

An Efficient Dual-Sampling Approach for Chest CT Diagnosis

Khalaf Alshamrani^{1,2}, Hassan A Alshamrani¹

¹Radiology Sciences Department, College of Medical Sciences, Najran University, Najran, Saudi Arabia; ²School of Medicine and Population Health, University of Sheffield, Sheffield, UK

Correspondence: Khalaf Alshamrani, Email Kaalshamrani@nu.edu.sa, k.alshamrani@sheffield.ac.uk

Background: This paper aimed to enhance the diagnostic process of lung abnormalities in computed tomography (CT) images, particularly in distinguishing cancer cells from normal chest tissue. The rapid and uneven growth of cancer cells, presenting with variable symptoms, necessitates an advanced approach for accurate identification.

Objective: To develop a dual-sampling network targeting lung infection regions to address the diagnostic challenge. The network was designed to adapt to the uneven distribution of infection areas, which could be predominantly minor or major in different regions.

Methods: A total of 150 CT images were analyzed using the dual-sampling network. Two sampling approaches were compared: the proposed dual-sampling technique and a uniform sampling method.

Results: The dual-sampling network demonstrated superior performance in detecting lung abnormalities compared to uniform sampling. The uniform sampling method, the network results: an F1-Score of 94.2%, accuracy of 94.5%, sensitivity of 93.5%, specificity of 95.4%, and an area under the curve (AUC) of 98.4%. However, with the proposed dual-sampling method, the network reached an F1-score of 94.9%, accuracy of 95.2%, specificity of 96.1%, sensitivity of 94.2%, and an AUC of 95.5%.

Conclusion: This study suggests that the proposed dual-sampling network significantly improves the precision of lung abnormality diagnosis in CT images. This advancement has the potential to aid radiologists in making more accurate diagnoses, ultimately benefiting patient treatment and contributing to better overall population health. The efficiency and effectiveness of the dual-sampling approach in managing the uneven distribution of lung infection areas are key to its success.

Keywords: lung cancer, SVM, KNN, dual sampling, uniform sampling, balanced class sampling, under sampling and over fitting

Introduction

Deep learning has been shown to be effective in identifying abnormalities in medical imaging.^{1,2} However, medical professionals are concerned about the capacity of models to absorb and perceive information,³ which is crucial for lung diagnosis. Class stimulation mapping⁴ and gradient-weighted class stimulation mapping⁵ both make localized heat maps that show key areas that are strongly connected to expected outcomes. This makes it easier to choose a model. These examples are provided in order to facilitate the process of selecting models. Both patients and physicians are put at serious risk when healthcare providers exploit false data. Misdiagnoses and improper treatments can result from data that is incorrect in the first place, in which the medical records of a patient are altered to display erroneous measurements or outcomes of tests. This could lead to misdiagnosing a patient's disease or giving them the wrong treatment.

Deep learning has the ability to diagnose diseases,⁶ classify organs,⁷ and perform other tasks.^{8–11} Convolutional neural networks (CNNs) have been put to use in the process of analyzing and assessing CT images.^{12–14} CNN detects adaptive enhancement agent-enhanced CT imaging of liver lesions.¹⁵ Chest CT is capable of diagnosing pulmonary nodules^{16,17} and pulmonary tuberculosis.¹⁸

The universal paradigm is not capable of selecting long-tailed datasets in a reliable manner for network training.^{19,20} Because of the uniform sampling, the network may choose to disregard these occurrences. This is achieved via the use of class re-sampling.^{21–24} Mini-batches are used to update class examples in order to improve the performance of long-tailed datasets.

Over-sampling causes data to be repeated for minority-labeled classes,²⁵ but under-sampling causes samples to be arbitrarily eliminated in order to bring class numbers closer together.²⁶ There is a possibility that oversampling may overfit minority groups.²⁷

In datasets that have an uneven distribution of classes, classification algorithms have a difficult time distinguishing between minor classes and the principal class. The challenges frequently involve datasets that contain this uneven distribution of classes. In the beginning, efforts were made to correct the class imbalance by using under- and oversampling to reduce the size of the information sample in the majority group and to increase the size of the sample in the minority group, respectively. The sampling that was carried out using the ensemble classifier had a positive outcome.

When the sample distribution is not balanced, it indicates that one class has a greater number of examples than the other classes. The authors of²⁸ suggested a clustering-SVM unbalanced data classification technique as a solution to the problems of insufficient classification accuracy and significant time requirements. The approach suggests under-sampling the samples that make up the majority in order to consider the distributions of the samples that make up the minority.

Data analysis relies on clustering, a basic approach that aims to group data items that are similar.

Clustering does not make use of any labels or categories that have been previously established for the data points; rather, it is an unsupervised learning technique. Instead, it automatically groups them based on the inherent commonalities in the data. This is why clustering is so effective for finding patterns in unlabeled data.

Every one of the several clustering methods has its own set of advantages and disadvantages. By reducing the distances between data points and their assigned cluster centers (centroids), the K-means cluster is suggested to be a better fit in such a situation. In order to provide a method for sampling and selecting members of minority categories, K-means clusters are used.^{29,30} In this scenario, the initial step in training the network involves carrying out uniform sampling throughout the whole of the dataset. Utilizing the suggested sample and selection strategy may help enhance the diagnostic performance of uniform sampling for a subset of anomalies that is underrepresented in the population. As a result, the dual-sampling network will be established as a result of this study. After the K-means approach has clustered and grouped the samples of the minority type, we will construct each cluster using the new data, and then valid confirmation will be completed. In conclusion, the classifiers are put to use to demonstrate the validity of the experimental findings. Both the SVM and the KNN classifiers have had their performance evaluated and compared.

Materials and Methods

This study employs SVM and KNN to provide a successful approach for uniform and balanced class sampling that may be used for chest X-ray and CT diagnostics. A data set that may be utilized for training and testing purposes is constructed with the help of the characteristics of a chest X-ray or a CT scan image. Following that, SVM and KNN will be used to categorize the images. The approach that is advised delivers a reliable diagnosis based on chest X-ray or CT scan images by making use of a dataset that is accessible to the general public. For the purpose of minority class formation in unequal data sets, this research suggests an oversampling technique that is based on K-means clustering. This strategy enhances the accuracy of minority class classification, whose performance has been examined using K-nearest neighbor (KNN) and support vector machine (SVM) classifiers. The data sets focus on how to better categorize members of minority classes. There is a possibility that the algorithm layer and data layer approaches will improve the performance of minority class classification in unequal data sets.

Proposed Methodology

The proposed method for minority class classification might work better if it includes an algorithm layer that sorts data by changing the cost function of samples from different classes, as well as the probability density, category boundary, and other factors. After preprocessing the training set using the data layer technique, classifier training should be performed using the discarded data set. Data layer resampling could be oversampled or undersampled. Duplicating samples from minority groups is the most straightforward approach to oversampling because doing so improves the accuracy of those groups' classifications. It does not give any extra knowledge about the minority class, which might narrow the decision-making scope of the classifier and lead to overlearning. Under-sampling will limit the number of examples of the

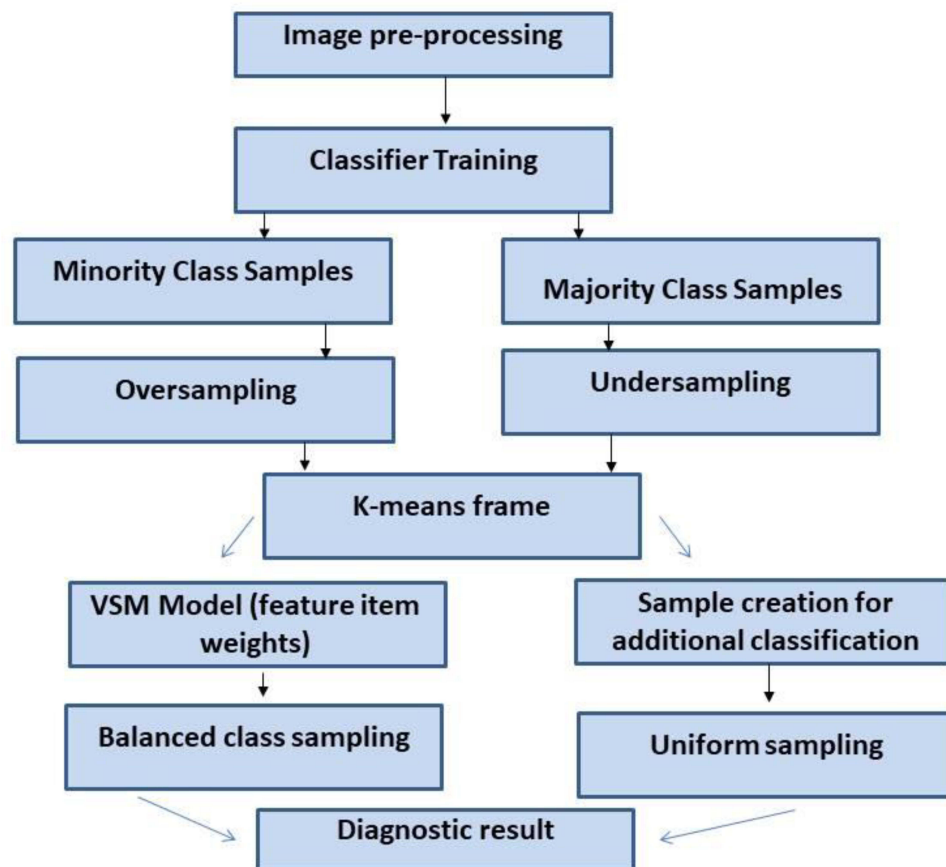


Figure 1 Block diagram of the proposed method.

majority class in order to increase the classification performance of the minority class. The most direct approach to under-sampling is the arbitrary removal of certain examples from the majority of types of class samples. As a result, more complex resampling methods were suggested. Figure 1 shows the block diagram of the proposed method.

Clustering and K-Means Framing

Clustering groups data or objects by attributes to make them as similar as feasible and increase variety. K-means method is a standard clustering technique that uses distance as an indicator of similarity. The closer two items are, the more similar they are. The adjustment rule is the distance between a particular data point and the working prototype classification centre) and the optimised goal function. K-means technique uses To obtain a particularly comprehensive categorization of any matching starting clustering centre vector, the Euclidean distance is used as a measure of similarity which is written as $V = (v_1, v_2, \dots, v_k)^T$, resulting in the minimal evaluating standard J_c . The algorithm always clusters using criteria function and squared error, which are specified as

$$J_c = \sum_{i=1}^k \sum_{p \in C_i} \|p - M_i\|^2$$

M_i is the mean data object value in class C_i , and p is its spatial location. The K-means frame is defined using following steps:

- 1) Fix the data set at n , $I = 1$, choose K pieces of initial cluster centre $Z_j(I)$, $j = 1, 2, 3, \dots, k$;

- 2) Calculate the distance that exists between each data item and the cluster centre stage $D(x_i, Z_j(I))$, $i = 1, 2, 3, \dots, n, j = 1, 2, 3, \dots, k$, if it fulfils
 $D(x_i, Z_k(I)) = \min \{D(x_i, Z_k(I)), j = 1, 2, 3, \dots, n\}$, so $x_i \in w_k$;
- 3) Computer Square error and criteria function J_c

$$J_c = \sum_{i=1}^k \sum_{k=1}^{n_j} \|x_k(j) - Z_j(I)\|^2$$

- 4) Judge; if $|J_c(I) - J_c(I-1)| < \xi$, until end;
else the new cluster center (k) is computed as, $I = I + 1$, $\left| Z_j(I) = \frac{1}{n} \sum_{i=1}^j x_i(j) \right|$, return 2.

Thus, the cluster algorithm may be used to create new samples by crossing minority class samples. Before clustering, the document is vectorized in n-dimensional characteristic space.

Vector Space Model (VSM)

Vector Space Model (VSM) is a popular document model. VSM vectors include feature item weights for each dimension. Formal description:

- 1) Add $n \times n$ pixel entries from the image as a vector expression of any image X. Such that $X = \{X_{1,1}, \dots, X_{n,n}\}$, $X_{1,1}$ is pixel representation of image.
- 2) The image can be represented with the weights for each pixel such that $X = (w_{1,1}X_{1,1}, \dots, w_{n,n}X_{n,n})$, where $w_{1,1}, \dots, w_{n,n}$ are the weights for corresponding image pixels.

The weight in the image, $w_{n,n}$ is 1 or 0. This weight is calculated using TF2IDF formulas

$$w_{i,j} = \frac{tf_{i,j}^d \times \log\left(\frac{N}{n_{i,j}} + 0.01\right)}{\sqrt{\sum_{i,j \in d} \left[tf_{i,j}^d \times \log\left(\frac{N}{n_{i,j}} + 0.01\right)\right]^2}}$$

Here, N denotes the overall pixels present, $n_{i,j}$ is the sample set that includes $X(i,j)$, and denominator is normalisation factor. $w_{i,j}$ value is a real integer number from 0–1.

Sample Creation for Additional Classification

First, use K-mean to cluster related samples. Then, code the cluster samples by excluding incorrect samples is validation. New samples should have randomization and additional category information. It should be as representative as feasible. Some new samples will be invalid and outside the cluster, so they are rejected.

Suppose class C_p is cluster centroid is $C_p = \{C_1, C_2, \dots, C_n\}$. The sample distance between x_i and centroid: $d_{ji}^p = \|x_j^p - C_i^p\|$, $1 \leq i \leq N_p$, $1 \leq j \leq m$

Algorithm steps:

- 1) input $k = 5$, apply K-means method to partition class $C_p(i)$ into five clusters, the i^{th} cluster is ($1 \leq i \leq 5$);
- 2) Randomly pair each m_i in the $C_p(i)$ with a group of $\frac{m_i}{2}$;
- 3) Validate new individuals in $C_p(i)$, reject invalid samples;
- 4) In case the simulation does not satisfy the criteria, then return to step 2) to proceed on the next validation till sufficient samples are clustered, or the algorithm terminate and output C_p as new sample space.

Balanced Class Sampling (BCS)

Balanced class sampling involves repeating data sampling for small instances in each mini batch during training. For training using balanced class sampling (BCS), uniform sampling is applied throughout the dataset. In particular, the network receives each training dataset sample once, with the same probability value for each epoch. Therefore, the proposed model may examine the whole dataset while retaining the inherent data distribution. A second network is

trained using BCS due to an abnormal size distribution imbalance. It increases mini-batch sampling of small- and large-nodule areas. The dataset is divided into three groups, depending on the nodule's proportion of volume to the lung:

1. small-nodule-area
2. large-nodule-area
3. Normal lung.

The nodule instances are classified as small-area or large-area based on a <0.030 threshold. A ratio >0.001 is for large-nodule areas, whereas the remainder are modest. The three groups have [small nodule, big nodule, normal] samples. The weights for the three groups are defined as small nodules, big nodules, and normal lung. Small and large nodules are greater than 1 because the threshold levels are low. Each training fold has a weight of 1.5. In any batch, it is random to choose among [small nodule, big nodule, normal] and a uniform sample from the chosen group. This technique increases the likelihood of sampling accuracy with minor abnormalities and with major abnormalities. All mini-batches are size-balanced during training the proposed model.

Uniform Sampling

Class-balanced sampling might improve minority representation and reduce nodule area bias in lung cancer patients. It may overfit minority groups. Uniform sampling may reliably comprehend depictions of features from data that was originally collected. An ensemble learning layer weights prediction outputs from both models in a dual-sampling technique. The estimates of two models are combined in an ensemble learning layer trained with different sampling strategies to provide the diagnostic result.

Statistical Analysis

Statistical data analysis to report the robustness of a detection strategy for lung tumour detection utilising mean values and standard deviations of performance indices. The mean values and standard deviations of the performance indices for a detection technique for lung tumour detection are shown in Table 1.

It is possible to draw the conclusion that the detection method is running efficiently based on the fact that the mean values of all of the performance indicators are pretty close to 1. This may be inferred from the data. It is possible to draw the conclusion that the detection method maintains a consistent level of performance overall by noting that all of the performance indices have extremely small standard deviations. This may be inferred from the fact that all of the performance indices have very modest standard deviations. Examining the distribution of performance indicators is another way to find out how reliable the current detection system is. If the performance indices are distributed normally, we can say that the detection system is trustworthy. The fact that the performance indices in the given scenario follow a Gaussian distribution supports the validity of the detection system. Thus, it appears that small discrepancies in the input data are unlikely to have a significant impact on the detection technique's performance. To find out if the method works

Table 1 The Mean and Standard Deviation of the Performance Index

Performance Index	Mean	Standard Deviation
Sensitivity	0.9	0.05
Specificity	0.95	0.02
Positive predictive value	0.8	0.07
Negative predictive value	0.98	0.01

as a detection strategy, it is crucial to assess how well it holds up under repeated testing. A robust detection method is one that can endure disturbances without significantly reducing its performance. This implies that factors like noise are less likely to have an impact on it. The performance indicators' mean values and standard deviations were calculated using the number of samples that should be included when reporting on a detection approach's robustness utilising statistical data analysis.

Experimental Results

In order to evaluate the model's capacity for generalization, this study used a dataset including 150 images that was retrieved from an internet database (<https://www.kaggle.com/datasets>). Under conditions of balanced class sampling, the suggested model achieved an area under the curve of 95.5%, accuracy of 95.2%, sensitivity of 94.2%, specificity of 96.1%, and F1-score of 94.9%. The suggested model was successful in achieving an F1-score of 94.2%, accuracy of 94.5%, sensitivity of 93.5%, specificity of 95.4%, and area under the curve of 98.4%. Both balanced class sampling and uniform sampling were used to investigate two different classifiers. After the data were normalized, the SVM classifier showed much superior average classification accuracy as compared to KNN. SVM was not able to identify with greater overall average classification accuracy when the data were unbalanced relative to the optimal classifier. A dual sampling method has been employed to discover significant diagnostic regions for the purpose of gaining a better understanding of the suggested model's conclusion. If the training set is heavily skewed towards small-nodule areas, then the model may be better at detecting small nodules than large nodules. This could lead to the model making inaccurate predictions on cases with large nodules. The error rates for the training set, the validation set, and the test set are shown in Figure 2.

This study examined three training, validation, and testing set proportions: 80:10:10, 70:15:15, and 75:10:15. Figure 3 shows training, validation, and test dataset error histogram plots. The training set error histogram plot shows how effectively the model predicts the goal result from training data. It shows the model's performance and shortcomings by showing the distribution of errors between expected and actual values. The validation set error histogram shows how well

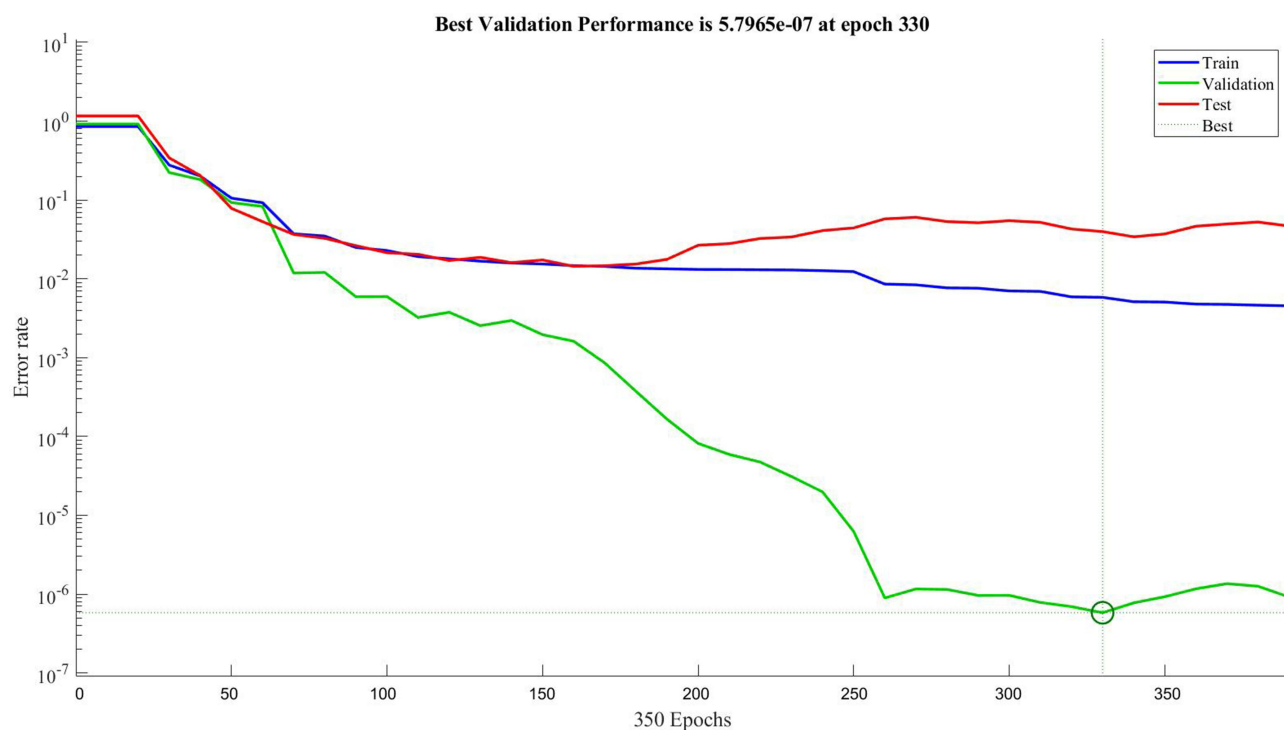


Figure 2 Plot of error rates for training, validation and test set.

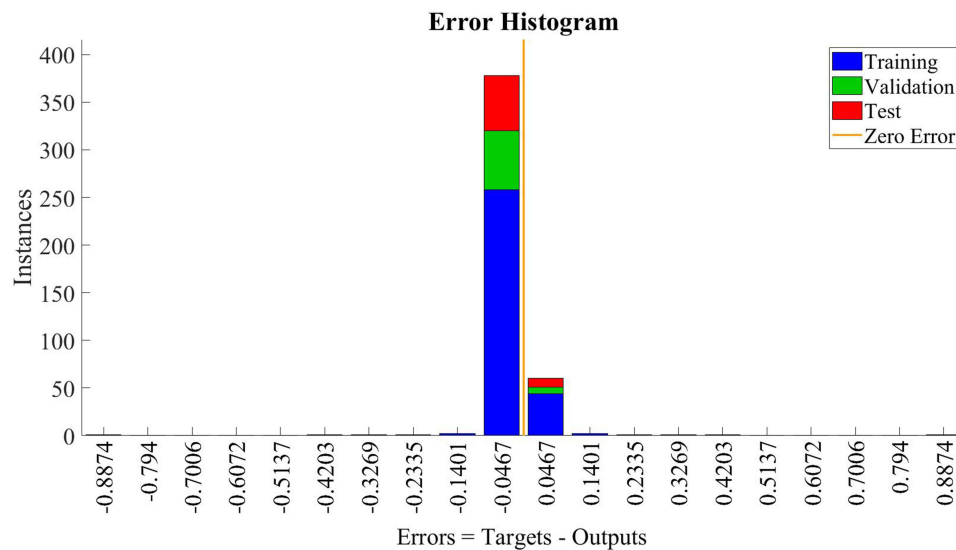


Figure 3 Error histogram plot for training, validation and test set.

the model works on untested data. This data determines whether the model overfits the training data and if the hyperparameters need to be modified. The error histogram for the test set illustrates how well the model functions when applied to data that it has not seen before. The performance of the model in terms of generalization may be evaluated with the use of this data, as can the identification of any possible problems with the design of the model (see Figures 4 and 5).

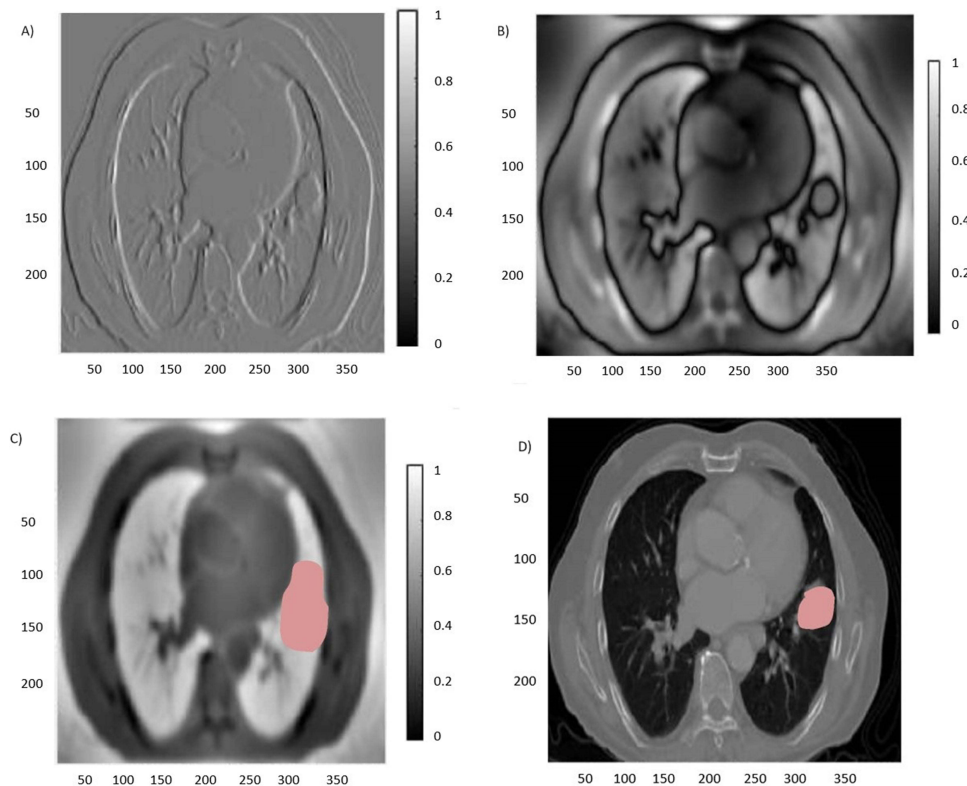


Figure 4 Qualitative results (A) Balanced Class sampling of small nodule area (B) Uniform Sampling of large nodule area (C) ROI For Large nodule area (D) For small nodule area.



Figure 5 Confusion matrix for (A) training (B) validation (C) test set (D) overall matrix.

The ROC plots for the training dataset, the validation dataset, the test dataset, and the total ROC are shown in Figure 6.

The number of images taken from the dataset that make up the training, validation, and test sets is shown in a variety of proportions in Table 2 and Table 3. The results of an analysis of numerous performance measures are shown in Table 4, which uses balanced class sampling. In most cases, the error rate for the training set is somewhere between 10 and 15%. In most cases, the error rate for the validation set is somewhere between 15 and 20%. In most cases, the error rate for the test set falls somewhere between 20 and 25%. In order to improve the accuracy of the trained model via the use of the suggested method, two distinct sampling strategies are utilized. When utilizing balanced class sampling, the amount of training error for a big or tiny lung nodule is dependent on the complexity of the model as well as the size of the training set. On the other hand, it is more likely to be less than the error found in a balanced class when uniform

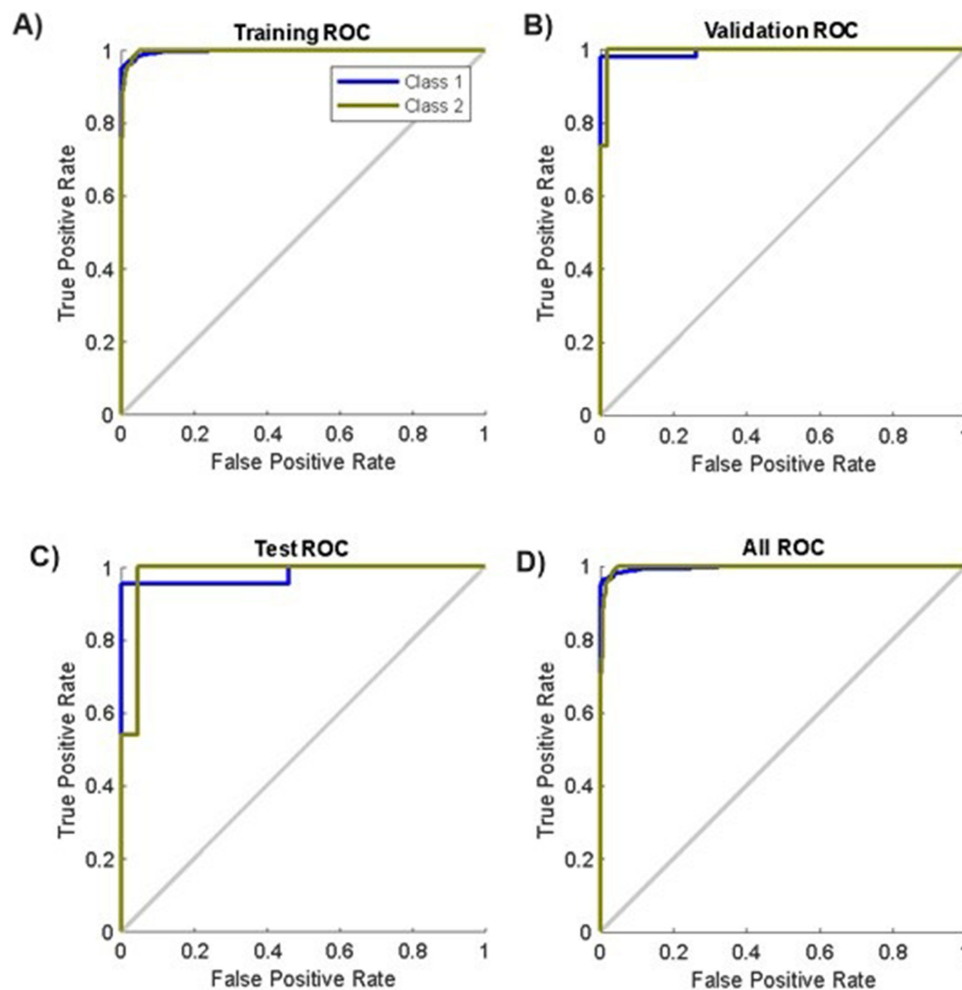


Figure 6 ROC plot for (A) training, (B) validation, (C) test set and (D) all ROC.

sampling is compared to it. Table 5 is available here, and it compares the accuracy of various classifiers by category. The results of evaluating several performance indicators using a variety of classifiers are shown in Table 6.

It has been discovered that it is feasible to build classes that are uniform and well-balanced by utilizing a support vector machine classifier. By using dual sampling, one is able to exercise control over the tradeoff that exists between the size of the margin and the misclassification rate. To acquire the required results, adjust the sampling approach to control the border form. By changing the number of neighbors, a KNN classifier may create consistent, balanced classifications. KNN classifiers allow this. Adding neighbors improves categorization consistency and balance. To achieve

Table 2 Number of Images from the Dataset for Training, Validation and Test Sets in Different Ratios

Characteristic of Dataset	Number of Images		
	Training dataset (80%)	Validation dataset (10%)	Test dataset (10%)
Small Nodule Area	57	8	8
Large Nodule Area	39	5	5
Normal Lung	40	5	5

(Continued)

Table 2 (Continued).

	Number of Images		
Characteristic of Dataset	Training dataset (80%)	Validation dataset (10%)	Test dataset (10%)
Small Nodule Area	53	7	11
Large Nodule Area	36	5	8
Normal Lung	37	5	8
	Number of Images		
Characteristic of Dataset	Training dataset (70%)	Validation dataset (15%)	Test dataset (15%)
Small Nodule Area	49	11	11
Large Nodule Area	35	7	7
Normal Lung	36	7	7

Table 3 Evaluation of Various Performance Metrics Using Uniform Sampling

Parameter	Training Dataset	Validation Dataset	Test Dataset
FI-Score	94.2	81.6	80.3
Accuracy	94.5	85.4	86.7
Sensitivity	93.5	87.3	86.7
Specificity	95.4	87.5	86.4
AUC	98.4	93.4	94.3

Table 4 Evaluation of Various Performance Metrics Using Balanced Class Sampling

Parameter	Training Dataset	Validation Dataset	Test Dataset
FI-Score	94.9	82.2	80.9
Accuracy	95.2	86.0	87.3
Sensitivity	94.2	87.9	87.3
Specificity	96.1	88.1	87.0
AUC	95.5	94.1	95.0

Table 5 Class-Wise Comparison of Accuracy Using Different Classifiers

Dataset Original		KNN	SVM
	Minority Class Sample	10	12
	Majority Class Sample	87	65
	Accuracy	92.4%	90.3%

(Continued)

Table 5 (Continued).

Dataset Original		KNN	SVM
Dataset with new samples			
	Minority Class Sample	22	19
	Majority Class Sample	75	58
	Accuracy	94.3%	93.5%

Table 6 Evaluation of Various Performance Metrics Using Different Classifiers

Classifier	F1-Score	Accuracy	Sensitivity	Specificity	AUC
Uniform Sampling					
KNN	93.4	94.5	93.2	95.3	94.3
SVM	92.1	93.2	92.8	94.1	93.2
Balanced Class Sampling					
KNN	94.1	95.2	93.9	96.0	95.0
SVM	92.7	93.9	93.5	94.8	93.9

Table 7 Performance Evaluation of the Proposed Approach in Comparison to State-of-The-Art Methodologies

	Method Used	Performance in Terms of Accuracy
Huang B., et al, 2020 ²⁸	An under-sampling approach (cluster-based instance selection)	88.8%
Tsai C.-F., et. al, 2019 ³¹	SVM combined with resampling technology for imbalance data	95.1%
Jiang S. et. al, 2021 ³²	Imbalance data classification algorithm	90.12% (Average Negative F-values)
Proposed Work	Dual Sampling	94.3% (with KNN), 93.5% (with SVM)

categorization homogeneity and equilibrium, the distance metric might be adjusted. [Table 7](#) compares the recommended technique to state-of-The-art methods.^{28–31}

Discussion

In general, the likelihood of malignancy increases proportionately with the size of the nodule. On the other hand, there are little nodules that may develop into malignant tumors, and there are huge nodules that are completely benign.

In medical practice, lung nodules are often categorized as either being tiny (less than 5 millimeters in diameter), intermediate (5–10 millimeters in diameter), or big (more than 10 millimeters in diameter). The nodule's size and the likelihood that it will turn cancerous are what determine these criteria.

In the academic fields of data mining and machine learning, class-imbalance learning is a subject that draws a significant amount of interest from researchers and students alike. If the sample distribution is not balanced, then one of the classes will have a larger number of examples than the others. This will be the case if there is a difference in the number of instances between the classes. The authors of the paper²⁸ propose a clustering-SVM unbalanced data classification approach as a solution to the problems of low classification accuracy and considerable time requirements, respectively. This solution can be found in the article. The strategy recommends using a smaller-than-usual sample size of the majority of the samples wherever it is feasible in order to provide an accurate representation of the distributions of the minority samples.

Cluster-based instance selection,³¹ or CBIS for short, is a unique approach to the problem of under-sampling that integrates the analysis of clusters with the process of selecting cases. CBIS is most often referred to as CBIS. The acronym CBIS is another common name for CBIS. When selecting an instance of the dataset, clustering generates subsets of the dataset that belong to the majority class, but when picking an instance of the dataset, data samples that are not relevant to each subclass are removed.

The hybrid improved adaptive SVM methodology presented in³² is a strategy that may be used for the management of imbalanced data. With an accuracy of 95.11%, the data revealed that their technique was superior.

The ability to recognize small and large lung nodules independently of one another facilitates the ability to distinguish between benign and malignant nodules. Because a greater likelihood of cancer is associated with bigger nodules, one might assume that smaller nodules are less likely to be malignant. This indicates that larger nodules may need removal, whereas smaller nodules may be examined more carefully.

Currently, chest CT is capable of performing diagnostic procedures. The success of these procedures is significantly correlated with the grade of the photos that are generated. Excessive noise, artifacts, and inadequate contrast in images may significantly impede the precision of diagnoses. An asymmetrical dataset, which is characterized by an imbalance in the number of normal occurrences compared to aberrant ones, may induce the model to exhibit bias, which could impact its performance. The current model architecture must be executed effectively by considering a multitude of distinct factors. Feature extraction, sampling techniques, and fusion approaches are among the aspects that are discussed. The computational resource requirements of the CT techniques that are currently available may be substantial, necessitating the provision of high-performance hardware.

Having shown its effectiveness in the diagnosis of chest computed tomography (CT), the dual-sampling network approach has a significant potential for use in other medical imaging diagnostic jobs. These responsibilities encompass a variety of fields, such as cardiology, orthopedics, oncology, and radiology. By improving the image quality and diagnostic accuracy in a number of applications (for example, prenatal imaging and abdominal ultrasonography), it contributes to the enhancement of lesion identification and characterization., CT images have wider applications, such as: differentiating types of pneumonia, predicting tumor genotypes, and assessing treatment efficacy.^{33–36} This is accomplished through the thorough examination of functional and anatomical data.

The use of the dual-sampling network strategy has the capacity to produce a range of inventive characteristics and benefits in the realm of chest CT diagnostics. These tactics allow people to maximize their use of available resources. The approach efficiently employs imaging resources by acquiring two separate sets of CT images with varying spatial resolutions throughout the image collection process. This facilitates the efficient application of imaging resources. Obtaining the complementary information is possible after it has been received. A comprehensive depiction of the patient's state may be attained by merging the data acquired from high-resolution and low-resolution images. Unlike the commonly used single-resolution methods, the dual-sampling methodology allows for the retrieval of a wider range of information. The use of the dual sampling technique has the capacity to augment the precision of diagnostic processes by capitalizing on the benefits that picture sets provide. This is because photo collections provide a multitude of advantages. The radiation dose required by these approaches is decreased due to the need for pictures with a lower resolution. A patient's radiation exposure may be significantly reduced by administering a lower amount of radiation. By using image reconstruction methods, the dual-sampling methodology has the capability to improve the quality of pictures.

Lung nodules are treated differently depending on the size of the lesion as well as the likelihood that the tumor is cancerous. For example, although smaller nodules may simply require regular CT scans for monitoring, larger nodules may necessitate surgical excision. No matter how small, lung nodules enhance a patient's risk of developing lung cancer. It is therefore probable that a more thorough evaluation of these patients for signs of cancer is required. By imaging tiny and large lung nodules independently, we can improve the precision of diagnosis and treatment for these conditions.^{37,38} Discerning between large and tiny lung nodules is clinically significant for this reason. This might improve the prognosis for patients with lung nodules.

Compared to large lung nodules, the chances of small lung nodules being carcinogenic are comparatively low. Having said that, the chance that a small nodule is malignant remains small. For this reason, it is common practice to track the development of small nodules with repeated CT scans.

Conclusions

The dual-sampling network is an especially useful diagnostic tool in chest CT. It excels in comparison to single-sampling networks and has the potential to enhance clinical diagnostics. The examination of the chest with a dual-sampling CT is more reliable. This may lead to better diagnostic results and outcomes for patients. Compared to other methods, the technique that has been developed is more accurate. In addition, the dual-sampling network may be simply adapted to many medical imaging applications, and it has the potential to be advanced even further with the introduction of other machine learning methodologies. The recommended model, on the other hand, needs to be validated for its performance and monitoring before it can be used on a very intricate dataset. When compared to the earlier methods that have been reviewed in this work, the method that is currently being provided has the potential to improve the interpretability of lung abnormality diagnoses. Additionally, the created model is able to focus on linked regions more effectively and avoid distortions brought about by the generation of visual feedback to classification performance. When it comes to the region that contains the very small nodules, the precision is rather low.

Data Sharing Statement

The data used to support the findings of this study are included in the article.

Ethical Approval Statement

The study titled “Dual-Sampling Technique for Chest CT Scans Diagnosis” has received ethical approval from the Scientific Research Ethical Committee at Najran University, Kingdom of Saudi Arabia (Reference No.: 443-42-66321-DS). The principal investigator for this study is Dr. Khalaf Alshamrani from the College of Applied Medical Sciences at Najran University.

The aim of the study is to improve the diagnostic process and accuracy of lung abnormalities in imaging modalities techniques, specifically computed tomography (CT) images. The study focuses on distinguishing cancer cells from normal chest tissue using the dual-sampling network technique. Participants and their families will not be involved in the design or conduct of this study as a dataset of chest X-ray or CT scan images will be utilized for training and testing purposes.

The ethical approval was granted under the authority of Dr. Mater H. Mahnashi, Head of the Research Ethics Committee at Najran University.

This committee is accredited as a local committee by the National Committee for Bioethics at King Abdulaziz City for Science and Technology (HAPO-11-N-102).

Acknowledgments

The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work under the Distinguished Research Funding Program grant code (NU/DRP/MRC/12/30).

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work under the Distinguished Research Funding Program grant code (NU/DRP/MRC/12/30).

Disclosure

The authors declare no conflict of interest.

References

1. Irvin J, Rajpurkar P, Ko M, et al. Chexpert: a large chest radiograph dataset with uncertainty labels and expert comparison. *Proc AAAI Conf Artif Intell.* 2019;33:590–597. doi:.
2. Cruz-Roa AA, Ovalle JEA, Madabhushi A, Osorio FAG. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. In: *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer; 2013:403–410.
3. Zhang Q-S, Zhu S-C. Visual interpretability for deep learning: a survey. *Front Inf Technol Electron Eng.* 2018;19(1):27–39. doi:10.1631/FITEE.1700808
4. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A, “Learning deep features for discriminative localization,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, June 2016, pp. 2921–2929.
5. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D, “Grad-CAM: visual explanations from deep networks via gradient-based localization,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.
6. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM, “ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2097–2106.
7. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer; 2015:234–241.
8. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436–444. doi:10.1038/nature14539
9. He K, Zhang X, Ren S, Sun J, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, June 2016, pp. 770–778.
10. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, “Densely connected convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, July. 2017, pp. 4700–4708.
11. Nie D, Cao X, Gao Y, Wang L, Shen D. Estimating CT image from MRI data using 3D fully convolutional networks. In: *Deep Learning and Data Labeling for Medical Applications.* Cham, Switzerland: Springer; 2016:170–178.
12. LeCun Y, Boser B, Denker JS, et al. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* 1989;1(4):541–551. doi:10.1162/neco.1989.1.4.541
13. Pang T, Guo S, Zhang X, Zhao L. Automatic lung segmentation based on texture and deep features of HRCT images with interstitial lung disease. *BioMed Res Int.* 2019;2019:1–8.
14. Park B, Park H, Lee SM, Seo JB, Kim N. Lung segmentation on HRCT and volumetric CT for diffuse interstitial lung disease using deep convolutional neural networks. *J Digit Imag.* 2019;32(6):1019–1026. doi:10.1007/s10278-019-00254-8
15. Yasaka K, Akai H, Abe O, Kiryu S. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. *Radiology.* 2018;286(3):887–896. doi:10.1148/radiol.2017170706
16. Huang P, Park S, Yan R, et al. Added value of computer-aided CT image features for early lung cancer diagnosis with small pulmonary nodules: a matched case-control study. *Radiology.* 2018;286(1):286–295. doi:10.1148/radiol.2017162725
17. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Med.* 2019;25(6):954–961. doi:10.1038/s41591-019-0447-x
18. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology.* 2017;284(2):574–582. doi:10.1148/radiol.2017162326
19. He K, Gkioxari G, Dollár P, Girshick R, “Mask R-CNN,” in *Proc. IEEE Int. Conf. Comput. Vis.* Oct. 2017, pp. 2961–2969.
20. Van Horn G, Perona P. The devil is in the tails: fine-grained classification in the wild. *arXiv:1709.01450.* 2017. Available from: <http://arxiv.org/abs/1709.01450>. Accessed January 04, 2025.
21. Zhou B, Cui Q, Wei X-S, Chen Z-M. BBN: bilateral-branch network with cumulative learning for long-tailed visual recognition. *arXiv:1912.02413.* 2019. Available from: <http://arxiv.org/abs/1912.02413>. Accessed January 04, 2025.
22. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw.* 2018;106:249–259. doi:10.1016/j.neunet.2018.07.011
23. Shen L, Lin Z, Huang Q. Relay backpropagation for effective learning of deep convolutional neural networks. In: *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer; 2016:467–482.
24. He H, Garcia EA. Learning from imbalanced data. *IEEE Trans Knowl Data Eng.* 2009;21(9):1263–1284. doi:10.1109/TKDE.2008.239
25. Japkowicz N, Stephen S. The class imbalance problem: a systematic study. *Intell Data Anal.* 2002;6(5):429–449. doi:10.3233/IDA-2002-6504
26. Cui Y, Jia M, Lin T-Y, Song Y, Belongie S, “Class-balanced loss based on effective number of samples,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, June 2019, pp. 9268–9277.
27. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* 2002;16:321–357. doi:10.1613/jair.953
28. Huang B, Zhu Y, Wang Z, Fang Z. Imbalanced data classification algorithm based on clustering and SVM. *J. Circuits Syst. Comput.* 2020.
29. H. L., He H, Wen Y. Dynamic particle swarm optimization and K-means clustering algorithm for image segmentation. *Optik.* 2015;126(24):4817–4822. doi:10.1016/j.ijleo.2015.09.127
30. Khrissi L, Akkad NE, Satori H, Satori K. Image Segmentation Based on K-means and Genetic Algorithms. *Embed Syst Artif Intell Fz.* 2020;489–497.
31. Tsai C-F, Lin W-C, Hu Y-H, Yao G-T. Under-sampling class imbalanced datasets by combining clustering analysis and instance selection. *Inf Sci.* 2019;477:47–54. doi:10.1016/j.ins.2018.10.029
32. Shen J, Jiachao W, Man X, Gan D, Bang A, Liu F. A hybrid method to predict postoperative survival of lung cancer using improved SMOTE and adaptive SVM. *Comput Math Methods Med.* 2021;2021:1–15. doi:10.1155/2021/2213194
33. Kuenzi BM, Park J, Fong SH, et al. Predicting drug response and synergy using a deep learning model of human cancer cells. *Cancer Cell.* 2020;38(5):672–684. doi:10.1016/j.ccell.2020.09.014
34. Shao J, Feng J, Li J, et al. Novel tools for early diagnosis and precision treatment based on artificial intelligence. *Chin Med J Pulmonary Critical Care Med.* 2023;1(3):148–160. doi:10.1016/j.pccm.2023.05.001

35. Reay WR, Geaghan MP, Agee M, et al. The genetic architecture of pneumonia susceptibility implicates mucin biology and a relationship with psychiatric illness. *Nat Commun.* 2022;13(3756). doi:10.1038/s41467-022-31473-3
36. Passaro A, Al Bakir M, Hamilton EG, et al. Cancer biomarkers: emerging trends and clinical implications for personalized treatment. *Cell.* 2024;187(7):1617–1635. doi:10.1016/j.cell.2024.02.041
37. Maci E, Comito F, Frezza AM, et al. Lung nodule and functional changes in smokers after smoking cessation short-term treatment. *Cancer Invest.* 2014;32(8):388–393. doi:10.3109/07357907.2014.919308
38. Xuanzhuang LU, Qiu Q, Yang C, et al. Low-dose CT lung cancer screening results and high-risk factors for women in Guangzhou. *Zhongguo Fei Ai Za Zhi.* 2024;27(5):345–358. Chinese. doi:10.3779/j.issn.1009-3419.2024.101.14

Journal of Multidisciplinary Healthcare

Publish your work in this journal

The Journal of Multidisciplinary Healthcare is an international, peer-reviewed open-access journal that aims to represent and publish research in healthcare areas delivered by practitioners of different disciplines. This includes studies and reviews conducted by multidisciplinary teams as well as research which evaluates the results or conduct of such teams or healthcare processes in general. The journal covers a very wide range of areas and welcomes submissions from practitioners at all levels, from all over the world. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/journal-of-multidisciplinary-healthcare-journal>

Dovepress
Taylor & Francis Group