Computer Methods and Programs in Biomedicine 243 (2024) 107879

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

DeepXplainer: An interpretable deep learning based approach for lung cancer detection using explainable artificial intelligence

Niyaz Ahmad Wani, Ravinder Kumar, Jatin Bedi *

Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology, Patiala (PIN: 147004), Punjab, India

ARTICLE INFO

Keywords:
Artificial intelligence
Smart healthcare systems
Lung cancer
Deep learning
Explainable artificial intelligence (XAI)
SHAP

ABSTRACT

Background and Objective: Artificial intelligence (AI) has several uses in the healthcare industry, some of which include healthcare management, medical forecasting, practical making of decisions, and diagnosis. AI technologies have reached human-like performance, but their use is limited since they are still largely viewed as opaque black boxes. This distrust remains the primary factor for their limited real application, particularly in healthcare. As a result, there is a need for interpretable predictors that provide better predictions and also explain their predictions.

Methods: This study introduces *"DeepXplainer"*, a new interpretable hybrid deep learning-based technique for detecting lung cancer and providing explanations of the predictions. This technique is based on a convolutional neural network and XGBoost. XGBoost is used for class label prediction after *"DeepXplainer"* has automatically learned the features of the input using its many convolutional layers. For providing explanations or explainability of the predictions, an explainable artificial intelligence method known as "SHAP" is implemented.

Results: The open-source "Survey Lung Cancer" dataset was processed using this method. On multiple parameters, including accuracy, sensitivity, F1-score, etc., the proposed method outperformed the existing methods. The proposed method obtained an accuracy of 97.43%, a sensitivity of 98.71%, and an F1-score of 98.08. After the model has made predictions with this high degree of accuracy, each prediction is explained by implementing an explainable artificial intelligence method at both the local and global levels.

Conclusions: A deep learning-based classification model for lung cancer is proposed with three primary components: one for feature learning, another for classification, and a third for providing explanations for the predictions made by the proposed hybrid (ConvXGB) model. The proposed *"DeepXplainer"* has been evaluated using a variety of metrics, and the results demonstrate that it outperforms the current benchmarks. Providing explanations for the predictions, the proposed approach may help doctors in detecting and treating lung cancer patients more effectively.

1. Introduction

Healthcare is one of the most important factors in maintaining the health and quality of life for almost all people and communities. It is now a basic right to have access to decent healthcare, which promotes health and prevents a variety of ailments. Healthcare systems play a crucial role in tackling several public health challenges and meeting the needs of vulnerable populations concurrently. Numerous technological and medical breakthroughs continue to modify and improve healthcare systems, making them more cost-effective and enhancing the delivery of treatment. Today's quick progress has resulted in the discovery of several novel medicines, medications, and cures for severe diseases such as cancer. When it comes to global mortality rates, cancer ranks second. Recent estimates from the United States National Cancer Institute (NCI) project that 1,93 million people will be diagnosed with cancer and 609,360 will lose their lives to the disease in 2023 [1]. With an estimated 130,180 deaths in 2023, lung cancer is the main cause of cancer-related mortality and the second most common cancer overall.

Additionally, men have a 1 in 13 chance of acquiring lung cancer in their lifetime, while women have a 1 in 16 chance. Detection or prediction may thus assist both patients and healthcare providers with cost control, treatment intensity, etc. Artificial intelligence (AI) has the

* Corresponding author.
E-mail addresses: nwani60_phd20@thapar.edu, neyaz32@gmail.com (N.A. Wani), ravinder@thapar.edu (R. Kumar), jatin.bedi@thapar.edu (J. Bedi).

https://doi.org/10.1016/j.cmpb.2023.107879
Received 1 June 2023; Received in revised form 17 October 2023; Accepted 20 October 2023
Available online 24 October 2023

ability to revolutionize the detection and treatment of lung cancer by analyzing massive amounts of medical data and providing individualized therapy. Major strides have been made in the detection of lung cancer [2], but more in-depth research is needed to develop approaches that are generally and practically applicable to medical applications and could ultimately lead to more informed decisions about medical professional's efforts and the efforts of the family to improve a cancer patient's health. For these reasons, the challenge of diagnosing lung cancer has attracted the interest of scientists, doctors, and computer engineers.

Nowadays Machine learning (ML) and Deep Learning (DL) technologies are booming in domains, such as Intelligent Transportation Systems [3] [4], and smart healthcare systems [5] [6]. Realistically, machine learning algorithms do a variety of particular jobs by identifying very narrow data associations. Understanding complex input, especially images [7] [8] [9], is challenging for ML algorithms because of the work involved in tasks like feature engineering and selection. In contrast, DL algorithms, which fall under the umbrella of ML, are composed of not just one but several hidden layers between the input and output layers. To run these models, a large amount of data and powerful workstations are required. Due to its many characteristics, including adaptability, high performance and capacity, interdisciplinary application, and other attributes, the scientific community is now concentrating on deep learning [10]. The DL methods outperform their ML-based predecessors, especially when it comes to classifying biological pictures [6]. However, both ML and DL are considered "Black Boxes" since neither explains the predictions being made [11,12]. As a consequence, end users (i.e., medical practitioners) are unable to fully comprehend the prediction process and validate the findings generated by these models. This eventually resulted in less adoption of artificial intelligence models by healthcare professionals, such as physicians, who continue to depend on evidence-based diagnoses to guide treatment predictions. Therefore, explanations and interpretations of model output are essential to instill confidence in and enhance the usage of these systems in diverse sectors of application, particularly when a single erroneous decision might threaten the lives of humans (autonomous driving, medical domain). In addition, when a medical professional explains the therapeutic option to his patient, the explanations of one's decision are typically a prerequisite for establishing a trusting relationship.

Explainable artificial intelligence suggests a change from black box to transparent artificial intelligence to address several methods for generating models with high accuracy that are simple for people to grasp [13]. In 2004, Van Lent et al. invented the phrase and defined it as the "ability of a system/model to characterize the behavior of AI-controlled entities in the simulation games". Explainability is now even mandated by regulatory requirements in sensitive domains such as medicine, which demand traceability, transparency, and interpretability [14]. According to the Defense Advanced Research Projects Agency (DARPA), "XAI attempts to construct more explainable models while maintaining a high level of learning performance (prediction accuracy), enabling humans to grasp, trust, and manage the next generation of AI partners". XAI's application domains include almost every important and ordinary field, such as medicine, transportation, the military, the law, security, education, etc. With the use of AI, several significant challenges in the medical or healthcare industry may be tackled. Recent interest has been drawn to drug creation and testing as a result of the extensive research into automated diagnosis [15]. Although several AI-based systems have been validated in real-world clinical settings for conditions including diabetic retinopathy, histology-based breast cancer metastases, and congenital defects, some of these systems have shown alarmingly large false positive rates [16]. Unfairness, security, confidentiality, and a lack of transparency, informational ness, and trust are some of the issues that have been raised regarding the use of ML in healthcare. Because the choices made or influenced by such systems affect human health, it is essential to comprehend how they are made [17].

As a result, explainability, along with the associated ideas of interpretability and transparency, have emerged as an essential aspect of artificial intelligence in healthcare in recent years. In order to use AI to address problems in medicine outside of the laboratory, and in routine settings, people must do more than simply enhance the efficacy of extant AI methods. Robust AI solutions must be able to handle imprecision, absence and incorrect data, and convey both the result and the process by which it was obtained to a medical professional [18]. Explainability is becoming more crucial to healthcare AI model developers, and the results are encouraging. Consequently, the requirement for XAI in the healthcare industry is of the highest significance, so AI-based solutions for healthcare are not only accurate in terms of prediction but also transparent and explainable. In this study, we focus on employing a deep-learning algorithm to detect lung cancer, and we use XAI methods, notably SHAP, to explain our predictions.

2. Related works

Machine learning and deep learning algorithms apply various transformations to various small and large databases allowing them to be used in a variety of fields of application, such as in healthcare for cancer detection, disease diagnosis, object detection, feature extraction, and image classification, among others. Support Vector Machines (SVMs), Random Forests (RF), and Nearest Neighbors (NN) [19] solve diverse classification and regression issues. Due to their outstanding prediction, performance, and task handling, these methods have been applied to several datasets.

2.1. Cancer prediction

Several detection methods for lung cancer have previously been developed by diverse systems. Nascimento et al. presented SVM-based classification for lung cancer diagnosis and attained an accuracy of 92.78% [20]. To classify nodules, Sheway et al. [21] applied geometric and histogram features. Using the idea of feature fusion, Farag et al. [22] inquired about the standardization of lung cancer classification. Local Binary Pattern (LBP), Gabor filter, and Signed-distance transform shape-based feature descriptors were employed for feature extraction, followed by k-nearest neighbors (kNN) and support vector machines (SVMs). Using the High Order Markov Gibbs Random Field (MGRF) technique and the 3D HOG filter, Shaffie et al. [23] extracted features from the LIDC-IDRI database, fused those features, and then categorized the data using a stacked auto-encoder. Using laboratory data such as CT scans and MRIs, many researchers in the area concentrated on the diagnosis, prognosis, and survival of patients. Macyszyn et al. [24] employ a mix of linear SVM and Kaplan-Meyer for glioblastoma MRI prediction and survival curves, respectively. Five-year survival rates of those with chronic renal illness were predicted using the multi-layer perceptron (MLP) model, according to another study [25]. Sesen et al. [26] presented a recommendation system for therapeutic guidance for patients with lung cancer using the LUCADA dataset. The SEER dataset was used by Aggarwal et al. [27] in their survival prediction calculator for patients with lung cancer over the course of 6 months, 1 year, and 5 years. A combined voting classification was utilized in this model. Patients' survival was predicted with 90% accuracy using data from the SEER database, and C4.5 and Naive-Bayes were compared for their performance [28]. Authors in [29,5] conducted an ensemble vote of 5-decision tree basis classifiers for the best lung cancer prediction.

Many researchers [30] have used various deep neural networks on the problem of lung cancer classification in light of the recent successes of Deep CNNs on diverse image classification benchmarks like ImageNet and MS COCO database. Another Deep Learning model was presented by Kumar et al. [31], who employed a stacked auto-encoder and attained an accuracy of 75.01%. Kuruvilla et al. and Dandil et al. suggested ANNs based on texture features for detection, achieving 93.30% and 90.0% accuracy, respectively [32,33]. The LIDC-IDRI dataset is used for training and testing the CMixNet that was proposed for lung nodule recognition and classification by Nasrullah et al. [34]. For lung cancer classification using the LIDC-IDRI dataset, Tran et al. [35] offer a 2D deep convolutional neural network (DCNN) architecture with 15 layers. They increase accuracy by using a focused loss function during model development. The majority of earlier research has suggested models with considerable results, however, they suffer from the black box issues. These models do not let researchers determine the basis for the prediction. For crucial issues where the cost of misclassification is large, it is necessary to determine why the certain forecast was made. Consequently, it is necessary to explore the strategies that generate substantial results in terms of accuracy, precision, recall, etc., as well as those that are more interpretable.

2.2. *Explainability*

With the advancement of diverse machine learning (ML) approaches used in the field of healthcare, the interpretability of models has gained a great deal of importance in healthcare applications [36], as a potentially incorrect decision could be fatal to a patient's health, making interpreting such results crucial [37]. Consequently, there has been significant growth in the usage of Integrated Machine Learning (IML) models in the healthcare arena for different difficulties, such as illness diagnosis in cardiology, neurology, and other fields. [38,39,37] marks the beginning of the preparatory work for the creation of these Ml models. Existing explainable artificial intelligence (XAI) technologies suited to the healthcare industry are diverse. Ante-hoc and post-hoc are the two most common classifications. Since they are not reliant on the model, post-hoc approaches are the most prevalent. Among the post-hoc strategies of XAI, SHapley Additive exPlanation (SHAP) is the most often used explanation-providing method. SHAP has been utilized in a variety of healthcare contexts by various research.

2.2.1. *SHapley Additive exPlanation (SHAP)*

Based on the game theory, Lundberg et al. [40] proposed the approach, in which each player's (feature's) contribution to the goal is quantified. Its core premise is that each feature contributes to a model's ability to forecast a given value. Based on each feature's unique contribution to the overall pay-out, a value called Shapley value is allocated to it. The mathematical modeling of the SHAP method is shown in Equation (1).

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j \tag{1}$$

Where (g) is the explanation model, $z' \in \{0,1\}^M$ is the feature vector that has been simplified, M is the vector's maximum coalition size, and $\phi_j \in R$ is the weight assigned to each feature (Shap value) for feature j. By evaluating the relative importance of several biomarkers or clinical features (the "players") in achieving a desired disease outcome (the "reward"), SHAP may find use in healthcare settings. More than 75% of SHAP-based studies have presented machine learning models, with XGBoost (17) studies being the most common and Random forest studies coming in second (RF) (12). Fig. 1 illustrates that SHAP may explain a variety of healthcare applications, including surgical complications, ICU patient states, hospital readmissions, etc. SECG (1D), photoplethysmography (PPG), and other forms of high-dimensional data may also be interpreted with the help of SHAP. Different researchers have used SHAP as an explanation for their healthcare model predictions. SHAP was utilized by [41] to explain the predictions of their suggested technique for the requirement for critical care for infants with pneumonia. Chen et al. [42] employed SHAP to explain the prediction in their study on predicting unfavorable surgical occurrences. In [43], researchers implemented SHAP for analyzing or understanding cancer categorization findings. Explanation of the COVID-19 patient outcomes, Lu et al. [44] and Alves et al. [45] both used SHAP to analyze the findings of the created approach. For interpretability purposes, Dissanayake's study [46] used the same approach for the identification of cardiac anomalies. The authors of studies [47–49] applied SHAP for evaluating the ophthalmic diagnosis, stroke outcome, and lung cancer forecasts. The authors illustrate subnetwork detection based on multimodal node features using a novel interpretable Greedy Decision Forest (GDF) [50]. This will be essential for retaining experts and gaining their trust in these algorithms. The proposed explainable method can aid in identifying disease-causing network modules from multi-omics data to gain a deeper understanding of complex diseases such as cancer.

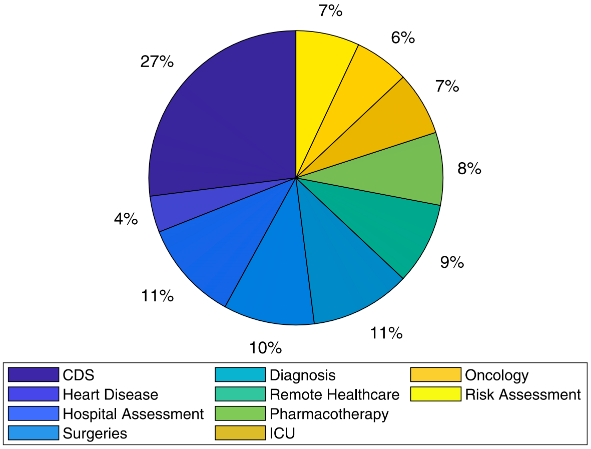


Fig. 1. Diagram of SHAP-studied healthcare applications.

2.3. *Motivation*

The main motivation of this paper is as follows: The recent advances in AI are generating strong outcomes in a variety of hard tasks, however, these activities are often black boxes since the model does not disclose how it arrives at a certain conclusion. When it comes to real-world applications that are extremely crucial, such as healthcare, where a single incorrect choice might be deadly, a lack of transparency or interpretability is in no way acceptable. Explainable AI attempts to tackle these issues by offering human-comprehensible explanations for these black box models. This topic focuses on strategies and approaches that aid in comprehending and analyzing the processes of a certain model. Consequently, in the context of medicine, XAI must take into account the possibility that several facts may contribute to a meaningful conclusion. In order to enhance patient care, medical practitioners need to understand the reasoning behind a machine's choice. Consequently, something similar to explainable medicine is critically needed in the medical industry for a range of fields, such as research and clinical decision-making.

2.4. *Contributions of this article*

The following are the article's major contributions:

1. The development of an efficient hybrid model (ConvXGB) based on the combination of CNN and XGBoost for the detection of lung cancer, results in enhanced prediction outcomes.
2. In the proposed hybrid method, a well-known explainable AI approach is used to provide the medical practitioner with explanations or interpretations to help them make better judgments.
3. Using the standard dataset "Survey Lung Cancer", the suggested hybrid strategy and XAI approaches are applied and evaluated.

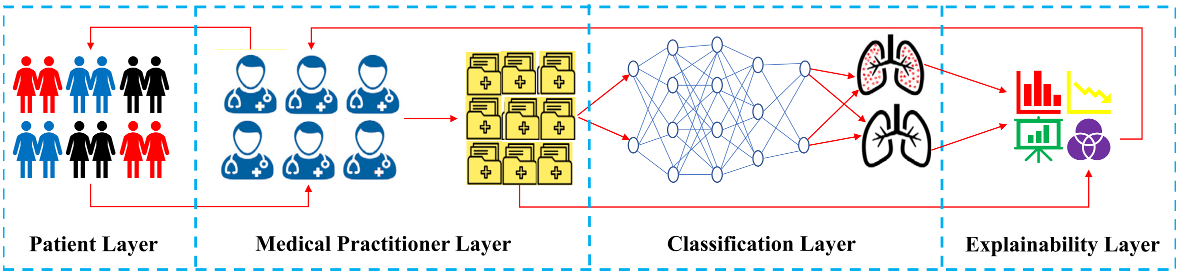


Fig. 2. System model.

2.5. Organisation

The rest of the paper is organized as follows. The whole model of the system is described in Section 3. The notations and objective functions are defined in Section 4. Patient predictions are predicted and explained in Section 5 using the proposed *DeepXplainer*. The results of the proposed system are shown in Section 6. Finally, Section 7 concludes the article.

3. System model

There are four distinct levels to the proposed system model shown in Fig. 2. These layers are discussed below.

3.1. Patient layer

The patient layer of the system model is made up of an 'n' number of patients who are experiencing a variety of symptoms, such as a cough, chest discomfort, shortness of breath, yellow fingers, and difficulty swallowing, among other things. These 'n' number of patients will make an appointment to see the physician to get a diagnosis.

3.2. Medical practitioner layer

At this level, 'n' doctors are supported by several nurses, technicians, computers, and other support personnel. When a patient has an appointment with a physician, he or she will be examined, and the information acquired during the examination, including vital signs and symptoms, will be put into a database for future processing.

3.3. Classification layer

At this layer, the patient's acquired data will be classified using a model that has been pre-trained. After the data is inputted into the model, it takes some time for the model to generate the result. If the patient has been diagnosed with cancer, the patient's data and model results are passed to the explainability layer. If the patient is not diagnosed with cancer by the model, the patient's data and results are nevertheless passed to the explainability layer so that explanations may be generated.

3.4. Explainability layer

Following the classification layer, which determines whether or not a patient has been diagnosed with cancer, the model output, vitals, and symptoms are transmitted to the explainability layer for explanations. This layer gives reasons for both outputs, namely whether or not a patient has cancer. Finally, these explanations are passed to the medical practitioner layer, where the corresponding doctor understands those explanations for the respective outputs and explains to the patient why he/she has been diagnosed with cancer or why not.

4. Problem formulation

This part formulates and presents as objective functions the problems that have been addressed in this study.

4.1. Rationale behind the proposed work

Two assumptions underpin the method that is being proposed here.

1. The accuracy requirements of different problems may be too high for a single model to handle.
2. Any model will have advantages and disadvantages if it is thoroughly and objectively analyzed; identifying these aspects will help better models be developed and the drawbacks avoided. We thus seek these advantages.

4.2. Notations

Let X_p contain symptoms and vitals given by the patient to the doctor. Let X_m contain X_p and the doctor's observations. Let X_d contain historical data of many patients. Let $(X_d)_0$ be the header of data or names of attributes/features in the dataset. Let $(X_d)_{0,j}$ be the name of the particular j^{th} attribute/feature of the dataset. $(X_d)_{0,1:n} = \left\{ Gender,\ Age,\ Smoking, Yellow\ fingers,\ Anxiety, Peer\ Pressure,\ Chronic\ disease,\ Fatigue,\ Allergy,\ Wheezing,\ Alcohol\ Consuming,\ Coughing,\ Shortness\ of\ Breath,\ Swallowing\ Difficulty,\ Chest\ Pain,\ Lung\ Cancer \right\}$.

The size of X_d is m, n. Where 'm' denotes the number of rows and 'n' denotes the number of columns. $(X_d)_{i,j}$ be the value of j^{th} attribute/feature present at i^{th} row. The doctor will add X_m data to the X_d dataset as per equation Equation (2).

$$X_d = X_d \cup X_m \tag{2}$$

Let D(·) and R(·) are the two functions that provide the domain and range of the input attribute/feature. Thus, the domain of each attribute is fixed in the 'X_d' dataset. For example, let $(X_d)_{0,k}$ ='whether patient smokes' then

$$D\left(\left(X_d\right)_{0,k}\right) = \left\{Yes, No\right\}$$

Let S(·) be the function that represents the number of rows that are present in the input data. Thus,

$$K = S\left(X_d\right)$$

$\Longrightarrow$ X_d dataset has 'K' rows.

Let the 'X_d' dataset be a breakdown into training and testing datasets denoted by '$X_d \cdot$ train' and '$X_d \cdot$ test' respectively. Let $\left(X_d\right)_{0,n-1}$ be the input attributes/features and $\left(X_d\right)_{0,n}$ be the output attribute. Let f_c define the training of the classification model denoted by 'ConvXGB', then,

$$ConvXGB = f_c\left(X_d \cdot train\right)$$

Let Y_c^j be the output or predicted data given by the trained 'ConvXGB' model on any new or testing data present at j^{th} row.

$$Y_c^j = ConvXGB\left(\left(X_d \cdot test\right)_{j,1:n-1}\right)$$

Moreover, the range of the ConvXGB model will be the domain of the output data feature 'Y_c'

$$D\left(Y_c\right) = R\left(ConvXGB\right)$$

Let $\hat{Y}_c^j$ represent the actual value of feature predicted as Y_c^j. Then,

$$\hat{Y}_c^j = \left(X_d \cdot test\right)_{j,n}$$

Let p_i denote the probability of i^{th} instance that it belongs to one particular class say '1'. This probability is provided by the trained 'ConvXGB' classification model. Let $X_d\cdot$ new are the vitals of the new patient observed by the medical practitioner along with the class label given by the ConvXGB model. Let $\mathrm{I}(\cdot)$ be a function which denotes the importance of the j^{th} individual feature $(X_d)_{i,j}$, where $j \in [1, n-1]$ in its class $(X_d)_{i,n}$ identified by the 'ConvXGB' model. Let 'EM' denotes the explainability model used for depicting the importance of an individual attribute or feature in its predicted class label. The negative mean of the log of the corrected predicted probability is the Binary Cross Entropy. The actual class output, which might be either 0 or 1, is compared to each of the projected probabilities. Then, it determines a score that discounts the odds in accordance with how far they are from the predicted value. That indicates how near or distant the value is from the true one. Then, let it is denoted by E_c.

$$E_c = \frac{1}{\mathrm{S}(X_d \cdot \mathrm{test})} \sum_{j=1}^{S(X_d \cdot \mathrm{test})} -\left(\hat{Y}_c^j * \log(p_j) + (1-\hat{Y}_c^j) * \log(1-p_j)\right)$$

4.3. *List of objectives*

In light of the foregoing statements, the following are the primary objectives of the current study:

$\mathcal{O}1)$ *Minimize*

$$\frac{1}{\mathrm{S}\left(X_d \cdot \mathrm{test}\right)} \sum_{j=1}^{S(X_d \cdot \mathrm{test})} -\left(\hat{Y}_c^j * \log\left(p_j\right) + \left(1-\hat{Y}_c^j\right) * \log\left(1-p_j\right)\right)$$

$$\mathcal{O}2)\ \forall j \in [1, n-1]\ \mathrm{I}\left(\left(x_d \cdot\ \mathrm{new}\right)_{0,j}\right) = \mathrm{EM}\left(\mathrm{ConvXGB}\left(\left(X_d \cdot \mathrm{new}\right)_{1,1:n-1}\right), \left(X_d \cdot \mathrm{new}\right)_{0,j}\right)$$

4.3.1. *Constraints*

Solution of $\mathcal{O}1$ and $\mathcal{O}2$ is constrained by $C1$, $C2$ and $C3$

$$C1\ :\ Y_c^j \in \left\{0, 1\right\}$$

$$C2\ :\ p_i \in [0, 1]$$

$$C3\ :\ \mathrm{X}_d \cdot \mathrm{train} \cup \mathrm{X}_d \cdot \mathrm{test} = \mathrm{X}_d$$

The constraint $C1$ defines that the predicted data must belong to either of two classes. $C2$ represents the probability of any instance 'i' belonging to one particular class should belong to the interval [0,1]. $C3$ represents the meaningful splitting of whole patient data.

5. ***DeepXplainer*: the proposed method**

This section describes the *DeepXplainer* implemented for detecting lung cancer and provides explanations of the predictions made by the detection (ConvXGB) model. The architecture of the proposed method consists of three major components, each of which is made of multiple distinct layers as shown in Fig. 3. The first deals with the loading and preparation of data, the second with learning and classifying features, and the third with explaining things via the use of an XAI model. Each part of the architecture is built up from many layers, each of which performs a specific task.

5.1. *Lung cancer detection*

The proposed procedure begins with data input, followed by data preprocessing. X_d represents the imported data with a capacity of m, n, where 'm' represents the number of rows and 'n' represents the number of columns. Once the X_d has been effectively imported, various preprocessing steps are carried out. By employing label encoding, existing categorical data is supplanted with numerical values. Two columns in X_d are categorical, with the target column containing the values 'YES' and 'NO' and the gender column containing the values 'M' and 'F'. After label encoding, 'YES' becomes '1' and 'NO' becomes '0'. Similarly, 'M' becomes '0' and 'F' becomes '1' (algorithm 1: *lines 1-3*). Following the application of label encoding, categorical data is transformed into a format that can be utilized by a variety of machine learning algorithms. After label encoding, another preparatory phase consists of removing any duplicate values from X_d (algorithm 1: *lines 4-6*).

After label encoding and removal of duplicate values, X_d is divided into two sets: one containing the target value, i.e., the final class label, and the other containing the remaining columns. For applying the proposed CNN, the data is partitioned into training and testing sets using the 80-20 strategy (algorithm 1: *lines 7-10*), followed by data transformation that includes scaling with MinMaxScaler, LabelBinarizer, etc. The processed data is transmitted to the second component of the proposed model to apply CNN to the data. After data transformation, X_d is divided into x-train, y-train, x-test, and y-test, which are represented by $X2_d^{tr}$, $Y2_d^{tr}$, $X2_d^{ts}$, and $Y2_d^{ts}$, respectively (algorithm 1: *lines 11-15*).

Now, $X2_d^{tr}$ and $Y2_d^{tr}$ are fed into the proposed CNN model, which consists of multiple convolutional and max-pooling layers. Table 1 outlines the proposed CNN's architectural layout. The proposed CNN is only used for feature learning and not for data classification. The proposed CNN's classification layer is substituted by the well-known machine learning classifier XGBoost. The proposed CNN has four total layers, two of which are convolutional and the other two are max-pooling. However, adding too many layers might have the opposite effect of what was intended, thus it's important to balance potential improvements in accuracy against the resources necessary to implement them on a given device. To reduce overfitting, a 0.20 dropout is applied after the second max-pooling layer.

After the dropout layer, the data is flattened before transmission to the first dense layer. There are three dense layers in total. The first two dense layers are made up of 128 and 50 filters, respectively. Once the architecture of the proposed CNN has been built, the model is trained using the training data, and validation is performed using validation data with batch size = 160 and epochs = 120 (algorithm 1: *lines 16-17*). After training, and before transmitting the data to the classification layer, XGBoost, the reshape layer is used to reformat the data. This layer executes a few elementary operations to transform the output of the convolutional layer into the vector required by the XGBoost layer (algorithm 1: *lines 18-22*). As described in 4.1, there are several reasons to use XGBoost as the classifier layer. In addition, when four machine learning algorithms are applied to the same dataset, XGBoost provides the highest accuracy, as shown in Table 2 and Fig. 8 in the result section. The classification layer will forecast the output class. The accuracy of these classes, indicating the learning performance of the model, is then evaluated. Certainly, our proposed method plays a crucial role as a determining factor and initial predictor in directing subsequent examinations of cancer patients. Subsequently, these patients are subjected to a series of essential, diverse clinical procedures in order to conduct an exhaustive evaluation. It is essential to emphasize that the primary focus of our work is determining whether or not a patient has cancer.

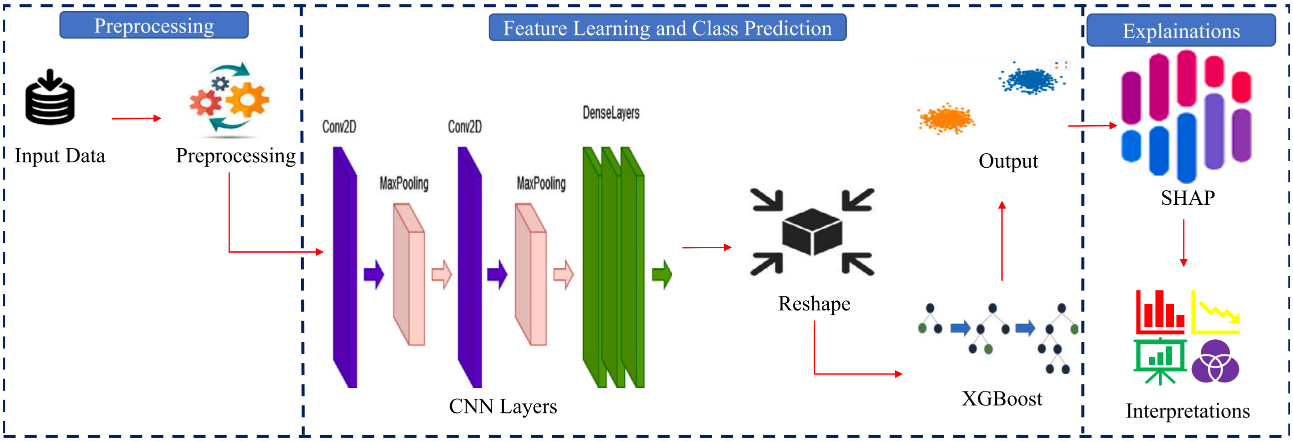


Fig. 3. Architecture of *DeepXplainer* model.

Algorithm 1 : *DeepXplainer.*

```
1: $X_d = \text{Load}\left(X_d\right)$
2: $X_{d\,:,1} = X_{d\,:,1} \cdot \text{replace}\left(\left\{\text{‘M’}: 0, \text{‘F’}: 1\right\}\right)$
3: $X_{d\,:,16} = X_{d\,:,16} \cdot \text{replace}\left(\left\{\text{‘YES’}: 0, \text{‘NO’}: 1\right\}\right)$
4: $X_d = X_d \cdot \text{drop\_duplicates}$
5: $X_d^0 = X_{d\,:,:} \mid \left(:, n == 0\right)$
6: $X_d^1 = \text{Concat}\left(X_d, X_d^0\right) \cdot \text{drop\_duplicates}\left(\text{all}\right)$
7: $X_d^{tr} = \text{Concat}\left(\left[X_d^1, X_d^0\right]\right)$
8: $X_d^{ts} = \text{Concat}\left(\left[X_d^0, X_d^1\right]\right) \cdot \text{drop\_duplicates}\left(\text{all}\right)$
9: $X1_d^{tr} = \left(X_d^{tr}\right)_{:,1:n-1}, Y1_d^{tr} = \left(X_d^{tr}\right)_{:,n}$
10: $X1_d^{ts} = \left(X_d^{ts}\right)_{:,1:n-1}, Y1_d^{ts} = \left(X_d^{ts}\right)_{:,n}$
11: $X1_d^{tr} = \text{Scale}\left(X1_d^{tr}, \left(0,1\right)\right), X1_d^{ts} = \text{Scale}\left(X1_d^{ts}, \left(0,1\right)\right)$
12: $X2_d^{tr} = \text{reshape}\left(X1_d^{tr} \cdot \text{shape}\left[0\right], 5, 3,1\right)$
13: $X2_d^{ts} = \text{reshape}\left(X1_d^{ts} \cdot \text{shape}\left[0\right], 5, 3,1\right)$
14: $Y2_d^{tr} = \text{LabelBinarizer}\left(Y1_d^{tr}\right)$
15: $Y2_d^{ts} = \text{LabelBinarizer}\left(Y1_d^{ts}\right)$
16: $\text{M1} = \text{Build\_Model}$
17: $\text{M1} \cdot \text{fit}\left(X2_d^{tr}, Y2_d^{tr}, X2_d^{ts}, Y2_d^{ts}, \text{epochs, batch\_size}\right)$
18: $\text{M2} = \text{Model}\left(\text{inp} = \text{M1} \cdot \text{input, output} = \text{M1} \cdot \text{get\_layer}\left(\text{cnn\_layers}\left[-1\right]\right.\right.$
    $\left.\left.\cdot\text{output}\right)\right)$
19: $X3_d^{tr} = \text{M2} \cdot \text{prediction}\left(X2_d^{tr}\right)$
20: $X3_d^{ts} = \text{M2} \cdot \text{prediction}\left(X2_d^{ts}\right)$
21: $\text{ConvXGB} = \text{XGBClassifier}\left(\right)$
22: $\text{ConvXGB} \cdot \text{fit}\left(X3_d^{tr}, Y2_d^{tr}\right)$
23: $Y_c^m = \text{ConvXGB} \cdot \text{predict}\left(\left(X3_d^{tr}\right)_{m,:}\right)$ {Prediction on Patient Data}
24: $\text{exp\_ConvXGB} = \text{shap.DeepExplainer}\left(\text{M1}, X2_d^{tr}\right)$
25: $\text{vshap\_ConvXGB} = \text{exp\_ConvXGB.shap\_values}\left(X2_d^{ts}\right)$
26: $\text{shap.forceplot}\left(\text{exp\_ConvXGB.expected\_value}\left[0\right],\right.$
    $\left.\text{vshap\_ConvXGB}\left[0\right], \left(X2_d^{ts}\right)_{m,:}\right)$
```

Following the prediction, the focus of our research is on elucidating the significance of distinct characteristics responsible for diagnosing the patient within that particular category.

5.2. Explainability

After the class prediction component has made its predictions, the results or predictions are forwarded to the model's XAI layer, where the XAI model is implemented to provide explanations. In the proposed scheme, SHAP is implemented, which receives input from XGBoost and provides numerous explanations. SHAP offers both local and global explanations that can be interpreted, as depicted in various figures given in subsubsection 6.5.2. These explanations are not interpretable to the patient, but they are to the medical professional. Once the medical practitioner interprets these explanations, the patient can be provided with a variety of reasons for having a specific disease, and the practitioner can then make a variety of recommendations (algorithm 1: *lines 24-26*). After the detection part of our proposed method is completed, we employed the SHAP (SHapley Additive exPlanations) technique to provide various kinds of explanations for the model's predictions. SHAP is a powerful and widely used method in explainable artificial intelligence (XAI), designed to elucidate the contributions of individual features in influencing model outputs. By leveraging SHAP, we aimed to offer insightful and interpretable explanations to enhance the transparency of our model's decision-making process. To utilize SHAP, we first obtained the predictions from our model for each patient in the dataset. Then, we computed the SHAP values for every feature used in the model, quantifying the impact of each feature on the prediction. These SHAP values represent how much each feature's presence or absence influenced the model's output, providing a measure of feature importance. Next, we carefully configured the SHAP values into various forms of explanations for the detected cases. For instance, we generated individual patient-level explanations, presenting a breakdown of the most influential features in each prediction. These patient-specific explanations allowed doctors to discern the critical factors contributing to a particular patient's diagnosis, thereby gaining deeper insights into the model's decision. Furthermore, we also aggregated SHAP values across the entire dataset to identify common patterns and trends. By doing so, we extracted global-level explanations, unveiling the overarching features consistently associated with positive or negative predictions for lung cancer. These global explanations served to highlight general patterns in the data and provided a broader understanding of the model's behavior. The incorporation of SHAP explanations after the detection phase of our proposed method proved invaluable for offering a comprehensive and interpretable assessment of the model's predictions. By providing both patient-specific and global-level insights, the SHAP technique empowered doctors and researchers to gain trust in the model's decisions and fostered a deeper understanding of the complex relationships between features and disease outcomes. This transparency and interpretability are critical for fostering confidence in the model's real-world application and its potential impact on improving lung cancer diagnosis and patient care.

6. Performance evaluation

This section presents a performance evaluation of the proposed *DeepXplainer* approach.

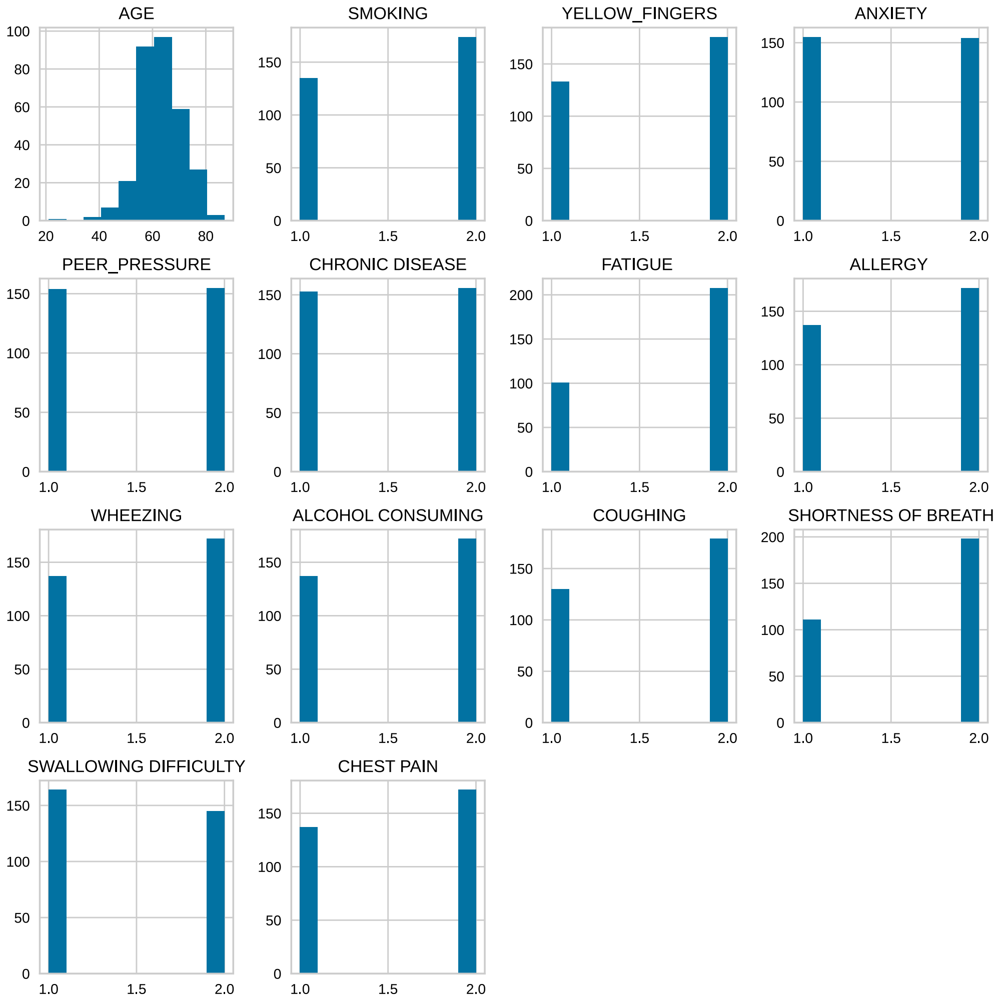


Fig. 4. Graphical view of dataset attributes.

6.1. Environment setup for experiments

Tests are performed using a machine with a 2.40 GHz Intel(R) Core(TM) i5-6300U CPU, 16 GB of RAM, a 256 GB SSD, an NVIDIA GeForce 940MX, and a 64-bit operating system. The experimental code for the study is developed using TensorFlow and Keras respectively.

6.2. Data-set description

The dataset utilized in this research, “Survey Lung Cancer (SLC),” was collected from an open source platform [51]. The dataset comprises a total of 16 features, including Smoking, Anxiety, Shortness of Breath, Chest Pain, and Coughing, among others. The dataset is imbalanced, necessitating the application of data-sampling techniques, such as SMOTE. Fig. 4 depicts a visual representation of the features. Before this dataset can be provided to the proposed model, a number of its features must undergo different preprocessing steps.

6.3. Experimental parameters

In the proposed model, two convolutional layers (Conv2D) in addition to three dense layers are used which are presented in Table 1

Table 1
Summary of CNN architecture.

Layer type	Kernel Size	Number of Filters
First Conv2D	3 × 3	30
MaxPooling 2D	1 × 1	-
Second Conv2D	1 × 1	15
MaxPooling 2D	1 × 1	-
Dense	-	128
Dense	-	50
Dense	-	1

The suggested model contains a first convolutional layer with 30 filters and a second convolutional layer with 15 filters. Max-pooling has been implemented after each convolutional layer in the proposed model. In a max-pooling setup, the pool size is always (1,1). Kernel size (3,3) and “Relu” as an activation function make up the first convolutional layer. The second convolutional layer uses a (1,1) sized kernel with the “Relu” activation function. After the second max-pooling layer in this model, a 0.20 dropout has been introduced. With the addition of dropout, the data are flattened before transmission to the first dense layer. The first and second dense layers have 128 and 50 filters with “Relu” as the activation function, respectively, but the final dense layer has just one filter with the “Sigmoid” activation function. In model compilation, binary-cross entropy is used as a loss function and optimizer as “Adam”.

6.4. State-of-the-art-techniques

Results from the proposed approach were compared to those from other research in terms of precision, sensitivity, F1-score, and other

Table 2
Comparison with state-of-the-art-techniques.

Model	Dataset	Number of Samples	Accuracy%	Sensitivity%	F1 Score
Nascimento et al. [20]	LIDC	73	92.78	85.64	-
Kuruvilla et al. [32]	LIDC	110	93.30	91.40	-
Dandil et al. [33]	Private	128	90.63	92.30	-
Gupta et al. [52]	Private	120	90.00	86.66	-
Mahmut Dirik [53]	SLC	309	91.00	-	-
Alpre et al. [54]	SLC	309	96.08	-	-
Naseer et al. [55]	SLC	309	96.67	-	-
XGBoost	SLC	309	92.36	95.93	95.66
Naive Bayes	SLC	309	90.31	95.84	94.53
Random Forest	SLC	309	91.90	97.67	95.46
CATBoost	SLC	309	88.65	98.60	93.82
Proposed Method (ConvXGB)	SLC	309	**97.43**	**98.71**	**98.08**

metrics for analysis. Table 2 compared the proposed scheme to other schemes based on a variety of relevant criteria.

1. Gupta et al. [52]: This method detects lung cancer by extracting features from CT scan images of lung cancer using the Curvelet transformation.
2. Dandil et al. [33]: In this study, using computed tomography (CT) images, a Computer Aided Diagnosis (CAD) system for lung cancer diagnosis and differentiation of benign from malignant tumors has been developed.
3. Kuruvilla et al. [32]: In this study, a computer-assisted segmentation and classification model employing morphological operations for segmentation and a neural network for classification is proposed.
4. Nascimento et al. [20]: In this study, the Shannon and Simpson diversity indices serve as texture descriptors for lung nodules in CT images, classifying them as benign or malignant. A dataset containing 73 nodules, of which 47 are benign and 27 are malignant, is classified using SVM.
5. Mahmut Dirik [53]: The study's overarching goals are to improve disease detection through symptom analysis, to speed up and reduce the cost of case evaluation, and to produce results in novel situations that are on par with or superior to those produced by human experts, all through the use of data mining and other algorithmic techniques.
6. Alpre et al. [54]: Several machine learning methods are used in this article to make predictions about lung cancer. Based on experimental data, random forest is superior, with an accuracy of 96.08% being achieved.
7. Naseer et al. [55]: The authors of this study used a neural network in the computer to help determine whether a human body movie had lung cancer or not. Model verification found a 96.67 percent accuracy rate. The results of this research demonstrate the neural network's diagnostic efficacy for lung cancer, suggesting its potential application by medical professionals.

6.5. Results and discussions

This section presents the prediction and explainability results of the proposed technique.

6.5.1. Lung cancer detection

In this subsection, we present the forecast findings of the proposed model (ConvXGB). The trial findings are shown in Fig. 5, Fig. 6, Fig. 7, Fig. 8 and tabulated in Table 2. As the model proposed in this study is a binary classifier, the various evaluation metrics used for ConvXGB are graphically represented in Fig. 7 for both classes, i.e., ‘0’ and ‘1’.

Fig. 5 illustrates the accuracy of the proposed model. It is evident from the graph that after 120 epochs, both training and validation accuracy is static. *K*-fold cross-validation was utilized during experiments with $K = 10$. Fig. 6 depicts the training and validation loss in terms of epochs. The proposed method obtained 97.53% accuracy, 98.71% sensitivity, and 98.08 F1-score, outperforming numerous state-of-the-

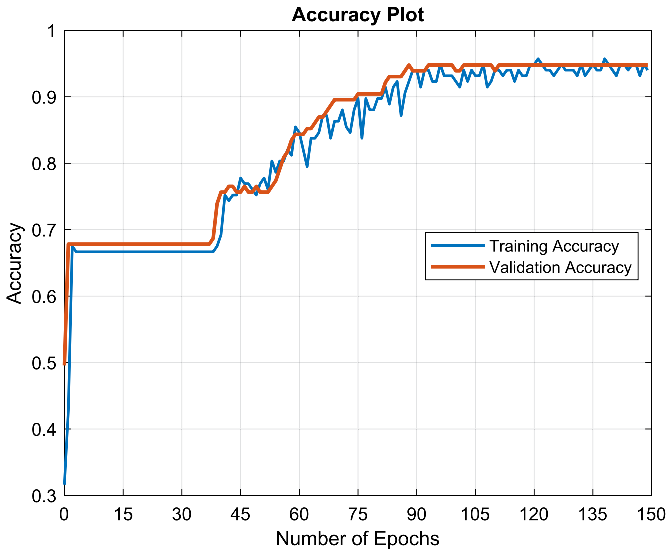


Fig. 5. Plot of training and validation accuracy obtained with respect to epochs.

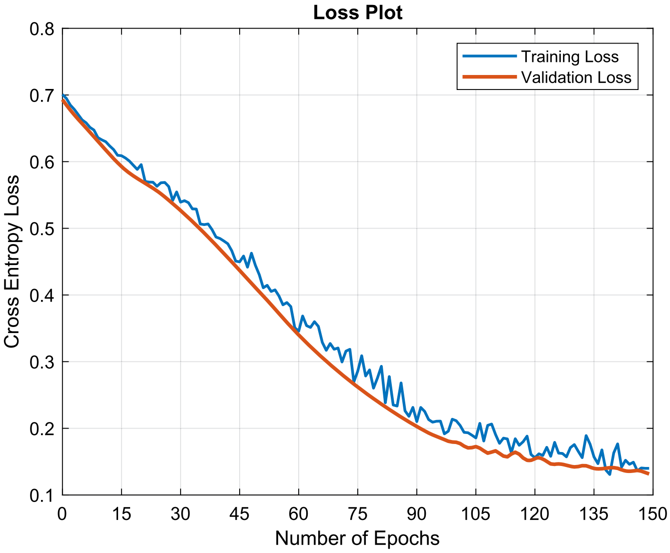


Fig. 6. Plot of training and validation loss obtained with respect to epochs.

Table 3
Computational time of various algorithms.

Algorithm	Time in Secs
Naive Bayes	174
XGBoost	207
Random Forest	389
CatBoost	315
Proposed Method	**110**

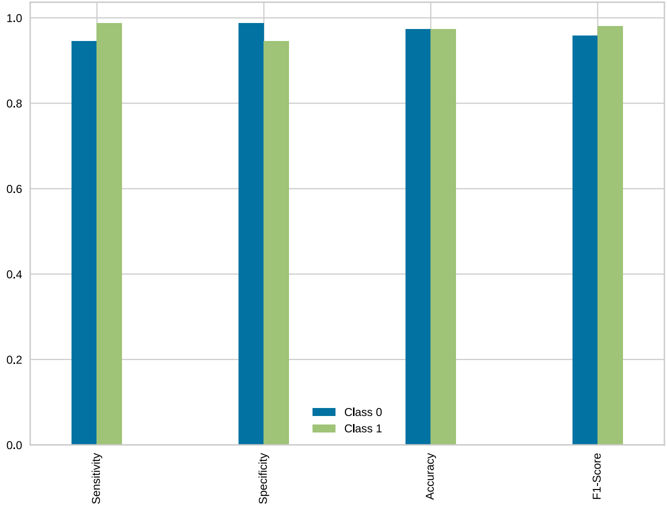


Fig. 7. Graphical representation of various parameters of ConvXGB for target classes.

art techniques. Table 2 displays the results of a comparison between the suggested method and many other state-of-the-art methods. The suggested method was evaluated using a range of metrics, including sample size, precision, and sensitivity, as well as the F1-score. In addition to Naive Bayes, XGBoost, Random Forest, and CATBoost, several machine learning algorithms were tested on the dataset used in this study. The proposed (ConvXGB) method also outperformed these in terms of various evaluation parameters such as accuracy, F1-score, etc. Fig. 8 depicts a comparison between the proposed (ConvXGB) method and the machine learning algorithms utilized throughout the research. The Computational Table Table 3 functions as an all-encompassing database containing the critical findings of our research as well as the numerical data obtained from our rigorous experiments. The Table 3 functions as an indispensable visual assistance, enabling our readers to efficiently comprehend the essence of our discoveries and conduct significant comparisons among various outcomes. We present a comprehensive analysis of the time distribution throughout different stages of the algorithm's implementation in Table 3. The present exhaustive examination includes the phases of training and assessment, in addition to the subsequent implementation of the SHAP (Shapley Additive Explanations) method of explainability. The aim of this endeavor is to produce insightful explanations for the forecasts produced by our model. It is noteworthy, for instance, that the Naive Bayes algorithm requires a total of 174 seconds to complete its execution, which includes training, testing, and explanation generation. When contrasting the execution times of XGBoost, Random Forest, CatBoost, and our proposed method, we observe that they are, respectively, 207, 389, 315, and 110 seconds. The effectiveness of our suggested methodology is clearly demonstrated in this table when combined with the "SHAP" explainable artificial intelligence technique. Significantly, our results demonstrate a considerable decrease in execution time when compared to alternative algorithms such as XGBoost, Naive Bayes, and Random Forest. This time-saving advantage is especially significant in the field of medical diagnostics, where prompt decision-making is of the utmost importance. The methodology we employ is particularly designed to accelerate the diagnostic process, consequently decreasing the duration that clinicians need to acquire critical insights into the conditions of patients. The expedited identification of medical conditions has the capacity to enable more prompt and accurate medical interventions, which could have a profound effect on patient outcomes and care.

6.5.2. *Explainability results*

The explainability results are categorized into two categories, i.e., local and global which are explained below.

6.5.2.1. Local explanations Local interpretability emphasizes the clarification of a single instance as opposed to the representation of all data. Fig. 9a, Fig. 9b, Fig. 9c, and Fig. 9d represent the individual force plots (local explanations) of a few patients. Each force plot yields its own set of SHAP values, from which it is evident why a case received a particular prediction, i.e., whether it is being classified as class 1 (having lung cancer) or class 0 (without lung cancer). Fig. 9b and Fig. 9d represent the two extreme cases, while Fig. 9a and Fig. 9c demonstrate a reasonable equilibrium between the two sets of characteristics that ultimately contribute to the prediction. In Fig. 9, red attributes push the prediction toward class 1 while blue attributes push it toward class 0. From Fig. 9, it is evident that the base value obtained by the proposed model is 0.6641, indicating that any instance whose predicted value is greater than 0.6641 will be classified in class 1 and values below or equal to 0.6641 will be classified in class 0. Fig. 9 force plots also indicate which patient class-predicting characteristic is most prominent or significant. As depicted in Fig. 9d, Peer Pressure, Allergy, Yellow fingers, Anxiety, and Chronic disease, among others are significant for classifying the patient into class 1, as they are present in the patient, i.e., their value is greater than the normal range, while as in Fig. 9a, Fig. 9b and Fig. 9c it is evident that the patients selected do not have lung cancer, as the majority of the key characteristics that contribute to class 1 are absent.

Similarly, Fig. 10 is only relevant for a limited number of patients. The illustrations show how the evidence for each characteristic affected the model's prediction. These charts are meant to graphically demonstrate how the SHAP values (evidence) of each feature change the model's prediction from our initial expectation under the background data dispersion to our final expectation given the evidence of all features. In the waterfall plot, red plots represent values contributing to a positive diagnosis, while blue plots represent values contributing to a negative diagnosis. In Fig. 10, features are ordered according to the magnitude of their SHAP values, with the smallest magnitude features concentrated at the bottom of the plot. Fig. 10 depicts the anticipated model output over the background dataset as the base/reference value, and each row shows how the positive (red) or negative (blue) contribution of each feature transfers the value from the expected model output to the model output for this prediction.

6.5.2.2. Global explanations Global interpretability defines the entirety of the data set. SHAP can facilitate the global and local interpretation of ML models. Fig. 11, Fig. 12 represent the various types of global explanations provided by the proposed model. The influence of characteristics on the prediction is shown graphically in Fig. 11, which also serves as the global explanation. Unlike the bottom characteristics, such as age and gender, the top features, such as chronic illness, exhaustion, peer pressure, and alcohol intake, have a larger impact on the model's predictions. The colors in Fig. 11 accurately depict the value of the feature, with blue representing low and red representing high. Blue dots indicate a negative SHAP value, while red dots indicate a positive SHAP value. Fig. 11 also illustrates the significance of a feature in terms of horizontal expansion. The greater the spread along the horizontal axis, the more significant the feature is for the model's overall

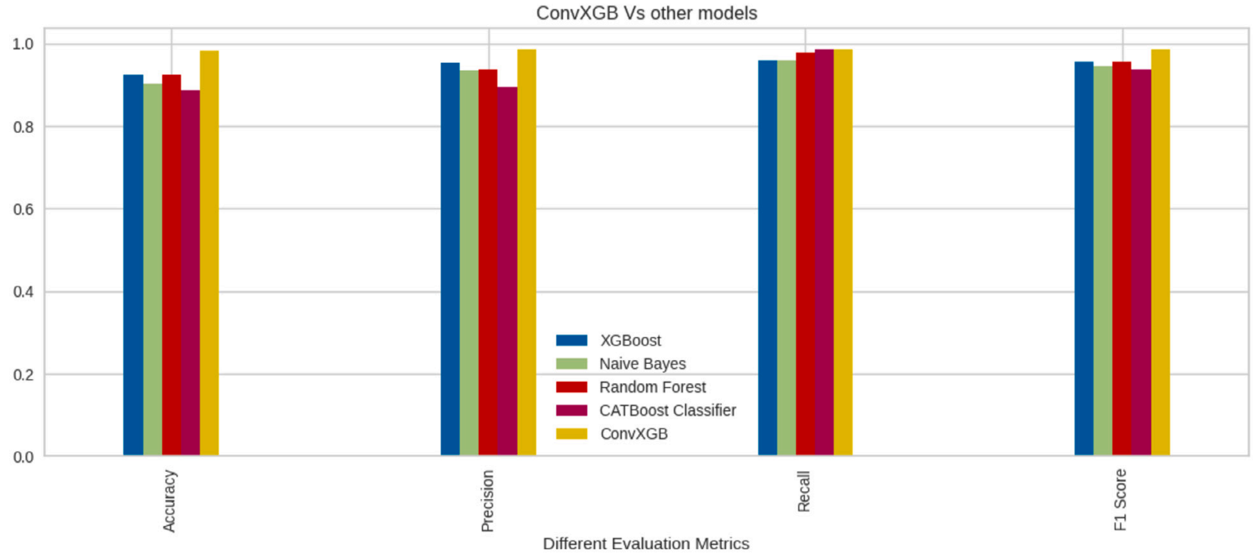
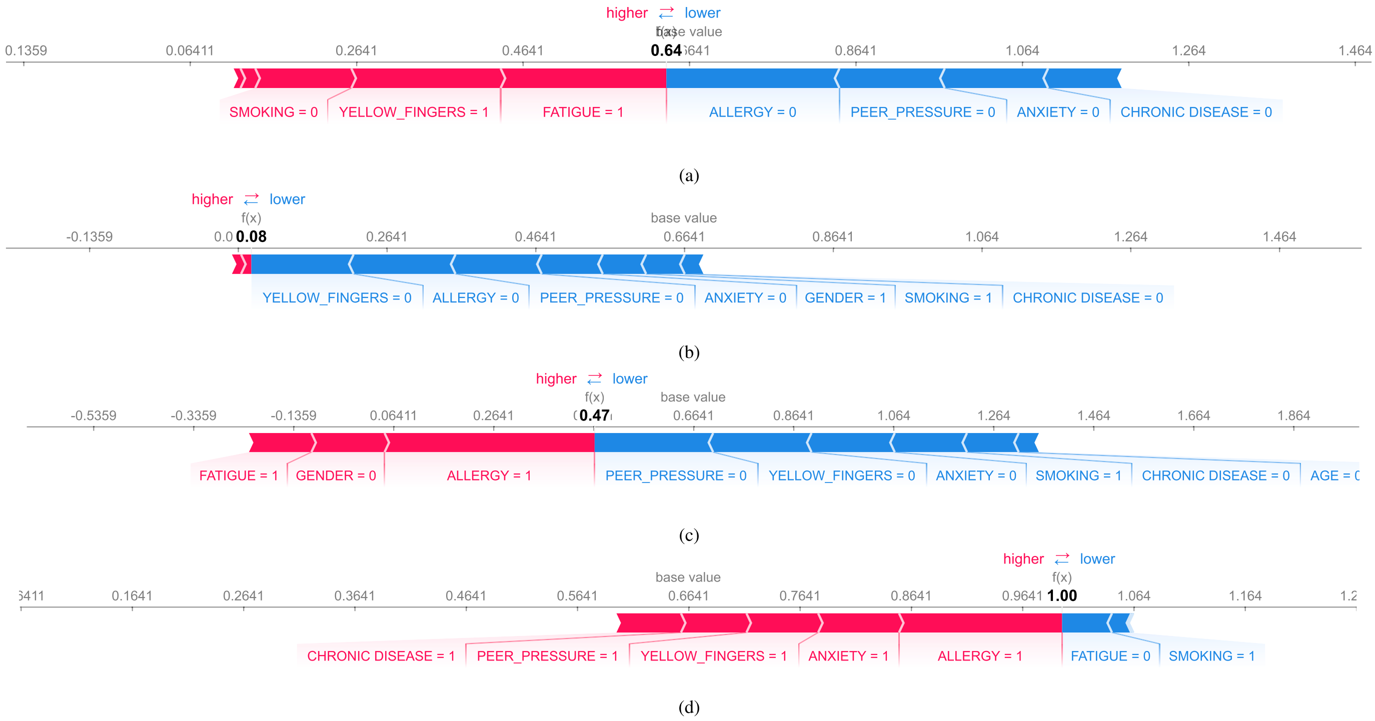


Fig. 8. Graphical representation of parametric comparison of state-of-the-art-techniques with ConvXGB.



(a)

(b)

(c)

(d)

Fig. 9. SHAP force plots of various tested instances for local interpretability by ConvXGB.

predictions, whereas features with points closer to zero are less significant. Fig. 12 is a clustering of the Shapley values for each instance. It depicts the grouped global explanations of 113 patients based on explanation similarity. If the cumulative value exceeds 0.6641, the patient will be classified as class 1; otherwise, the patient will be classified as class 0. There are numerous ways to visualize global explanations in the form of a graph. Fig. 12a depicts a sample ordered by similarity, which groups patients with the most similar characteristics. As shown in the graph, instances between 10 and 50 are grouped into class 1 due to the greater area covered by the red color, while instances between 50 and 90 are grouped into class 0 due to the greater area covered by the blue color. Fig. 12b is a visual representation of the original sample ordering, where samples 0-113 are arranged in ascending order.

6.6. Discussion

As described in section 5, an interpretable lung cancer detection model has been developed in this article. In addition to the detection model, the XAI model is implemented to provide explanations for the model's predictions. Explanations are necessary for medical practitioners to comprehend the actual reason behind a prediction so that an accurate diagnosis can be made and the patient can receive effective treatment.

This article presents a hybrid detection model, combining a convolutional neural network (CNN) with XGBoost. Here, CNN is used for feature learning and the final layer of the proposed CNN is replaced with the well-known machine learning classifier XGBoost, which has the highest classification accuracy among the applied machine

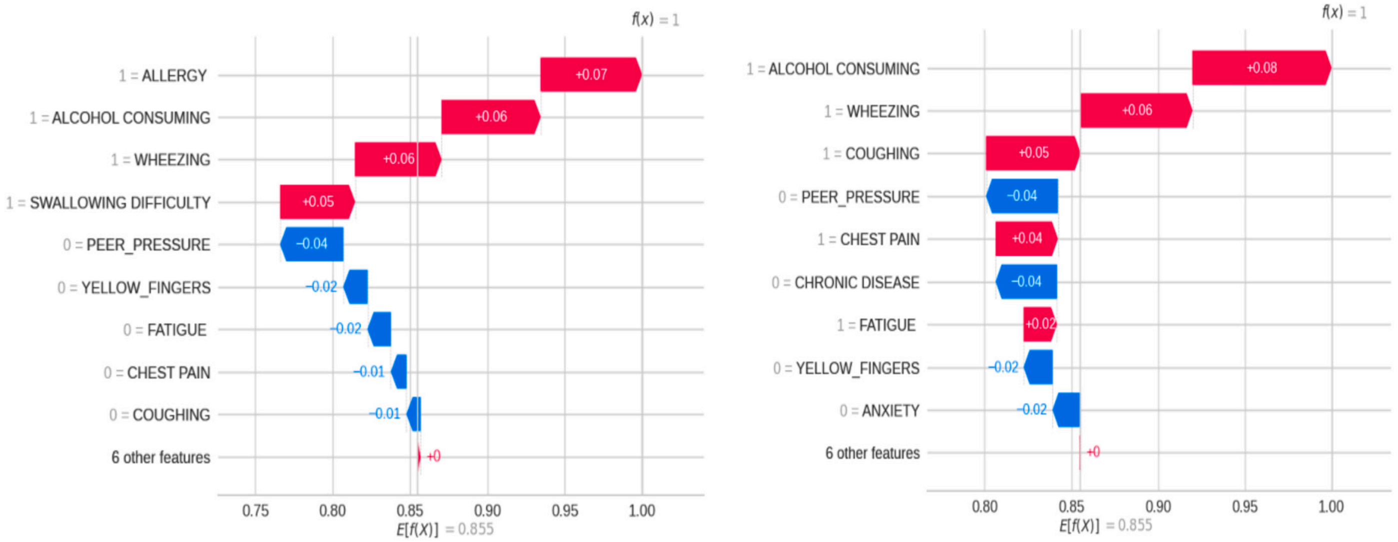


Fig. 10. Local explanation (waterfall) using Shapley values by ConvXGB.

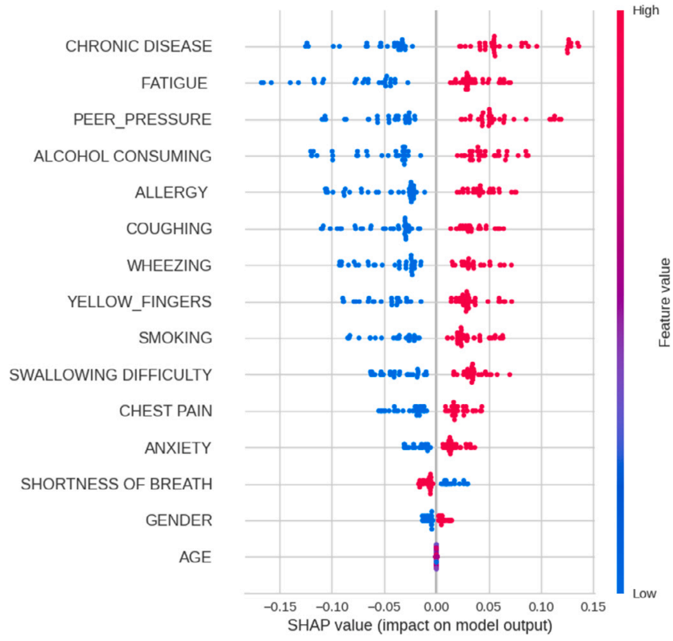


Fig. 11. Global explanation generated by SHAP summary plot.

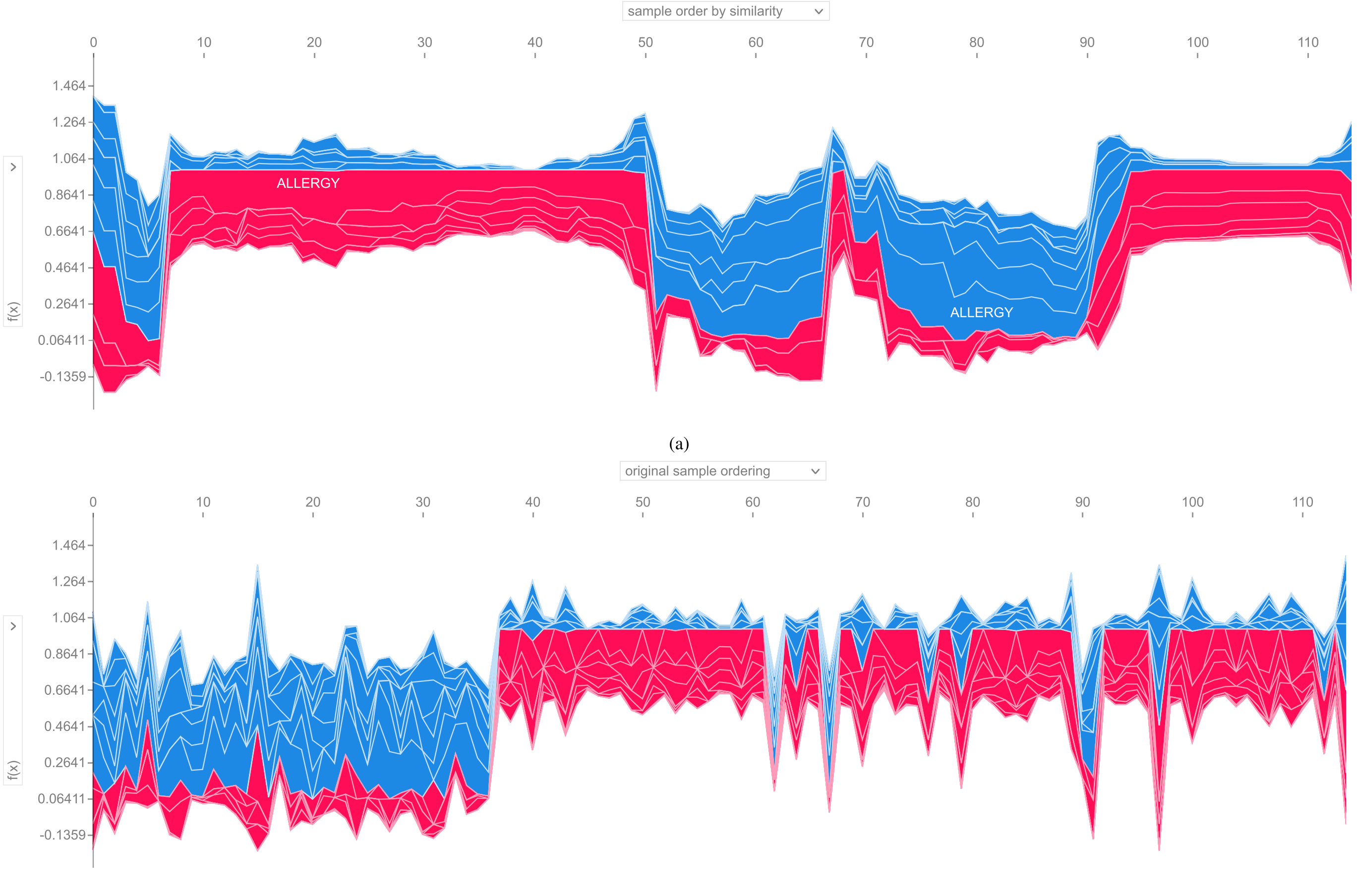
learning algorithms, as shown in Table 2. Table 2 shows how the suggested approach or methodology outperforms many state-of-the-art methods. After classification, the most widely used XAI technique, SHAP, is implemented to provide explanations. Results of the detection module are described in subsubsection 6.5.1 and explainability as in subsubsection 6.5.2. The significance of this study lies in the fact that, on the one hand, lung cancer of the patient is diagnosed and, on the other hand, predictions are explained. Both aid the medical practitioner in diagnosing and treating the patient more effectively. The following are two potential limitations of the research study: Limited Hardware Capabilities: The procurement and upkeep of essential hardware, such as high-performance GPUs and TPUs, which can be expensive. Especially for smaller research institutions and medical facilities with limited budgets, this requirement's high price tag may present obstacles. Therefore, it may be challenging for these institutions to reproduce the study's findings. Accessibility Restriction: In certain regions or institutions, the availability of high-performance hardware may be restricted. This dearth of accessibility may hinder the ability of researchers to implement and validate the proposed deep-learning method.

7. Conclusion

In crucial fields such as healthcare, it is not sufficient to merely make accurate predictions with machine learning. Interpretability is also necessary to comprehend how ML should operate internally to minimize the risk of error or bias in the results. Interpretable Machine Learning (IML) is an interesting approach to interpretability that makes machine learning (ML) more transparent for its consumers. With IML, the understanding of outcomes guarantees that ML users (medical practitioners) can comprehend the values that influence the outcome. In this article, a lung cancer classification based on interpretable deep learning is developed. The overarching goals of the formulated problem is minimization of loss during training and testing of the model so that more accurate predictions can be made, and interpretation or explainability of the made predictions. An efficient deep learning-based model for lung cancer classification is proposed with three primary components: one for feature learning, another for classification, and the third for offering explanations of the predictions made by the proposed hybrid (ConvXGB) model. The proposed *DeepXplainer* has been examined on a variety of metrics, and the findings show that it outperforms the current gold standard in the field. *DeepXplainer's* feature learning and classification are accomplished by combining a CNN with an ML classifier (XGBoost). Lastly, we added interpretability to our model by implementing a model-agnostic model (SHAP), which estimates feature importance for a single prediction and the global level for the entire dataset. Important determinants of lung cancer presence or absence have been identified. Improving an early planning process that lays the groundwork for arranging treatments, money, and resources needed by a cancer patient at the moment of identification is one of the main contributions of this study's expanded findings. In addition to facilitating better therapy, this may also boost post-treatment care for patients and reduce barriers for medical staff.

Code availability

The code of this research is available at the following link: https://github.com/niyazwani/DeepXplainer.git



(a)

(b)

Fig. 12. SHAP results for global interpretability by ConvXGB.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] N. C. Institute, Seer stat fact sheets: all cancer sites, https://seer.cancer.gov/statfacts/html/all.html. (Accessed January 2003).

[2] Y.-H. Lai, W.-N. Chen, T.-C. Hsu, C. Lin, Y. Tsao, S. Wu, Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning, Sci. Rep. 10 (1) (2020) 1–11.

[3] S. Saharan, S. Bawa, N. Kumar, Dynamic pricing techniques for intelligent transportation system in smart cities: a systematic review, Comput. Commun. 150 (2020) 603–625.

[4] S. Saharan, N. Kumar, S. Bawa, DyPARK: a dynamic pricing and allocation scheme for smart on-street parking system, IEEE Trans. Intell. Transp. Syst. 24 (4) (2023) 4217–4234.

[5] J.A. Bartholomai, H.B. Frieboes, Lung cancer survival prediction via machine learning regression, classification, and statistical techniques, in: 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), IEEE, 2018, pp. 632–637.

[6] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90.

[7] S. Saharan, N. Kumar, S. Bawa, An efficient smart parking pricing system for smart city environment: a machine-learning based approach, Future Gener. Comput. Syst. 106 (2020) 622–640.

[8] S. Saharan, S. Bawa, N. Kumar, OP^3 S: on-street occupancy based parking prices prediction system for its, in: 2020 IEEE Globecom Workshops (GC Wkshps), IEEE, 2020, pp. 1–6.

[9] B. Mirza, W. Wang, J. Wang, H. Choi, N.C. Chung, P. Ping, Machine learning and integrative analysis of biomedical big data, Genes 10 (2) (2019) 87.

[10] H.W. Loh, W. Hong, C.P. Ooi, S. Chakraborty, P.D. Barua, R.C. Deo, J. Soar, E.E. Palmer, U.R. Acharya, Application of deep learning models for automated identification of Parkinson's disease: a review (2011–2021), Sensors 21 (21) (2021) 7034.

[11] J.-G. Lee, S. Jun, Y.-W. Cho, H. Lee, G.B. Kim, J.B. Seo, N. Kim, Deep learning in medical imaging: general overview, Korean J. Radiol. 18 (4) (2017) 570–584.

[12] J. Varghese, Artificial intelligence in medicine: chances and challenges for wide clinical adoption, Visc. Med. 36 (6) (2020) 443–449.

[13] F. Doshi-Velez, B. Kim, Towards a rigorous science of interpretable machine learning, arXiv preprint, arXiv:1702.08608, 2017.

[14] A. Holzinger, The next frontier: AI we can really trust, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2021, pp. 427–440.

[15] E.J. Hwang, S. Park, K.-N. Jin, J. Im Kim, S.Y. Choi, J.H. Lee, J.M. Goo, J. Aum, J.-J. Yim, J.G. Cohen, et al., Development and validation of a deep learning–based automated detection algorithm for major thoracic diseases on chest radiographs, JAMA Netw. Open 2 (3) (2019) e191095.

[16] E.J. Topol, High-performance medicine: the convergence of human and artificial intelligence, Nat. Med. 25 (1) (2019) 44–56.

[17] A.B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion 58 (2020) 82–115.

[18] A. Holzinger, M. Dehmer, F. Emmert-Streib, R. Cucchiara, I. Augenstein, J. Del Ser, W. Samek, I. Jurisica, N. Díaz-Rodríguez, Information fusion as an integrative cross-cutting enabler to achieve robust, explainable, and trustworthy medical artificial intelligence, Inf. Fusion 79 (2022) 263–278.

[19] Q. Liu, A. Puthenputhussery, C. Liu, Novel general KNN classifier and general nearest mean classifier for visual classification, in: 2015 IEEE International Conference on Image Processing (ICIP), IEEE, 2015, pp. 1810–1814.

[20] L.B. Nascimento, A.C. de Paiva, A.C. Silva, Lung nodules classification in CT images using Shannon and Simpson diversity indices and SVM, in: Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings 8, Springer, 2012, pp. 454–466.

[21] T.N. Shewaye, A.A. Mekonnen, Benign-malignant lung nodule classification with geometric and appearance histogram features, arXiv preprint, arXiv:1605.08350, 2016.

[22] A.A. Farag, A. Ali, S. Elshazly, A.A. Farag, Feature fusion for lung nodule classification, Int. J. Comput. Assisted Radiol. Surg. 12 (2017) 1809–1818.

[23] A. Shaffie, A. Soliman, H.A. Khalifeh, M. Ghazal, F. Taher, A. Elmaghraby, R. Keynton, A. El-Baz, Radiomic-based framework for early diagnosis of lung cancer, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE, 2019, pp. 1293–1297.

[24] L. Macyszyn, H. Akbari, J.M. Pisapia, X. Da, M. Attiah, V. Pigrish, Y. Bi, S. Pal, R.V. Davuluri, L. Roccograndi, et al., Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques, Neuro-Oncol. 18 (3) (2015) 417–425.

[25] H. Zhang, C.-L. Hung, W.C.-C. Chu, P.-F. Chiu, C.Y. Tang, Chronic kidney disease survival prediction with artificial neural networks, in: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2018, pp. 1351–1356.

[26] M.B. Sesen, T. Kadir, R.-B. Alcantara, J. Fox, S. Michael Brady, Survival Prediction and Treatment Recommendation with Bayesian Techniques in Lung Cancer, AMIA Annual Symposium Proceedings, vol. 2012, American Medical Informatics Association, 2012, p. 838.

[27] A. Agrawal, S. Misra, R. Narayanan, L. Polepeddi, A. Choudhary, A lung cancer outcome calculator using ensemble data mining on seer data, in: Proceedings of the Tenth International Workshop on Data Mining in Bioinformatics, 2011, pp. 1–9.

[28] J.D. Hunter, Matplotlib: a 2D graphics environment, Comput. Sci. Eng. 9 (03) (2007) 90–95.

[29] H. Nori, S. Jenkins, P. Koch, R. Caruana, InterpretML: a unified framework for machine learning interpretability, arXiv preprint, arXiv:1909.09223, 2019.

[30] M. Al-Shabi, B.L. Lan, W.Y. Chan, K.-H. Ng, M. Tan, Lung nodule classification using deep local–global networks, Int. J. Comput. Assisted Radiol. Surg. 14 (2019) 1815–1819.

[31] D. Kumar, A. Wong, D.A. Clausi, Lung nodule classification using deep features in CT images, in: 2015 12th Conference on Computer and Robot Vision, IEEE, 2015, pp. 133–138.

[32] J. Kuruvilla, K. Gunavathi, Lung cancer classification using neural networks for CT images, Comput. Methods Programs Biomed. 113 (1) (2014) 202–209.

[33] E. Dandıl, M. Çakiroğlu, Z. Ekşi, M. Özkan, Ö.K. Kurt, A. Canan, Artificial neural network-based classification system for lung nodules on computed tomography scans, in: 2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR), IEEE, 2014, pp. 382–386.

[34] N. Nasrullah, J. Sang, M.S. Alam, M. Mateen, B. Cai, H. Hu, Automated lung nodule detection and classification using deep learning combined with multiple strategies, Sensors 19 (17) (2019) 3722.

[35] G.S. Tran, T.P. Nghiem, V.T. Nguyen, C.M. Luong, J.-C. Burie, Improving accuracy of lung nodule classification using deep learning with focal loss, J. Healthc. Eng. (2019) 2019.

[36] M.A. Ahmad, C. Eckert, A. Teredesai, Interpretable machine learning in healthcare, in: Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, 2018, pp. 559–560.

[37] D. Dave, H. Naik, S. Singhal, P. Patel, Explainable AI meets healthcare: a study on heart disease dataset, arXiv preprint, arXiv:2011.03195, 2020.

[38] A. Holzinger, C. Biemann, C.S. Pattichis, D.B. Kell, What do we need to build explainable AI systems for the medical domain? arXiv preprint, arXiv:1712.09923, 2017.

[39] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, N. Elhadad, Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp. 1721–1730.

[40] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).

[41] C.-T. Kor, Y.-R. Li, P.-R. Lin, S.-H. Lin, B.-Y. Wang, C.-H. Lin, Explainable machine learning model for predicting first-time acute exacerbation in patients with chronic obstructive pulmonary disease, J. Person. Med. 12 (2) (2022) 228.

[42] H. Chen, S.M. Lundberg, G. Erion, J.H. Kim, S.-I. Lee, Forecasting adverse surgical events using self-supervised transfer learning for physiological signals, npj Digit. Med. 4 (1) (2021) 167.

[43] E. Withnell, X. Zhang, K. Sun, Y. Guo, Xomivae: an interpretable deep learning model for cancer classification using high-dimensional omics data, Brief. Bioinform. 22 (6) (2021) bbab315.

[44] J. Lu, R. Jin, E. Song, M. Alrashoud, K.N. Al-Mutib, M.S. Al-Rakhami, An explainable system for diagnosis and prognosis of Covid-19, IEEE Int. Things J. 8 (21) (2020) 15839–15846.

[45] M.A. Alves, G.Z. Castro, B.A.S. Oliveira, L.A. Ferreira, J.A. Ramírez, R. Silva, F.G. Guimarães, Explaining machine learning based diagnosis of Covid-19 from routine blood tests with decision trees and criteria graphs, Comput. Biol. Med. 132 (2021) 104335.

[46] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, H. Ghaemmaghami, C. Fookes, A robust interpretable deep learning classifier for heart anomaly detection without segmentation, IEEE J. Biomed. Health Inform. 25 (6) (2020) 2162–2171.

[47] A. Singh, J. Jothi Balaji, M.A. Rasheed, V. Jayakumar, R. Raman, V. Lakshminarayanan, Evaluation of explainable deep learning methods for ophthalmic diagnosis, Clin. Ophthalmol. (2021) 2573–2581.

[48] E. Zihni, V.I. Madai, M. Livne, I. Galinovic, A.A. Khalil, J.B. Fiebach, D. Frey, Opening the black box of artificial intelligence for clinical decision support: a study predicting stroke outcome, PLoS ONE 15 (4) (2020) e0231166.

[49] B. Alsinglawi, O. Alshari, M. Alorjani, O. Mubin, F. Alnajjar, M. Novoa, O. Darwish, An explainable machine learning framework for lung cancer hospital length of stay prediction, Sci. Rep. 12 (1) (2022) 1–10.

[50] B. Pfeifer, H. Baniecki, A. Saranti, P. Biecek, A. Holzinger, Multi-omics disease module detection with an explainable Greedy decision forest, Sci. Rep. 12 (1) (2022) 16857.

[51] DataWorld, Survey lung cancer datset, https://data.world/sta427ceyin/survey-lung-cancer. (Accessed January 2023).

[52] B. Gupta, S. Tiwari, Lung cancer detection using curvelet transform and neural network, Int. J. Comput. Appl. 86 (1) (2014).

[53] M. Dirik, Machine learning-based lung cancer diagnosis, Turk. J. Eng. 7 (4) (2023) 322–330.

[54] A.E. Celik, J. Rasheed, A. Yahyaoui, Machine learning approaches for lung cancer prediction, in: 2022 12th International Conference on Advanced Computer Information Technologies (ACIT), IEEE, 2022, pp. 540–543.

[55] I.M. Nasser, S.S. Abu-Naser, Lung cancer detection using artificial neural network, Int. J. Eng. Inf. Syst. 3 (3) (2019) 17–23.