

*Lung cancer is the fastest-growing and most dangerous type of cancer worldwide. It ranks first among cancer diseases in the number of deaths, and diagnosing it at late stages makes treatment more difficult. Artificial intelligence has played an essential role in the medical field in general, and early diagnosis of diseases and analyzing medical images in particular, as it can reduce human errors that may occur with the medical expert in medical image analysis. In this study, a hybrid framework is proposed between deep learning using the proposed convolutional neural network and multi-criteria decision-making techniques in order to reach an effective and accurate classification model for lung cancer diagnosis and select the best methodology to solve the problem of class imbalance datasets, which is a general problem in medical data that causes problems and errors in prediction. The IQ-OTHNCCD dataset that has a class imbalance was used. Three class balancing techniques were used separately and the data from each one enters the proposed convolutional neural network for feature extraction and classification. Then the Fuzzy-Weighted Zero-Inconsistency algorithm and VIKOR were used to make the ranking for the best classification approach and determine the best technique to balance the classes. This contributed to increasing the efficiency of the classification, where the best model got an accuracy of 99.27 %, sensitivity of 99.33 %, specificity of 99 %, precision of 98.67 % and F1-score of 99 %. This study can be applied to any data that suffers from the class imbalance problem to find the best technique that gives the highest classification accuracy*

**Keywords:** lung cancer, deep learning, multi-criteria decision making (MCDM), class imbalance

# HYBRID SELECTION FRAMEWORK FOR CLASS BALANCING APPROACHES BASED ON INTEGRATED CNN AND DECISION MAKING TECHNIQUES FOR LUNG CANCER DIAGNOSIS

**Mustafa Mohammed Jassim**

*Corresponding author*

MSc, Researcher

Informatics Institute for Postgraduate Studies (IIPS)

Iraqi Commission for Computers and Informatics (ICCI)

Al-Nidhal str., 062, Baghdad, Iraq, 10069

Department of Medical Instruments Engineering Techniques

Al-Farahidi University

Al-Jadriah str., 098, Baghdad, Iraq, 10070

E-mail: ms202010619@iips.icci.edu.iq

**Mustafa Musa Jaber**

PhD, Lecturer

Department of Medical Instruments Engineering Techniques

Dijlah University College

Al-Dora str., 100, Baghdad, Iraq, 10021

Institute of Informatics and Computing in Energy

Universiti Tenaga Nasional

Jalan Ikram-Uniten, Selangor, Malaysia, 43000

Received date 15.06.2022

Accepted date 23.08.2022

Published date 31.08.2022

**How to Cite:** Jassim, M. M., Jaber, M. M. (2022). Hybrid selection framework for class balancing approaches based on integrated CNN and decision making techniques for lung cancer diagnosis. *Eastern-European Journal of Enterprise Technologies*, 4 (9 (118)), 69–76. doi: <https://doi.org/10.15587/1729-4061.2022.263644>

## 1. Introduction

According to the latest global cancer statistics in 185 countries and for 36 cancer types from the American cancer society, there are an estimated 19.3 million new cancer diagnoses and 10 million deaths globally [1]. Lung cancer ranked second in terms of the number of patients, accounting for around 11.4 % of all cancer cases with an estimated 2.2 million lung cancer cases. Lung cancer is the leading cause of cancer-related mortality, accounting for 18 % of all cancer-related deaths. Smoking is a leading cause of lung cancer; in some nations, smoking rates have peaked or are rising. This predicts that lung cancer rates will continue to rise for the foreseeable future [1]. It has been stated that the patient's chances of living a long life rise if cancer is detected early, diagnosed, and treated well [2]. Evaluating medical data and diagnosing illnesses need a medical specialist, and due to the intricacy of

medical imaging, experts' opinions frequently conflict when analyzing medical images. In the medical area, artificial intelligence has played a crucial role. In recent years, machine learning and deep learning algorithms have been used for the analysis and processing of medical pictures and the diagnosis of illnesses, since they give innovative solutions for medical applications [3]. Developing a prediction system that provides accurate diagnosis is not easy, and research is still underway in this area. This study proposed a framework based on multi-criteria decision-making, which is an extension of decision theory that encompasses all complicated multi-attribute decision-making problems by evaluating alternatives based on distinct, competing criteria and merging them into a single overall evaluation [4]. It is used in this study to solve the problem of selecting better class balancing approaches to the problem of class imbalance in lung cancer data to increase the efficiency of classification, as the class imbalance problem

is one of the biggest and most common problems, which increases the false positive rate and false negative rate. It causes a wrong diagnosis of the disease and therefore the patient may be given unnecessary and potentially dangerous drugs, the misprediction due to data bias towards the larger class. It shows a bias towards the majority group in the classification and in extreme cases the minority group is ignored [5]. The main motivation for lung cancer diagnosis using deep learning algorithms is the difficulty of diagnosing by radiologists, it is a time-consuming, costly, tedious job, and error-prone task, the screening process requires very high concentration and skills owing to factors such as low contrast variation and heterogeneity [6]. This is the general motive of this study, but the specific motive is to propose a hybrid selection framework for finding the best technique to deal with the problem of the class imbalanced datasets to obtain a high-performance deep learning classification model. This scientific topic concludes the best methodology for dealing with class imbalance data using the proposed model to reach a model with high efficiency for predicting lung cancer. This work can be applied to any imbalanced data to choose the best data balancing technique.

---

## 2. Literature review and problem statement

---

The paper [7] used a convolutional neural network on the LIDC-IDRI dataset and employed a median filter and the Gaussian Process (GP) Regression to improve the images and data augmentation technique by rotation and noise increase to increase data samples. An accuracy of 92.31 % was achieved. But the paper [8] used the same dataset with synthetic minority over-sampling technique class balancing technique and achieved an accuracy of 91 %, meaning that the result is different with changes in the class balancing methods. Thus, there is no single method that is the best with different datasets. The work [9] used AlexNet with the Taguchi method for classification on SPIE-AAPM data and the data augmentation technique using a data image generator where 99 % accuracy was achieved. The work [10] used IQ-OTH/NCCD data, which suffers from class imbalance problems and few benign tumor samples. CNN was used for classification without using any class balancing method and achieved 93.55 % accuracy, 95.71 % sensitivity, and 95 % specificity. The paper [11] used the same with the SVM classifier but the lowest class was neglected and was not included in the training. An accuracy of 89.88 % was achieved. In [12], a deep neural network was used on the private dataset, and the SMOTE technique was used to address the class imbalance problem. A sensitivity of 70.27 % and a specificity of 64.23 % were achieved. The paper [13] used ANNs for classification and the dataset from the «National Cancer Institute Data Access System», and the data augmentation techniques are used by flipping and rotation. An accuracy of 90.29 % was achieved. The work [14] used DFCNet for classification and data augmentation techniques for improving the classification performance on the LIDC-IDRI dataset with an accuracy of 86 %. The paper [15] used CNN with R2MNet architecture for classification with the LUNA16 dataset and data augmentation technology by scaling, flip, and rotation and obtained an accuracy of 94.74 %. By studying the relevant papers, it was found that there is no static approach for each dataset to address the class imbalance problem, which would be the best to increase the classification efficiency. As lung cancer datasets often lack samples in cases of benign

tumors [16], there is an urgent need to find a decision-making methodology to determine the better class balancer that gives the best results to increase the efficiency of the prediction model.

---

## 3. The aim and objectives of the study

---

The aim of the study is to propose a framework to overcome the problem of choosing a method to deal with class imbalance in lung cancer datasets. This will make it possible to choose the best approach that gives the highest efficiency of the diagnostic model.

To achieve the aim, the following objectives must be accomplished:

- to balance and preprocess the dataset using 3 class balancing techniques;
- to classify the data with each balancing technique;
- to select and assign weights for criteria;
- to build a decision matrix and select the best class balancing technique by ranking the 3 classification models based on each technique to determine the best approach to obtaining the highest-accurate classification model for lung cancer diagnosis by CT scan using imbalanced data.

---

## 4. Materials and methods

---

### 4.1. Methodology of research

To develop a framework for selecting the best classification approach among the three approaches based on different techniques to address the class imbalance problem, the methodology has been proposed, which is divided into three stages. In the first stage, the IQ-OTH/NCCD dataset was selected, which consisted of 1097 samples of chest CT scan images for lung cancer divided into 3 classes (benign, malignant, and normal) [6]. Then data pre-processing was performed (image resizing, image filtering, image normalization) and after that the data has been entered separately into three class balancing techniques (SMOTE, class-weighted approach, and data augmentation). Each of these techniques produces different balanced data. The second stage is where features are extracted and classified using the proposed convolutional neural network architecture separately for the data obtained from each balancing technique. In this way, three classification models were made. In the third stage, multi-criteria decision-making is used in this study to determine which data balancing method is best to improve the classification efficiency of unbalanced data with the stability of the other methods used. Firstly, a decision matrix must be built. It is essential in the decision-making process and consists of the weights of criteria on the  $x$ -axis and alternatives on the  $y$ -axis. The fuzzy-weighted zero-inconsistency was used for giving weights to criteria depending on expert's opinions. VIKOR is used for ranking the alternatives in the decision matrix represented by the classification results based on each class balancing technique. Thus, the balancing methods are evaluated in terms of their effect on the classification results from best to least effect on the results. Fig. 1 shows the stages of the proposed methodology. Python was used with TensorFlow and Keras libraries [17] mainly to train the model using a convolutional neural network using a Core i7-10870H processor, Nvidia GeForce RTX 3070 GPU, and Python was also used to implement decision-making methods.

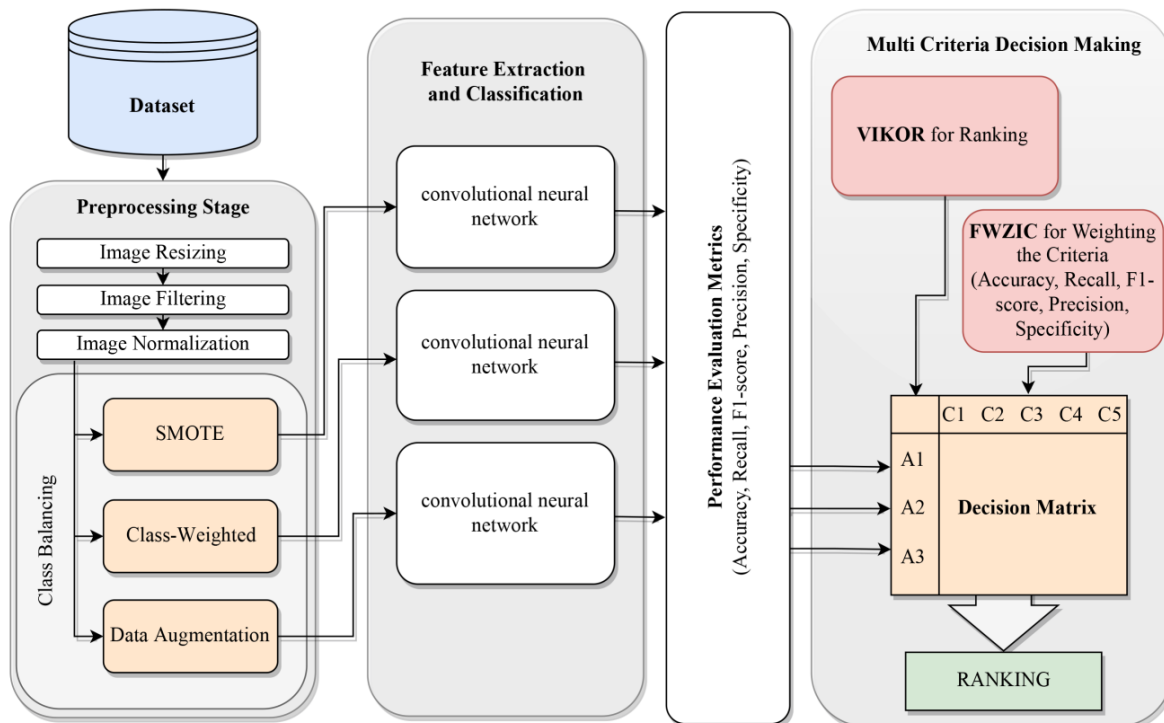


Fig. 1. Proposed methodology

This methodology is applicable to any class imbalanced dataset, so that class balancing techniques are evaluated based on the classification results for each technique, and therefore the approach that received the best score can be chosen.

**4. 2. Proposed convolutional neural network architecture**

The data generated from each balancing technique is entered separately to the convolutional neural network algorithm [18] for the purpose of feature extraction and classification by sequential CNN architecture as follows:

1. 2D convolutional layer with 64 filters of size (3×3) and ReLU activation function.
2. Max-Pooling layer with pooling size (2×2) and 1 stride value.
3. 2D-Convolutional layer with 64 filters of size (3×3) and RelU activation function.
4. Max-Pooling layer with pooling size (2×2) and 1 stride value.
5. Flatten layer.
6. Dense layer.
7. Dense layer with softmax activation function.

As shown in Fig. 2.

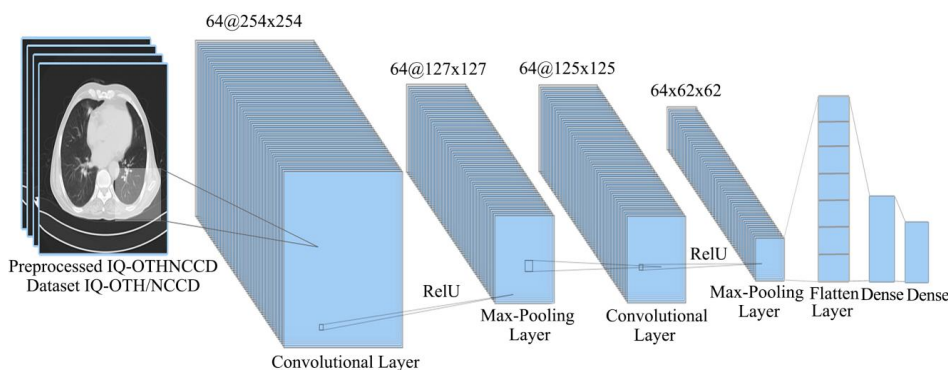


Fig. 2. Proposed convolutional neural network architecture

75 % of the data were used for training, 25 % of the data for testing with sparse categorical cross entropy loss function and Adam optimizer [19] and using batch size=8 and epochs=10.

**5. Results of the proposed selection framework**

**5. 1. Results of pre-processing and data balancing stage**

In the preprocessing stage, several steps were taken. In the first step, the images were scaled, and all of them became 256×256 in size after were of different large sizes as shown in Fig. 3, b. In the second step, the images were filtered using a Gaussian Blur filter with a 5×5 kernel [20] as shown in Fig. 3, c, and then the data images were normalized.

The data used suffered from class imbalance, so in the last step, the data that was processed in the previous steps was entered into three class-balancing techniques separately for each one. The first technique applied was SMOTE Oversampling [21]. The results of the implementation are shown in Fig. 4.

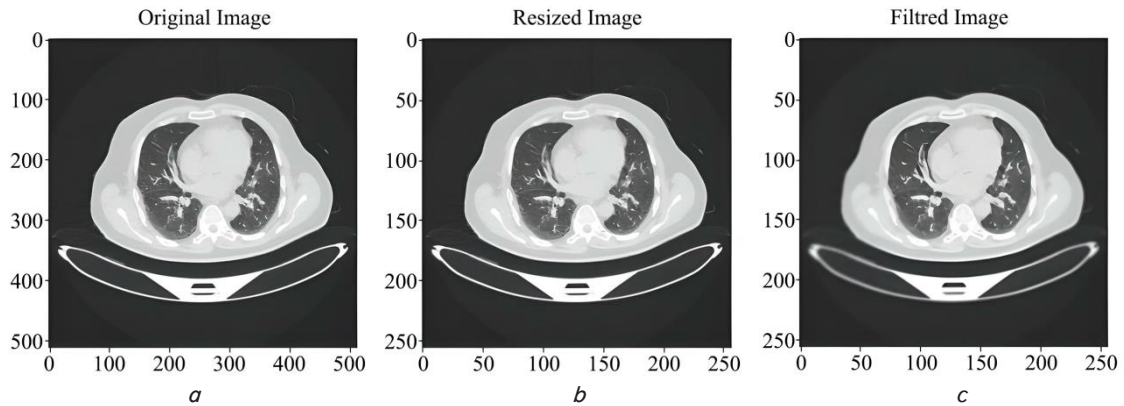


Fig. 3. Sample of data images: *a* – original image; *b* – resized image; *c* – filtered image

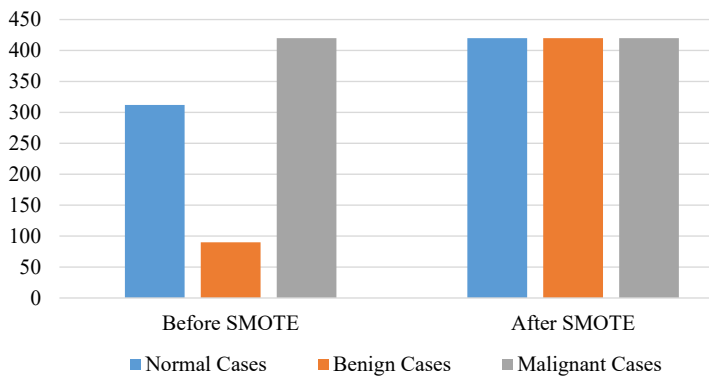


Fig. 4. Classes of data before and after the synthetic minority over-sampling technique

The second technique is the use of Class-Weighted Approaches [22] by assigning different weights to each class, where the weights of the class became as follows: Benign: 3.0444, Malignant: 0.6523, Normal: 0.8782. The third technique is data augmentation using the ImageDataGenerator function by generating new data making a horizontal and vertical flip of the original images [23].

**5. 2. Classification results**

The classification results are divided into three parts according to the type of class balancing technique that the data is subject to, which are as follows:

1. Classification based on SMOTE.

The classification results based on the data that were balanced using the SMOTE technique are shown in Table 1 and demonstrate the accuracy during model training in Fig. 5, *a* and the loss during model training in Fig. 5, *b*.

2. Classification based on class-weighted approach.

The classification results based on the data that were balanced using the class-weighted approach are shown in Table 2 and demonstrate the accuracy during model training in Fig. 6, *a* and the loss during model training in Fig. 6, *b*.

3. Classification based on data augmentation.

The classification results based on the data that were balanced using the data augmentation techniques are shown in Table 3 and demonstrate the accuracy during model training in Fig. 7, *a* and the loss during model training in Fig. 7, *b*.

The results of the three classification models are evaluated in the next step to determine the best model in terms of performance.

Table 1

Results of classification based on SMOTE

Accuracy	Sensitivity	Specificity	Precision	F1-score
0.9927	0.9933	0.9900	0.9867	0.9900

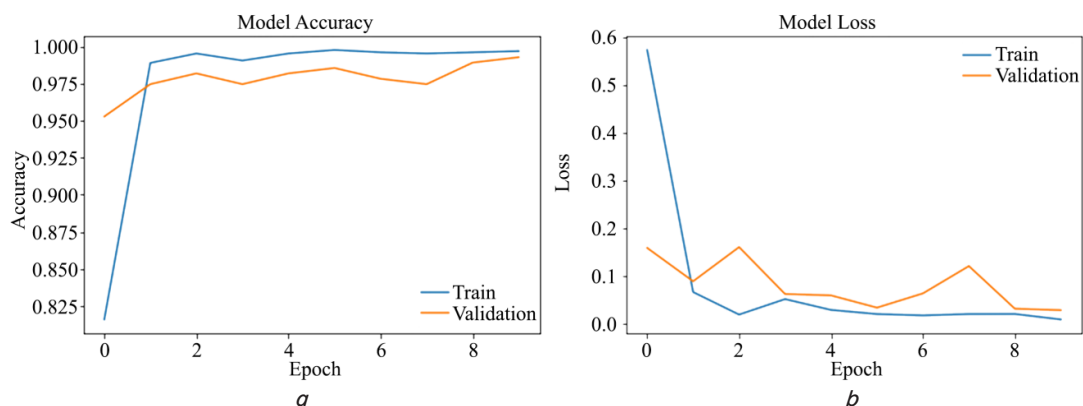


Fig. 5. Classification results based on synthetic minority over-sampling technique during model training: *a* – accuracy; *b* – loss

Table 2

Results of classification based on class-weighted approach

Accuracy	Sensitivity	Specificity	Precision	F1-score
0.9818	0.9933	0.9698	0.95666	0.97333

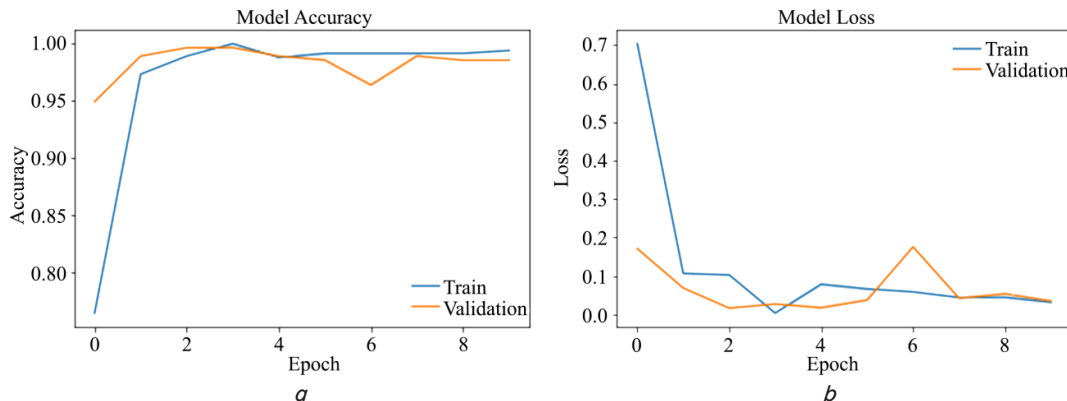


Fig. 6. Classification results based on class-weighted approach during model training: *a* – accuracy; *b* – loss

Table 3

Results of classification based on data augmentation techniques

Accuracy	Sensitivity	Specificity	Precision	F1-score
0.9636	0.94666	0.9500	0.953333	0.95666

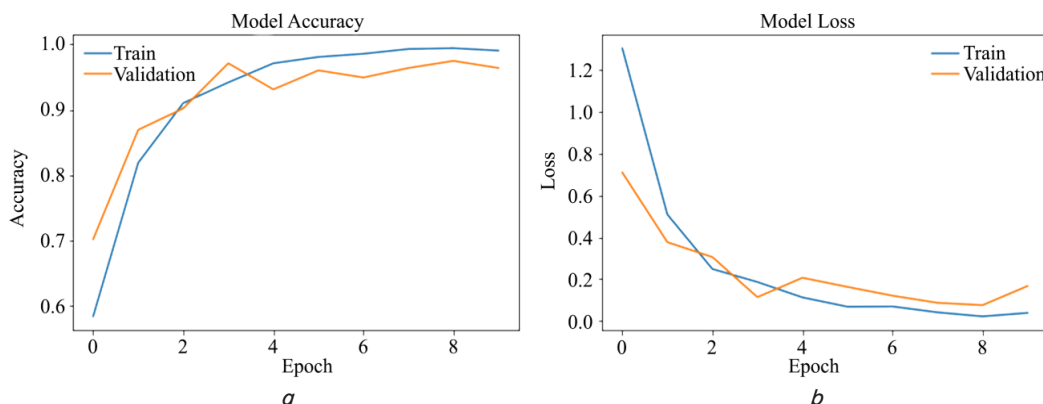


Fig. 7. Classification results based on data augmentation technique during model training: *a* – accuracy; *b* – loss

**5. 3. Results of weighting and selection of criteria**

The FWZIC approach was used to give weights to the criteria that will be used for building a decision matrix. This methodology includes five stages, the results of which were as follows:

1) the criteria by which the weights will be calculated have been selected, which are measures of classification efficiency (Accuracy, Sensitivity, Specificity, Precision, F1-score) [24];

2) six experts were selected based on their field of specialization and published research papers in the field of machine learning and data mining, and they are highly cited. A questionnaire form was prepared and their views were taken on the importance of each of the criteria for evaluating the efficiency of classification. The questionnaire form was prepared based on the five-point Likert scale, and their views were taken on the importance of each of the criteria of the

rating competency assessment scales. Then the opinions of the experts that were filled out in the form were converted into what is sufficient on the numerical scale;

3) the expert decision matrix (EDM) was built containing expert opinions on the importance of each criterion as shown in Table 4;

4) the fuzzy membership function is applied on the expert decision matrix generated from the previous step and the triangular fuzzy number was used, the results are shown in Table 5;

5) in the final step, the final values of the weights of the evaluation criteria were calculated based on fuzzification data generated from the previous step and the equations were applied to them, and in the end, defuzzification is done to get the final weight. The results of weighting the selected criteria (Accuracy, Sensitivity, Specificity, Precision, F1-score) are shown in Table 6.

Table 4

Results of expert opinions on the importance of classification performance metrics (expert decision matrix)

Expert \ Criteria	Accuracy	Sensitivity	Specificity	Precision	F1-score
Expert 1	4	3	4	3	4
Expert 2	4	3	3	3	4
Expert 3	5	5	5	5	4
Expert 4	4	5	4	5	3
Expert 5	5	5	4	5	4
Expert 6	5	3	3	3	3

Table 8

Ranking of the classification based on different class balancing techniques

Alternative	Q-value	Order
Classification Based on SMOTE technique	0.000	1
Classification Based on Class-Weighted Approaches	0.633	2
Classification Based on Data Augmentation technique	1.000	3

Table 5

Results of implementing the fuzzy membership function on the EDM

Accuracy			Sensitivity			Specificity			Precision			F1-score		
0.12	0.23	0.43	0.07	0.15	0.36	0.12	0.23	0.43	0.07	0.15	0.36	0.12	0.23	0.43
0.12	0.25	0.47	0.07	0.17	0.39	0.07	0.17	0.39	0.07	0.17	0.39	0.12	0.25	0.47
0.15	0.21	0.29	0.15	0.21	0.29	0.15	0.21	0.29	0.15	0.21	0.29	0.10	0.17	0.26
0.11	0.20	0.32	0.16	0.24	0.36	0.11	0.20	0.32	0.16	0.24	0.36	0.07	0.13	0.27
0.16	0.21	0.31	0.16	0.21	0.31	0.10	0.18	0.28	0.16	0.21	0.31	0.10	0.18	0.28
0.19	0.31	0.51	0.08	0.17	0.38	0.08	0.17	0.38	0.08	0.17	0.38	0.08	0.17	0.38

Table 6

Final results for weights of criteria of classification performance metrics

Criteria	Accuracy	Sensitivity	Specificity	Precision	F1-score
Final weights	0.2278	0.1952	0.1926	0.1952	0.1893

After the end of this stage, the necessary alternatives and weights were provided to build the decision matrix.

**5. 4. Results of building a decision matrix and ranking the alternatives**

After defining and extracting the weights of the criteria, a decision matrix can be built by placing the alternatives on the y-axis and the weights of the standards on the x-axis, as shown in Table 7.

Table 7

Decision matrix

Alternatives \ Weights	Accuracy	Sensitivity	Specificity	Precision	F1-score
		0.2278	0.1952	0.1926	0.1952
SMOTE	0.9927	0.9933	0.9900	0.9867	0.9900
Class-Weighted	0.9818	0.9833	0.9698	0.9567	0.9733
Data Augmentation	0.9636	0.9467	0.9500	0.9533	0.9567

Then the VIKOR algorithm [25] is applied to the decision matrix to rank the three alternatives represented by the results of classifying the CT images of lungs using different methods to balance the data classes. The results of the Q-value and order for each alternative are shown in Table 8.

The validity of the ranking results, according to the VIKOR method, is verified through acceptable advantage by verifying the validity of the mathematical relationship (1) where  $Q(a'')$  is the Q-value of the alternative that has an order of 2 and  $(a')$  is the Q-value of the alternative that has an order of 1 and  $J$  is the number of alternatives:

$$Q(a'') - Q(a') \geq \frac{1}{J-1}. \tag{1}$$

When applying the mathematical relationship,  $0.633 > 0.5$ , which means that the results are valid.

**6. Discussion of the experimental results of classification and decision making**

From the results of this study, it was found that after balancing the data in different ways, the classification based on SMOTE achieved the results shown in Table 1, the classification based on Class-Weighted Approach achieved the results shown in Table 2 and the classification based on Data Augmentation Technique achieved the results shown in Table 3. This shows that the classification results change a lot according to the balancing technique used, and for choosing the best approach to achieve the best performance, multi-criteria decision-making techniques were used. First, weights were assigned to each performance evaluation metric using the FWZIC method shown in Table 4. This means that the most important metric is accuracy followed by sensitivity and precision in the same importance, then specificity, and finally F1-score.

In the last step, the decision matrix was built and the VIKOR method was implemented to make the decision to choose the most efficient approach to classification. The classification based on SMOTE ranked first, followed by the classification based on Class-Weighted Approaches in the second place, and finally the classification based on Data Augmentation technique in the third place.

From the results of the evaluation of the three approaches, the proposed classification model based on SMOTE technology was selected as the most efficient in terms of prediction accuracy for the data set used. Table 9 shows a comparison of the results obtained for the model and the results of studies that used the same dataset.

According to the results shown in Table 9, the approach chosen through the proposed framework is the most efficient and its results are the highest for all performance metrics that were obtained in the previous studies.

Table 9

Benchmark of previous studies that used the same dataset and the current research

References	Method	Accuracy	Precision	Specificity	Sensitivity	F1-score
[10]	AlexNet-CNN	0.935	0.971	0.950	0.957	0.964
[11]	SVM	0.898	0.985	0.975	0.971	0.978
[26]	GoogLeNet-CNN	0.943	0.935	0.937	0.950	0.943
[27]	VGG16	0.46	0.31	–	0.34	–
	MobileNet	0.51	0.42	–	0.39	–
	DenseNet	0.53	0.43	–	0.42	–
	VGG19	0.56	0.43	–	0.42	–
	ResNet	0.48	0.41	–	0.37	–
Selected approach	CNN-Based SMOTE	0.992	0.986	0.990	0.993	0.9900

The limitation of this work is the lack of a lung cancer dataset due to the lack of clinical information, increasing the data set leads to an increase in the efficiency of diagnosis. The disadvantage of this research is the static of the convolutional neural network architecture and the decision-making process is not involved in determining the best architecture to improve efficiency, and the decision-making process is restricted to category balancing techniques. This work can be developed using the FWZIC method with a larger number of machine learning experts to obtain more accurate weights for the criteria, and the decision-making process can be further developed to include selecting the best CNN architecture such as ResNet, VGG16, AlexNet, GoogLeNet, etc., which

gives the highest classification efficiency. Thus, the decision-making process is in two locations within the framework and the under-sampling class balancing techniques can be used, which provides a better experience.

## 7. Conclusions

1. A framework has been developed to systematically identify the best classification approach of 3 approaches based on different class balancing techniques.

2. The class balancing techniques to improve the efficiency of the classification model for the lung cancer dataset were arranged based on the proposed CNN, where SMOTE technique got place one, the class-weighted approach was in the second place, and the data augmentation technique took the third place.

3. Weights were assigned using the FWZIC technology for each of the classification efficiency measures to determine the importance of each scale, where the highest importance was for accuracy, then sensitivity and precision got equal importance, devouring the specificity and finally the F1-score.

4. The most accurate classification model was the SMOTE technology-based model with an accuracy of 0.9927, sensitivity of 0.9933, specificity of 0.9900, precision of 0.9867, and F1-score of 0.9900, and this is the highest classification efficiency among the results of studies that used the dataset.

## Conflict of interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in the paper.

## References

- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71 (3), 209–249. doi: <https://doi.org/10.3322/caac.21660>
- Begum, S., Sarkar, R., Chakraborty, D., Maulik, U. (2020). Identification of Biomarker on Biological and Gene Expression data using Fuzzy Preference Based Rough Set. *Journal of Intelligent Systems*, 30 (1), 130–141. doi: <https://doi.org/10.1515/jisys-2019-0034>
- Razzak, M. I., Naz, S., Zaib, A. (2017). Deep Learning for Medical Image Processing: Overview, Challenges and the Future. *Classification in BioApps*, 323–350. doi: [https://doi.org/10.1007/978-3-319-65981-7\\_12](https://doi.org/10.1007/978-3-319-65981-7_12)
- Albahri, O. S., Albahri, A. S., Zaidan, A. A., Zaidan, B. B., Alsalem, M. A., Mohsin, A. H. et. al. (2019). Fault-Tolerant mHealth Framework in the Context of IoT-Based Real-Time Wearable Health Data Sensors. *IEEE Access*, 7, 50052–50080. doi: <https://doi.org/10.1109/access.2019.2910411>
- Johnson, J. M., Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6 (1). doi: <https://doi.org/10.1186/s40537-019-0192-5>
- Jassim, M. M., Jaber, M. M. (2022). Systematic review for lung cancer detection and lung nodule classification: Taxonomy, challenges, and recommendation future works. *Journal of Intelligent Systems*, 31 (1), 944–964. doi: <https://doi.org/10.1515/jisys-2022-0062>
- Hussein, S., Gillies, R., Cao, K., Song, Q., Bagci, U. (2017). TumorNet: Lung nodule characterization using multi-view Convolutional Neural Network with Gaussian Process. 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). doi: <https://doi.org/10.1109/isbi.2017.7950686>
- Apostolopoulos, I. D. (2020). Experimenting with Convolutional Neural Network architectures for the automatic characterization of Solitary Pulmonary Nodules' malignancy rating. arXiv. doi: <https://doi.org/10.48550/arXiv.2003.06801>
- Lin, C.-H., Lin, C.-J., Li, Y.-C., Wang, S.-H. (2021). Using Generative Adversarial Networks and Parameter Optimization of Convolutional Neural Networks for Lung Tumor Classification. *Applied Sciences*, 11 (2), 480. doi: <https://doi.org/10.3390/app11020480>

10. Al-Yasriy, H. F., AL-Huseiny, M. S., Mohsen, F. Y., Khalil, E. A., Hassan, Z. S. (2020). Diagnosis of Lung Cancer Based on CT Scans Using CNN. IOP Conference Series: Materials Science and Engineering, 928 (2), 022035. doi: <https://doi.org/10.1088/1757-899x/928/2/022035>
11. Kareem, H. F., AL-Huseiny, M. S., Mohsen, F., Khalil, E., Hassan, Z. (2021). Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset. Indonesian Journal of Electrical Engineering and Computer Science, 21 (3), 1731. doi: <https://doi.org/10.11591/ijeecs.v21.i3.pp1731-1738>
12. Li, J., Tao, Y., Cai, T. (2021). Predicting Lung Cancers Using Epidemiological Data: A Generative-Discriminative Framework. IEEE/CAA Journal of Automatica Sinica, 8 (5), 1067–1078. doi: <https://doi.org/10.1109/jas.2021.1003910>
13. Kaur, L., Sharma, M., Dharwal, R., Bakshi, A. (2018). Lung Cancer Detection Using CT Scan with Artificial Neural Network. 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE). doi: <https://doi.org/10.1109/icrieece44171.2018.9009244>
14. Masood, A., Sheng, B., Li, P., Hou, X., Wei, X., Qin, J., Feng, D. (2018). Computer-Assisted Decision Support System in Pulmonary Cancer detection and stage classification on CT images. Journal of Biomedical Informatics, 79, 117–128. doi: <https://doi.org/10.1016/j.jbi.2018.01.005>
15. Zheng, S., Shen, Z., Pei, C., Ding, W., Lin, H., Zheng, J. et. al. (2021). Interpretative computer-aided lung cancer diagnosis: From radiology analysis to malignancy evaluation. Computer Methods and Programs in Biomedicine, 210, 106363. doi: <https://doi.org/10.1016/j.cmpb.2021.106363>
16. Bansal, G., Chamola, V., Narang, P., Kumar, S., Raman, S. (2020). Deep3DScan: Deep residual network and morphological descriptor based framework for lung cancer classification and 3D segmentation. IET Image Processing, 14 (7), 1240–1247. doi: <https://doi.org/10.1049/iet-ipr.2019.1164>
17. Raschka, S., Patterson, J., Nolet, C. (2020). Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. Information, 11 (4), 193. doi: <https://doi.org/10.3390/info11040193>
18. Yamashita, R., Nishio, M., Do, R. K. G., Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. Insights into Imaging, 9 (4), 611–629. doi: <https://doi.org/10.1007/s13244-018-0639-9>
19. Kingma, D. P., Ba, J. (2014). Adam: A method for stochastic optimization. arXiv. doi: <https://doi.org/10.48550/arXiv.1412.6980>
20. Hu, X., Huang, C., Feng, R., Zhou, L., Zheng, L. (2021). Blind image blurring by Gaussian filtering extreme channels prior. Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering. doi: <https://doi.org/10.1145/3501409.3501510>
21. Kovács, B., Tinya, E., Németh, C., Ódor, P. (2020). Unfolding the effects of different forestry treatments on microclimate in oak forests: results of a 4-yr experiment. Ecological Applications, 30 (2). doi: <https://doi.org/10.1002/eap.2043>
22. Batista, G. E. A. P. A., Prati, R. C., Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorations Newsletter, 6 (1), 20–29. doi: <https://doi.org/10.1145/1007730.1007735>
23. Gao, J. (2020). Data Augmentation in Solving Data Imbalance Problems. Degree Project in Computer Science and Engineering. Stockholm. Available at: <https://www.diva-portal.org/smash/get/diva2:1521110/FULLTEXT01.pdf>
24. Jiao, Y., Du, P. (2016). Performance measures in evaluating machine learning based bioinformatics predictors for classifications. Quantitative Biology, 4 (4), 320–330. doi: <https://doi.org/10.1007/s40484-016-0081-2>
25. Sayadi, M. K., Heydari, M., Shahanaghi, K. (2009). Extension of VIKOR method for decision making problem with interval numbers. Applied Mathematical Modelling, 33 (5), 2257–2262. doi: <https://doi.org/10.1016/j.apm.2008.06.002>
26. Al-Huseiny, M. S., Sajit, A. S. (2021). Transfer learning with GoogLeNet for detection of lung cancer. Indonesian Journal of Electrical Engineering and Computer Science, 22 (2), 1078. doi: <https://doi.org/10.11591/ijeecs.v22.i2.pp1078-1086>
27. Mohite, A. (2021). Application of Transfer Learning Technique for Detection and Classification of Lung Cancer using CT Images. International Journal of Scientific Research and Management, 9 (11), 621–634. doi: <https://doi.org/10.18535/ijstrm/v9i11.ec02>