CAM) along with SHAP for text features, ensuring robust multimodal interpretability.
4.	P. Chaturvedi et al., *"Prediction and Classification of Lung Cancer Using Machine Learning Techniques"*, IOP Conf. Ser. 2021	Focused mainly on classification accuracy; did not address overfitting, dataset imbalance, or feature-level contributions.	Our invention applies advanced ensemble learning, deeper feature engineering, and interpretable AI techniques to ensure stable and generalizable predictions.
5.	Joy Chakra Bortty et al., *"Optimizing Lung Cancer Risk Prediction with Advanced Machine Learning Algorithms"*, JMHS 2024	Applied ML optimization but lacked transparency; no provision for clinical decision support or doctor-friendly outputs.	Our invention generates interpretable reports (feature contributions + heatmaps) that can be directly used by clinicians for decision-making.
6.	M. Mamun et al., *"Lung Cancer Prediction Model Using Ensemble Learning Techniques"*, AIIoT 2022	Used ensemble methods but restricted to textual datasets; no integration of CT scans or XAI frameworks.	Our invention fuses ensemble learning with multimodal (text + CT scan) inputs and explainability, giving superior accuracy and reliability.
7.	K. Moon & A. Jetawat, *"Predicting Lung Cancer with K-Nearest Neighbors (KNN): A Computational Approach"*, IJST 2024	Simpler algorithm (KNN) showed low scalability and poor performance on high-dimensional data; no explainability.	Our invention leverages advanced models (Random Forest, DenseNet121) optimized for large datasets, while ensuring interpretability with SHAP and Grad-CAM.
8.	S. Agarwal et al., *"Prediction of Lung Cancer Using Machine Learning Techniques and Their Comparative Analysis"*, ICRITO 2022	Comparative study of ML algorithms but lacked integration of real-world deployment and explainability.	Our invention not only compares algorithms but deploys the best-performing models in a real-time web-based clinical tool with interpretability.
9.	T. Kadir & F. Gleeson, *"Lung Cancer Prediction Using Machine Learning and Advanced Imaging Techniques"*, TLCR 2018	Applied imaging techniques but without integration of clinical features; limited real-world usability.	Our invention integrates imaging with clinical attributes for holistic multimodal prediction, enhancing both sensitivity and specificity.
10.	A. Ampuh Yunanto, *"International Electronics Symposium"*, IEEE 2020	Focused on conventional ML methods with limited dataset coverage; lacked multimodal framework and clinical adaptability.	Our invention employs both ML and DL models on larger, diverse datasets, combining multimodal inputs for robust and clinically viable predictions.

8. **List the Technical features and Elements of the invention along with the Description of your invention from start to end.**
The invention is a multimodal diagnostic system that will predict lung cancer both clinically and with imaging information at an early stage. The process is initiated by taking of data whereby patient attributes are taken including age, smoking history, oxygen saturation, family history and symptomatic features of the suspected cases accompanied with CT scan images. The data is first pre-processed to be high quality: clinical data is cleaned, normalized, and coded, and CT images are resized, normalized, and augmented to enhance the learning rate. The mechanism of feature extraction then proffers the identification of key attributes in the clinical data and it learns the image features automatically with the deep CNN layers.

Machine learning classifiers like Random Forest and Logistic Regression are trained on the clinical data and CNN models like DenseNet121 and ResNet50 are used to classify CT scans into benign, malignant, or normal scans. To solve the black box character of AI, the invention is built to include an explainability component: SHAP calculates feature-wise contributions to clinical predictions, and Grad-CAM produces heatmaps indicating abnormal lung areas in CT scans. Both streams of data make predictions that are combined in a decision fusion component, which increases reliability and minimizes misclassification.

The system produces a detailed diagnosis report to not only indicate the classification outcome, but also describe the contributing clinical features and identify suspicious lung areas. This provides clinical trust and transparency. Lastly, the invention is implemented using FastAPI as the backend and a web-based interface to interface with healthcare professionals by letting them enter patient information, submit CT scans, and receive real-time diagnostic findings. The invention is flexible to hospitals, diagnostic centers, and telemedicine sites, hence enhancing efficiency, accuracy, and early case of lung cancer.

9. **List out the features of your invention which are believed to be new and distinguish them over the closest technology.**
The invention is characterized by a number of new features unavailable to the current technologies. Contrasting with the previous methods that utilize only one of the two sources of data, such as clinical data or imaging data, this system is the first of its kind in which the two are coupled together to offer a multimodal paradigm of lung cancer prediction. It uses explainable AI techniques, such as SHAP used on clinical data and Grad-CAM used on CT scans, that improve accuracy and also allow medical professionals to understand and trust the predictions rather than models that are frequently black-box models with unknown parameters. The decision fusion module also improves reliability as it fuses machine learning and deep learning models and hence misclassification is minimized. Moreover, the invention offers an entire diagnostic pipeline, end-to-end, which includes data preprocessing, feature extraction, prediction, interpretability, report generation, running in a deployable web-based platform, which is driven by FastAPI. This makes it useable in real time clinically in any hospital, diagnostic centre and in a telemedicine setting. Moreover, the system can be used with other cancers or pulmonary diseases through retraining of the models with appropriate datasets, and this will give it an adaptability that surpasses the capabilities of the immediate prior technologies.

10. **Has the invention been built or tested or implemented? If yes please provide the Efficiency/Efficacy details of the invention**
Yes, both clinical and imaging datasets have been used to build and test the invention. Using machine learning models like the Random Forest and the Logistic Regression, the clinical dataset of 5000 records of patients comprising of 18 medical attributes has been run, with high predictive accuracy including balanced preciseness and recall. Random Forest had better performance and high overall accuracy and robustness on a variety of evaluation metrics.

DenseNet121, ResNet50, and EfficientNet-B0 deep learning architectures were trained and tested on CT scan images classification. Among them, DenseNet121 had the highest accuracy with 99 percent, 99.45 percent precision, and 99.44 percent recall and F1-score where it became able to label benign, malignant, and normal cases with minimum errors.

Explainability methods (SHAP and Grad-CAM) were incorporated to increase trust and clinical usability and allow physicians to see contributing factors and abnormal lung regions. This high efficacy and interpretability combination attest to the reliability of the invention to be used in the real world.

11. **Briefly state when and how you first conceived this idea?**

The idea was first conceived during the initial stages of our major project work in early 2025, while analysing the limitations of existing lung cancer diagnostic systems. We observed that most available solutions relied either on clinical data or on imaging alone, and often functioned as black-box models with little to no interpretability for doctors. Recognizing the urgent need for an accurate, early detection system that also offered transparency, we formulated the concept of a multimodal, explainable AI-based framework. This approach combines clinical attributes with CT scan analysis and integrates explainability tools such as SHAP and Grad-CAM to ensure both predictive efficacy and clinical trust.

12. **Have you sold, offered for sale, publicly used or published anything related to this invention? If yes, please briefly explain the dates and circumstances. List those individuals to whom you have revealed your invention. Were non-discloser documents signed prior to discloser in each case? Please state any deadlines of which you may be aware for filing an application on this invention.**

No

13. **Include any reasons that your invention would not have been obvious to someone of average skill in the art.**

It would not have made an intuitive discovery to a person of average ability in the field as the invention adds a novel mix of multimodal data integration, explainability, and real-time deployment, absent in the existing literature. Although the previous methods have used clinical datasets or CT imaging alone, this invention is a synergistic approach that uses both in the same diagnostic pipeline. In addition, clinical data with SHAP and CT scans with Grad-CAM offer dual-level interpretability, a step further than traditional accuracy-oriented models. The reliability is further improved with the implementation of a decision fusion module that reconcile predictions provided by machine learning and deep learning models to overcome the drawbacks of false positives and negatives in the previous art. Lastly, the practical clinical usefulness as shown in the real-time deployment with a FastAPI-based web platform would not have been evident in the current literature where offline experimentation is mostly emphasized.

14. **Additional comments by inventor (if you want to give more details out of scope of this IDF).**

Outside of the context of this disclosure, the invention can be expanded into a wider medical diagnostics application. Multimodal framework can be extended to predict other cancers (breast or cervical) or respiratory diseases by retraining the models using domain-specific data. The explainability does not only enhance clinical trust but is also used as an educational tool by medical trainees as it is used to point out correlations between symptoms, risk factors and imaging findings. It can be enhanced in the future by combining it with cloud-based health systems, remote-screening mobile health applications, and real-time monitoring dashboards, in the case of large-scale public health programs. Therefore, the invention is scalable and versatile diagnostic support system with broad clinical and societal effects.

15. Drawings/Flowchart/Table