IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 February 2025, accepted 1 April 2025, date of publication 16 May 2025, date of current version 1 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3570643

RESEARCH ARTICLE

A Novel Context-Aware Feature Pyramid Networks With Kolmogorov-Arnold Modeling and XAI Framework for Robust Lung Cancer Detection

YAJNASENI DASH[1], SUDHIR C. SARANGI[2], VINAYAK GUPTA[3], NAWEEN KUMAR[4], AND AJITH ABRAHAM[5,6]

[1]School of Artificial Intelligence, Bennett University, Greater Noida 201310, India
[2]Department of Pharmacology, All India Institute of Medical Science, New Delhi 110029, India
[3]Department of Computer Science and Engineering, ITS Engineering College, Greater Noida 201310, India
[4]School of Computer Science and Engineering Technology, Bennett University, Greater Noida 201310, India
[5]School of Artificial Intelligence, Sai University, Chennai 603104, India
[6]Research Center of the Artificial Intelligence Institute, Innopolis University, 420500 Innopolis, Russia

Corresponding authors: Yajnaseni Dash (yajnasenidash@gmail.com) and Ajith Abraham (abraham.ajith@gmail.com)

This work was supported by the Ministry of Economic Development of the Russian Federation (agreement No. 139-10-2025-034 dd. 19.06.2025, IGK 000000C313925P4D0002).

ABSTRACT The present work describes a new method for detecting lung cancer based on a Kolmogorov-Arnold Network (KAN) combined with a Context-Aware Feature Pyramid Network (FPN), denoted KAFPN and supported by XAI tools, including Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) for explanation purposes. The model is compiled based on an extended corpus containing 25,000 histopathological images from 15,000 lung tissue samples. The proposed KAFPN model outperformed the traditional models, such as Base CNN, Inception V3, VGG16, ResNet50, EfficientNet-B7, ResNet-101, and DenseNet-121, based on testing and validation accuracy of 99.63% and 99.60%, respectively. The integration of LIME and SHAP facilitated an advanced level of interpretability by providing detailed visual explanations of the model's predictive decisions, delineating the image regions that most significantly impacted the decision-making process. The proposed KAFPN framework is well-suited for image segmentation and object detection because it has multiple-scale feature responses. Stringent data augmentation and correct division of data sets make the model robust and efficient for real-world applications.

INDEX TERMS Computer-aided diagnosis (CAD), convolutional neural networks, image segmentation, Kolmogorov-Arnold network (KAN), lung cancer detection.

I. INTRODUCTION

Lung cancer remains the most lethal type of cancer with a high mortality rate attributed to the advanced stage that is primarily present at the time of diagnosis, leading to poor prognosis and intervention [1]. Early diagnosis is essential for better survival, but most lung cancers remain silent and, therefore, difficult to diagnose during the initial stages [2]. Outsourced diagnostic methods, including X-rays and CT scans, have the disadvantage of involving human interpretation that adds variability and mistakes [3]. Although accurate, these methods lack specificity and are prone to false positives, making early diagnosis even more challenging.

Nonetheless, the accuracy issue has been noted to be something that CNNs are very good at, but what is terrible is that they are referred to as 'black-box' models; that is, they do not have to explain why a given decision was made. This reduces credibility where it is undesirable, as in

The associate editor coordinating the review of this manuscript and approving it for publication was Sotirios Goudos.

medical applications where it is crucial to know why a particular decision has been made [4], [5], [6]. As a result of these challenges, there is an increased focus on a new field of study known as Explainable AI (XAI) that seeks to increase the amount of explainability in AI systems [7]. Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) models are the main Explainable AI (XAI) techniques that can help explain model predictions, enhancing trust while mitigating bias [8], [9]. Despite the progress made in the literature, the combination of XAI with lung cancer detection systems has limited solutions in terms of both impressive accuracy and interpretability.

The new AI system is being developed due to the lack of accuracy, interpretability, and clinical applicability of the existing methods and CNN-based models. Nevertheless, analogous CNNs demonstrate high accuracy for detecting lung cancer, but their poor interpretability hinders their application in medical practice [10]. Therefore, there is a significant opportunity for a new AI-based model that will be equally effective in performance and offer possibilities to interpret diagnostic results.

The suggested article presents a new method to resolve the lung cancer identification problem by designing a KAFPN model incorporated with the Context-Aware module. Kolmogorov-Arnold Network (KAN) utilizes mathematical transformations to update suggestive CNNs, increasing interpretability and accuracy regarding complex medical images [11], [12], [13], [14], [15]. This integration of context-aware mechanisms helps the model narrow to a region of interest in the image, improving the feature extractor's quality. The Feature Pyramid Network (FPN) component allows the model to combine features on multiple scales, improving the model's image segmentation and detection abilities [16], [17], [18]. Furthermore, the suggested work exploits LIME and SHAP to explain model decisions to overcome the 'black box' issue and enhance confidence in the system [19], [20], [21].

A. OBJECTIVES OF THE STUDY

Adopting a nuanced methodology, the article aims to achieve the following objectives:

- Develop a superior-performing KAFPN model to detect lung cancer with great sensitivity.
- Improve model interpretability through LIME and SHAP to present features meaningful to doctors and practitioners.
- Solve the most critical problem of model interpretability using artificial intelligence in lung cancer diagnosis.
- Integrating the proposed KAFPN framework with XAI methods can potentially transform the early detection of lung cancer since it generates a high-performance AI model and a properly interpretable model for practitioners.

The manuscript adopts a multifaceted approach to address the challenges of lung cancer detection using artificial intelligence, with a core emphasis on both performance and interpretability. The primary objective is to develop a Kolmogorov-Arnold Feature Pyramid Network (KAFPN) architecture that achieves high sensitivity and accuracy in classifying histopathological images of lung cancer. By leveraging the powerful feature extraction capabilities of CNN-based Feature Pyramid Networks (FPNs) and enhancing them with KAN-based nonlinear univariate transformations, the proposed model aims to capture rich multi-scale spatial and semantic information. This results in a more robust and generalizable classifier, essential for sensitive tasks like cancer detection, where false negatives can be critical. A key advancement in the proposed methodology lies in its integration with explainable AI (XAI) frameworks, specifically LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations). These tools are employed to decode the decision-making process of the model, highlighting which features or regions in the image are contributing to the prediction. This not only enhances the interpretability of the model but also aligns its outcomes with clinical reasoning, making the system more trustworthy and actionable for doctors and healthcare practitioners. Importantly, the study tackles one of the most pressing issues in medical AI: black-box model opacity. By combining the high-performance KAFPN architecture with transparent reasoning provided by XAI, the framework can support early-stage diagnosis while also ensuring that the predictions are understandable to non-technical medical professionals. This dual achievement in performance and interpretability makes the proposed system a significant advancement in applying AI for early lung cancer detection.

The article has been divided as follows, briefly introducing the issue presented in Section I. Section II surveys relevant studies and identifies the research gap. Section III introduces the suggested model and the various algorithms. The following Section IV discusses the results. Section V concludes the suggestive study.

II. LITERATURE SURVEY

Majaku et al. [22], taking cognizance of the importance of lung cancer detection, used CT scan images to develop a system that used watershed segmentation and SVM for classification. Their model was pretty advanced and comprised various data preprocessing steps, like using Gaussian and Median filters combined with Watershed to give an overall accuracy of 92%. The study was limited since there is scope for improvement in classification accuracy and better fundamentals to clear the interpretability of the model. Conducting an extensive review of the available literature on lung cancer detection and identification, Shailesh et al. [23] compared various CAD systems. They highlighted that the techniques used are usually based on the CNN network, showing their sophistication and importance. The highlighted range of accuracy of the different models is as low as 68.0%

TABLE 1. Summary of related works and their identified gaps.

Ref. No.	Author(s)	Description and Identified Gap
[22]	Majaku et al.	Used CT scans with watershed segmentation and SVM, achieving 92% accuracy. Lacked interpretability and scope for improved classification.
[23]	Shailesh et al.	Conducted a comparative study on various CAD systems. Emphasized CNN importance; noted variability in accuracy (68.0%–99.51%). Need for better generalization.
[24]	Togacar	Achieved 99.51% accuracy using a hybrid of AlexNet, LeNet, VGG-16 with feature optimization. Lacked discussion on interpretability.
[25]	Shakeel et al.	Proposed IDNN model with 96.2% accuracy and 0.974 precision. The aspect of model explainability was not addressed.
[26]	Hatuwal and Himal	Used ConvNet on histopathological images. Achieved 97.20% validation accuracy. Highlighted CNN effectiveness but lacked XAI applications.
[27]	Ahmed and Fatma	Tested VGG19 + SVM on LDCT images with 96.25% accuracy and 97.5% sensitivity. Needed a larger dataset.
[28]	Satvik and Somya	Implemented SmoothGrad and GradCAM achieving 97.5% accuracy. Focused on binary classification; comparison with counterparts showed improvement.
[29]	Hamida et al.	Used SegNet for pixel-wise segmentation across 3 datasets. Achieved 99.12% accuracy but lacked testing on noisy/augmented data and interpretability.
[30]	Talukder et al.	Proposed ensemble learning using SVM, LR, and MLP after transfer learning. Achieved 94.45% average accuracy. Lacked fine-grained interpretability.
[31]	Abtin et al.	Tested KimiaNet on TCGA slides with 96.38% accuracy (colorectal). This study did not focus on lung cancer.
[32]	Mehedi et al.	Made use of image transformation with 96.33% accuracy . But the study lacked in providing any aspect of interpretability.

and as high as 99.51%, given by Togacar [24], having a combination of AlexNet, LeNet and VGG-16 and specific feature optimization methods to configure these results. This highlights the importance of CNN as an architecture in image segmentation and the need for more sophisticated algorithms to be applied to the same for better generalization of results. Shakeel et al. [25] enhanced the CAD systems with the help of an improved deep neural network (IDNN), providing an accuracy of 96.2% on the dataset. Although the system fired well, there was a lower performance in Precision 0.974, and there was no discussion based on the explainability of the results. Hatuwal and Himal [26] implemented a similar kind of ConvNet to understand the efficacy of CAD on lung cancer detection using histopathological images. Their model had a training accuracy of 96.11% and a validation accuracy of 97.20%, showing the fairness of CNNs in lung cancer detection. Comparing various models on the low-dose computed tomography images, Ahemed and Fatma [27] showed the prominence of predefined architectures such as Alexnet, VGG16, and VGG19 on lung cancer detection. They concluded an accuracy of 96.25% and a sensitivity of 97.5% on a combination of VGG19 and SVM classifiers. The research demanded the use of a more extensive and better data set. Advanced models like SmoothGrad and GradCAM were used extensively by Satvik and Somya [28] to highlight their performance compared to their counterparts in lung cancer detection. Their implementation showed an output of 97.5% accuracy, showing an improvement in the field with the dataset focusing on binary classification. Hamida et al. [29] proposed a pixel-wise segmentation strategy, paired it with SegNet, and tested it on three different datasets. Their results achieved an accuracy of 99.12% but lacked in the testing on augmented and noisy data, proving uninterpretable in the decisions made. Using a hybrid model, Talukder et al. [30] proposed an ensemble feature extraction model for lung cancer classification. Their method encompassed transfer learning from known pre-trained models to extract the salient feature and the combination of SVM, LR, and MLP to create a voting classifier to achieve an average accuracy of 94.45% among all the classes. Refining denseness for cancer detection, Abtin et al. [31] made use of TCGA diagnostic slides to test the performance of KimiaNet as compared to others. Their findings revealed an accuracy of 96.38% on the colorectal dataset. Mehedi et al. [32] demonstrated the usage of two collective image transformation techniques, the two-dimensional Discrete Fourier transform (2D-DFT) and Single-level discrete two-dimensional wavelet transform (2D-DWT), to make their predictions. Their proposed model had an accuracy of 96.33%, which is advanced.

A. RESEARCH GAP

The literature mentioned above regarding lung cancer detection demonstrates several innovative developments in the field; however, it also exposes neglected areas in development that our proposed work fills. Majaku et al. [22] and Shakeel et al. [25] have reported high accuracy of this model, but the models are not very interpretable. They are discrete and have an inconsistent accuracy rate anywhere between studies [23]. Most of them are also not optimal when utilizing enhanced feature extraction or dealing with noisy or augmented data [29]. Secondly, although CNN-based models have proven very effective, everyone knows the main downside. There is very little that can be explained in clinical practice, which means that people do not trust systems that use artificial intelligence. Mel-Frequency Cepstral Coefficients (M-FCC systems) have identified the following limitations: Unfavorable classification accuracy, lack of transparent decision-making, and model vulnerability to adversarial perturbations. A summary of the proposed models has been discussed in Table 1. These shortcomings are incorporated in the proposed KAFPN model, incorporating KAN, FPN, and XAI tools (LIME and SHAP), making the classification model more accurate, reliable, and clinically relevant than present M-FCC systems. Table 2 shows the list of major abbreviations used most frequently throughout the article.

III. MATERIALS AND METHODS

A. DATASET

The suggested work uses the Lung and Colon Histopathological Images dataset, which is available on the Kaggle

TABLE 2. Abbreviation table.

Abbreviation	Description
CAD	Computer Aided Diagnostic
CAM	Context-aware Mechanism
CNN	Convolutional Neural Network
FPN	Feature Pyramid Network
KAFPN	Kolmogorov-Arnold Network-based Context Aware Feature Pyramid Network
KAN	Kolmogorov-Arnold Network
LIME	Local Interpretable Model-agnostic Explanations
M-FCC	Mel-Frequency Cepstral Coefficients
SHAP	Shapely Additive Explanations
XAI	Explainable AI

network. This vast dataset contains about 25,000 images, which are standardized to a size of 768 × 768 pixels. These high-resolution images allow for easier detection of features and thus facilitate better classification by the proposed model. The images that have been included in the dataset were all sourced from HIPAA-compliant sources, ensuring the validity of the images and their sources. The dataset is divided into five classes, each containing 5,000 images:

- Lung benign tissue
- Lung adenocarcinoma
- Lung squamous cell carcinoma
- Colon adenocarcinoma
- Colon benign tissue

For this research, only lung tissue images are utilized. This subset includes 15,000 images, distributed evenly among the three lung tissue classes: benign lung tissue, lung adenocarcinoma, and lung squamous cell carcinoma (Fig 1). Each of these classes consists of 5,000 images, providing a balanced dataset that is crucial for training the CNN model effectively.

B. ALGORITHMS USED

1) KANs

The major aspect of the proposed architecture is the Kolmogorov-Arnold Networks (KANs). KANs are generic to the Kolmogorov Superposition Theorem, which encompasses the framework for the representation of multivariate continuous functions with the superposition of simpler univariate functions and their summations. This theorem, developed by Andrey Kolmogorov and refined further by Vladimir Arnold, states that any multivariate continuous function can be decomposed into a finite number of univariate continuous functions under the summation operation. The theorem can be formally expressed as given under Eq. (1).

Kolmogorov Superposition Theorem:

For any continuous function

$$f : [0, 1]^n \to \mathbb{R}$$

There exist continuous functions $\phi_p(x)$ and $\psi_q(y)$ such that:

$$f(x_1, \ldots, x_n) = \sum_{p=1}^{2n+1} \phi_p\Big(\sum_{q=1}^{n} \psi_q(x_q)\Big) \tag{1}$$

where $(x_1, \ldots, x_n)$ is an n-dimensional input vector, ϕ_{ij} are continuous univariate functions, and ψ_i are continuous aggregation functions. This shows that any multivariate continuous function can be determined with the help of the underlying univariate continuous function as well as its submission used in conjunction with each other [33]. This decomposition from the above equations allows the KANs to increase the efficacy and effectiveness of model performance [12], [34]. It allows the complex components to become more manageable as compared to the original equations, thus allowing the model to make more accurate generalizations, while also helping them to simplify the learning of the extracted features. This allows the KAN network to be more precise and improves the performance as a whole.

The inclusion of Kolmogorov-Arnold Networks (KAN) in the proposed architecture is a significant contribution, particularly in enhancing feature extraction and improving model robustness. However, an explicit discussion on the advantages of KAN over simpler transformations, supported by both theoretical and experimental evidence, is necessary to establish its practical utility in histopathological data analysis. KAN leverages the Kolmogorov-Arnold representation theorem, which states that any multivariate continuous function can be decomposed into a finite sum of univariate continuous functions. This principle allows KAN to systematically transform input features into highly informative representations by applying nonlinear transformations to subsets of the feature space. Compared to traditional linear or simple nonlinear transformations, KAN captures complex interactions between features without explicitly relying on high-dimensional mappings, thus reducing computational complexity and overfitting.

2) CNN

It is another facet of deep learning specifically tailored for processing structured grid data or images, in our case. The architecture of CNN takes its inspiration from the human visual system, thereby allowing it to be autonomous in terms of the learning of spatial hierarchies and processing them efficiently [35]. CNNs in the proposed model are an important aspect, where they support the feature extraction from the given histopathological images and thus help in the decision-making process. CNN architecture comprises various layers, namely the convolutional layer, the pooling layer, and the fully connected layer [36], [37]. Out of these, the convolutional layers are adept for the task of feature extraction as they convolve the image using various filters to gather information such as edges, textures, as well as colors in the initial screening. With the progression of the data in the deeper layers, these filters allow for the capturing of complex attributes, which comprise the various shapes and objects providing relevant information for the identification of cancerous tissues, as in our case. The pooling layers that follow the convolutional layers are generally embedded to reduce the spatial dimension of the feature maps generated by the previous layers [38]. This dimensionality reduction is achieved with varied operations like max pooling, aimed at

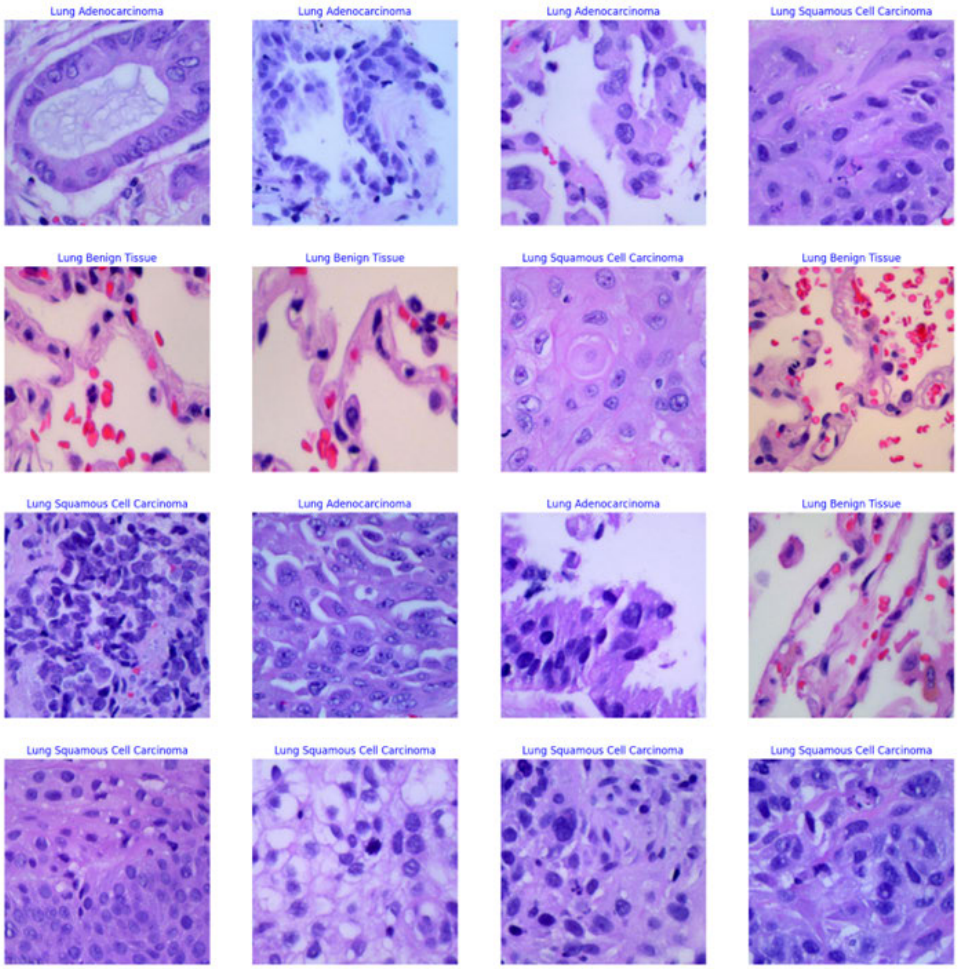

FIGURE 1. Various carcinoma tissues in the dataset.

reducing the complexity of the feature map while retaining the important features. This layer, in turn, helps to target the risk of overfitting and helps in enhancing the robustness of the feature selection process. These abilities of the CNN architecture to hierarchically and efficiently extract features from high-resolution histopathological images made it the optimal choice for the proposed architecture. The extracted feature maps from the CNN would allow the subsequent KAN architecture to work on a well-represented feature set.

3) KAN-ENHANCED FEATURE PYRAMID NETWORK (KAN-FPN)

The KAN-based feature pyramid is a novel aspect of this methodology that uses the mathematical transformations of KAN networks and enhances the capabilities of Feature Pyramid Networks (FPN). This dual-pronged approach boosts the efficacy of feature extraction and representation of the images across the various scales. As such, the development of the same model improves the performance of the CNN in detecting and classifying lung cancer from CT scan images. The FPN, the basis of the proposed KAN-FPN, is known for its prowess in handling the features at various scales, making it a prompt choice for tasks like object detection and image segmentation. Traditional FPNs have highlighted the issues of complex feature capturing and non-linear relationships, which are generally associated with medical images. To tackle this issue, the KAN mathematical basis was used. The aspect of KAN is to approximate and represent the complex functions in the form of univariate transformations, and the improved network allows for a more enhanced feature extraction mechanism. The architecture integrates KAN transformation at each level of FPN to change the feature map. This transformation allows the breaking down of the feature maps into their univariate components, thus reducing complexity and recombining them. The condensation process allows the network to capture the intricate patterns more effectively. In practice, the network allows the multi-scale feature extraction of FPN to be combined with the approximations of the KAN networks, thereby making a more robust and precise analysis. The advantage of the hybrid structure also lies in the ability to enable integrity with the help of high resolution

and semantic richness spread across the various layers. This processing allows for improved performance and is particularly valuable in lung cancer detection, where early and accurate identification of anomalies can significantly impact patient outcomes.

4) CONTEXT-AWARE MECHANISM (CAM)

The proposed architecture has been enhanced with the help of the context-aware mechanism, which is used to enhance the model's ability to focus on certain related features of an input image. Thus, the usage of the same has supported the refining and optimization of the feature extraction process. The working of CAM in the process of the dynamic generation of attention maps, which emphasizes the most important informative regions of a medical image. The said mechanism proves itself to be particularly valuable in the aspect of medical images, where subtle variations make it critical for accurate diagnosis. Taking from a practical example, the CT scans of the image for lung cancer may be inorganically distributed, meaning certain areas will be more important while the other areas might be irrelevant to the study for a while. As such, CAM allows the model to focus on the crucial areas and thus improves both the accuracy and interpretability of the predictions. The CAM methodology in this aspect has been applied after the integration of the KAN-based FPN. The FPN network is tasked with the extraction of features at multiple scales and then the CAM network enhances these features by highlighting the most contextually relevant parts. The multiplication of feature maps with attention maps allows CAM to suppress the noise and enhance the signal to make the model robust to the variations of the given input data. The integration of CAM within the proposed architecture will help to significantly improve the model's performance as well as interpretability. The attention maps that are generated by the CAM help to provide insights as to what parts are being highlighted for the model output. In summary, CAM significantly contributes to the novelty and effectiveness of the proposed model by ensuring that the network focuses on the most relevant features during both the learning and inference phases.

5) LIME

Local Interpretable Model-agnostic Explanations (LIME) is another aspect of known Explainable AI algorithms. It provides an efficient way to interpret the results provided by the complex machine learning algorithms by the production of locally interpretable explanations catering to specific predictions [39]. These allow the LIME model to augment the overall interpretability of the model and provide a better understanding of the decision-making process. In the context of image processing and interpretation, LIME progresses by perturbing the given image and then examining the resultant vibration in the predictions generated by the model [40], [41]. The technique in general encompasses the generation of a series of perturbed images that stem from the original image with the sophisticated usage of stochastic modifications. These modifications include, but are not limited to, occlusion and distortion, generating various perturbed images of the given input. These perturbed images are then used to document the model's prediction, with a dataset being created for the pairs of images and respective features.

Mathematically, let x represent the original image, and let $f(x)$ denote the model's prediction function. LIME generates perturbed images x' prime and computes their predictions $f(x')$. It then fits a local interpretable model g to these predictions, where g is typically a simple model such as a linear classifier or a decision tree. The goal is to approximate the complex model f in the vicinity of x with the simpler model g.

The explanation provided by LIME can be expressed as Eq. (2).

$$g(x) = argmin_{g \in G} \sum_{x' \in \mathcal{N}(x)} Loss\left(f\left(x'\right), g\left(x'\right)\right) . Weight(x, x') \quad (2)$$

where $N(x)$ is the neighborhood of perturbed images around x, Loss measures the discrepancy between the predictions of f and g, and $Weight(x, x')$ assigns weights to perturbed samples based on their proximity to x. With the generation of these local explanations, LIME bolsters the understanding of each feature in terms of its importance to a particular image. This allows us to understand the contribution of the features to the various decisions made by the models, thus bolstering transparency and trust in the CAD systems.

6) SHAP (SHAPLEY ADDITIVE ExPlanations)

SHAP is another sophisticated approach that is used for the interpretation of machine learning models, which provides theoretically sustained explanations for the various individual predictions [42]. In the aspect of image interpretations, SHAP makes use of the Shapley values, which stem from cooperative game theory, to discern the contribution of each feature to the model predictions, thereby allowing a more nuanced, granular insight into the specific features and their corresponding outputs. SHAP operates by the computation of the Shapley values attributed to each feature and thus quantifies the contribution of each to the predictions compared to the base model [43]. This method involves discerning even the marginal contribution of each feature across all possible permutations to ensure a comprehensive and equitable distribution of the same. For a given input image x and model f, SHAP computes the Shapley value ϕ_i for each feature i as follows Eq. (3).

$$\phi_{ij} = S \subseteq N \setminus \{i\} \, |S| \, N - |S^{-1}| \, |N| \, [f_S \cup \{i\} - f_S] \quad (3)$$

where N is the set of all features, S is a subset of features excluding i, and $f(S)$ is the model's prediction with features S. This Eq. (3) calculates the average marginal contribution of a feature across all possible subsets of features, ensuring a fair distribution of the prediction's explanation. In the context of

image segmentations, SHAP values elucidate the importance of features concerning individual pixels or regions that steer the models' decision-making process. It attributes numerical values to each feature, allowing SHAP to visualize which components predominantly influence the predictions. This sophisticated mechanism allows for the enhancement of the transparency of the model and enables clinicians to understand the rationale for the decision-making process.

C. PROPOSED MODEL-KAFPN

In the proposed model for the aspect of enhancing lung cancer detection, the KAFPN makes use of multiple advanced mechanisms like CNN, KAN, FPN as well as CAM. The basis of architecture is based on the usage of the CNN layer, which is used to extract the key features from the images and is then passed for further processing. Since CNNs are adept at understanding the various complex medical image analyses, the integration of the same with the model proves to be a significant advantage. The model works by taking an input image, the size of which is defined as per the size of the training image, and then it passes through a convolutional layer. These layers extract the basic features like the edges, textures, and patterns from the images. This extraction is followed by the spatial dimension, which allows for the retention of the important features while downscaling the data to reduce the computation. The next step of the architecture is the application of the FPN, which has been enhanced with the help of the KAN fundamentals. In the FPN, feature maps are extracted at multiple scales to provide a multi-level representation of the image. This representation allows to capture of both fine details and coarse features of the images. The model is further enhanced with the inclusion of the KAN transformation at each level of the feature pyramid. At each level of the pyramid, the extracted feature map is flattened, and then the KAN transformation is applied, where the univariate transformations are performed on the selected features. The transformation of the features with the help of the dense layers, allows the network to be able to capture the complex dependencies. The transformed features are then attributed to dimensionality reduction with the help of max pooling at each level of the FPN.

Once the KAN-FPN has processed the feature maps on varying scales, the CAM mechanism has been applied for further enhancement. The CAM integration works with the generation of attention maps that highlight the most contextually relevant part of the image. This map is then multiplied by the feature map, amplifying the important regions that contribute most to the classification aspect. The output of the CAM from each of the FPN levels is then concatenated. This allows for the creation of a rich-level feature representation, the following is then passed through another univariate KAN transformation. This step thus allows for the enhancement of the network's ability to detect intricate patterns in the image data. With the successful completion of the hierarchical aggregation and the final KAN transformation, the features are then passed through a dense layer and a softmax activation function to predict the class labels. The softmax function outputs the probability distribution across the classes, helping in classifying the input into its respective category.

An overview of the above work can be understood with the help of Algorithm 1, which emphasizes the flow of input and the generation of output.

Algorithm 1 KAFPN Working

Input: Image I
Output: Classification Probabilities
Function `KAFPN()`:
Step 1: Initialize the CNN layer and apply a convolution operation as:

$$F_1 = \sigma(W_1 * I + b_1)$$

where F_1 is the generated feature map, W_1 is the convolutional kernel, b_1 is the bias term, $*$ denotes the convolution operation, and σ is the activation function (ReLU).
Step 2: Apply Max Pool operation on the feature map F_1:

$$F_1 = \text{MaxPool}(F_1)$$

STAGE 1: Apply KAN-FPN
for *each level l in the Feature Pyramid* **do**
Step 3: Apply convolution operation:

$$F_l = \sigma(W_l * F_{l-1} + b_l)$$

Step 4: Flatten F_l.
Step 5: Apply KAN Transformation:
Step 6: Split F_l into g_l as:

$g_l = \lambda(x)$ where λ(lambda $x : x[:, i :: \text{num\_transformations}])(features)$

Step 7: Apply dense transformation $T(i)$.
Step 8: Summarize the $T(i)$ as:

$$\phi(F_l) = \sum_{i=1}^{2} T_i(g_i(F_l))$$

Step 9: Apply MaxPooling to downsample.
Step 10: Store F_l.
STAGE 2: Apply CAM
for *each level l in the Feature Pyramid* **do**
Step 11: Generate Attention Map A_l:

$$A_l = \sigma(W_{\text{cam}} * F_l + b_{\text{cam}})$$

where W_{cam} is the convolution operation with $1 \times 1 \times 1$ kernel and b_{cam} is the bias term.
Step 12: Create attention weighted feature map F'_l:

$$F'_l = F_l * A_l$$

STAGE 3: Feature Aggregation and Classification
Step 13: Concatenate all layers of Pyramid:

$$F_{\text{agg}} = \text{Concatenate}(F'_1, F'_2, F'_3)$$

Step 14: Apply final KAN Transformation:

$$\phi(F_{\text{agg}}) = \sum_{i=1}^{2} T_i(g_i(F_{\text{agg}}))$$

Step 15: Apply Dense Transformation:

$$D = \sigma(W_d * \phi(F_{\text{agg}}) + b_d)$$

Step 16: Classify using softmax:

$$P(y|I) = \text{Softmax}(W_d * D + b_d)$$

The above algorithm captures the essence of the algorithm, which gives the classification output after the usage of the FPN and the CAM aspects enhanced with the help of KAN, thereby generating the KAFPN proposed architecture.

To augment the explainability of the results and make the decisions more interpretable, the results from the KAFPN were subjected to both LIME and SHAP to attain a deeper understanding of the model's decision-making process. LIME helps to identify regions that most significantly influence the model's predictions. Specifically, LIME provides visual explanations by accentuating the various regions of lung histopathological images. This visual interpretability is crucial for clinicians as it allows for the foundations of a strong CAD system. Furthermore, the architecture incorporates the usage of SHAP, which provides SHAP values for the quantification of the contribution of each feature to the final predictions, the results of which are grounded in cooperative game theory. This bolsters the model by analyzing the influence of different regions of the image and explicating how each component impacts the model's output. This dual-pronged approach, grounded in the facets of Explainable AI (XAI), leveraging both LIME and SHAP, allows the model to perform accurately and be interpretable. The proposed model thus integrates the theoretical robustness of the KAN, the feature extraction capabilities of the CNN architecture, and the enhanced interpretability provided by LIME and SHAP. This amalgamation, as depicted in Fig. 2, results in a resilient framework that is more aligned with the medical requirements to provide explanations of the predictions and also supports the medical practitioners in the process of decision making.

D. DATA PREPROCESSING

Given that the dataset was clean and standardized, with images bearing a uniform size of 768 × 768 pixels, data augmentation was employed to be able to generate a diverse set of sample images. Data augmentation includes the artificial enlargement of the dataset, such that the model can learn more robust features and enhance its generalizability to novel data [44], [45]. The dataset originally comprised 15,000 images belonging to the three classes, which were then expanded to about 25,000 images. This number ensures that the risk of overfitting to a limited number of original images is limited, and the presence of stochastic variation in the training images allows the model to be more prepared for real-world data prone to variability and noise. To be better able to understand the performance of the proposed model, the dataset was split into training, testing, and validation datasets, with the ratio between them being 70%, 15%, and 15% in the particular order. This ensures that the model has sufficient data to learn from and also to assess the model's generalization of results [46]. This optimum split ensures that the model's performance is fairly evaluated, ensuring that it is capable of accurate predictions on new samples without being prone to overfitting data. The classification metrics with their corresponding formulas to make the suggested results more reliable, as follows: Eqs. (4), (5), (6), (7), (8), (9).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{6}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{AUC} = \int_0^1 \text{Sensitivity}\, d(1 - \text{Specificity}) \tag{8}$$

$$\text{F-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

In this paper, metrics like accuracy and precision were used to highlight the performance of the model, but to adhere to real-world applications, a detailed analysis of the computational overheads was also performed. The proposed model was therefore examined with the help of various aspects, like training time, which estimated that the model being trained on a T4 GPU of Google Cloud for 150 epochs, for a batch size of 64, took an approximate time of 2.5 hours to complete. This higher training time could be attributed to the fact that the model was complex in terms of the inclusion of aspects of KAN and CAM components. The model requires approximately 3.8 GB of GPU memory during training and 1.2 GB for inference. This is largely attributed to the hierarchical FPN structure and attention mechanisms, which increase parameter count and intermediate tensor sizes.

IV. RESULTS

This section explores the workings of the proposed model on the data and compares it with the literature to understand how well it fared. The proposed KAN-based model was tested extensively on both a testing dataset and a validation dataset. The results were fairly accurate, with the model giving a testing accuracy of 0.9963 and a validation accuracy of 0.9960. These metrics show that the model is proficient in terms of discerning the salient features of the histopathological images and thus accurately classifying them.

A. BENCHMARK COMPARISONS

To better facilitate and benchmark the performance of the model, various distinguished and established architectures were used to compare their performance. These include Base CNN, Inception V3, VGG16, and ResNet 50 model architecture. The various models were facilitated by a rigorous comparison which was subjected to both preprocessed data and processed data (Table 3 and Table 4). We have initially chosen basic CNN and VGG16 models for benchmark comparisons with our proposed model due to their frequent applications in prior research and historical significance. However, as the computer vision field is

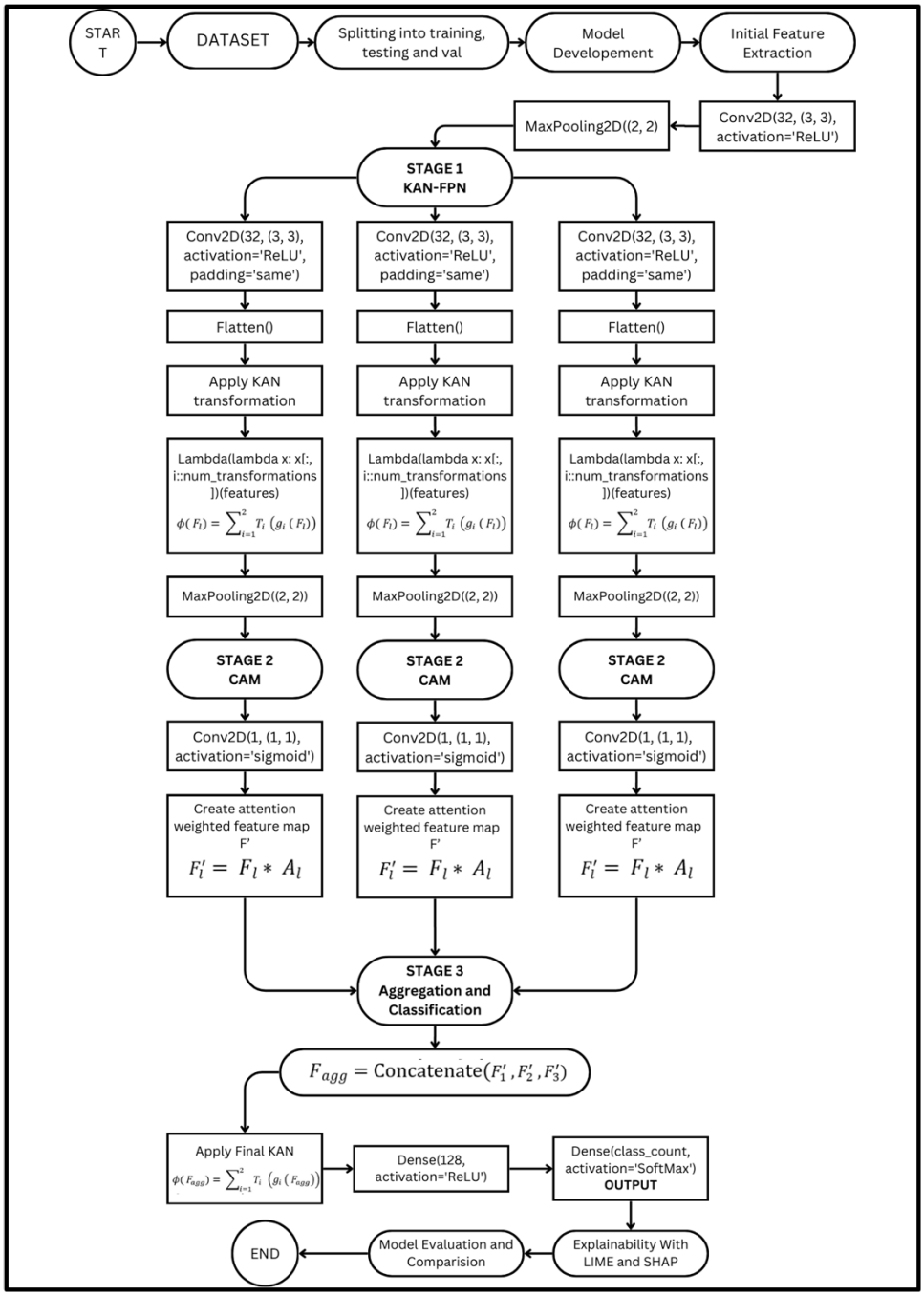

FIGURE 2. Flowchart of proposed model.

growing rapidly, we have also incorporated Inception V3, ResNet50, EfficientNet-B7, ResNet-101, and DenseNet-121, which have shown promising results in brain tumor segmentation and similar tasks. Furthermore, we have reviewed the recent literature and included some other models, such as the LungStat model, the LungNet model, and hybrid SVM-CNN models. [47], [48], [49]. The LungStat model achieved an accuracy of 90% combining the use of class activation mapping to show the usage of computer vision in lung cancer detection [47]. Further advancements in the field include LungNet, which makes use of wearable sensor data to diagnose lung cancer, showing a high accuracy of 96.81% [48]. The developed model has 22 layers of unique aspects of IoT-enabled detection. Integrating SVM with CNN to eliminate the aspect of irrelevant information and reported an accuracy of 97.91% [49]. It is found that our proposed KAFPN model outperforms these models in terms of accuracy scores.

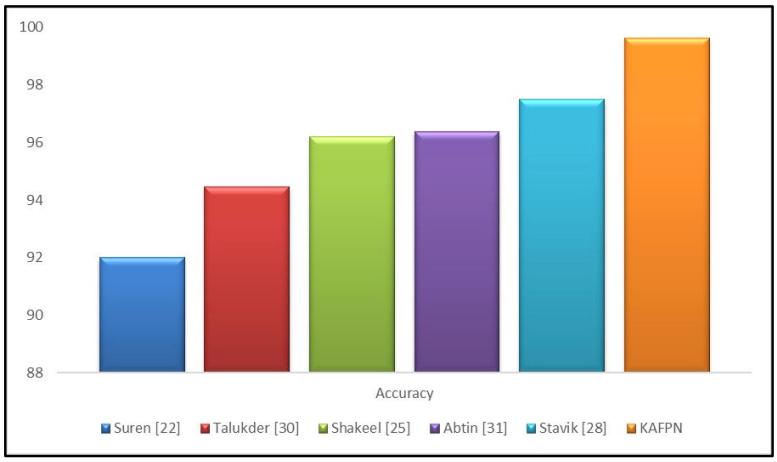

FIGURE 3. Accuracy comparison between various models.

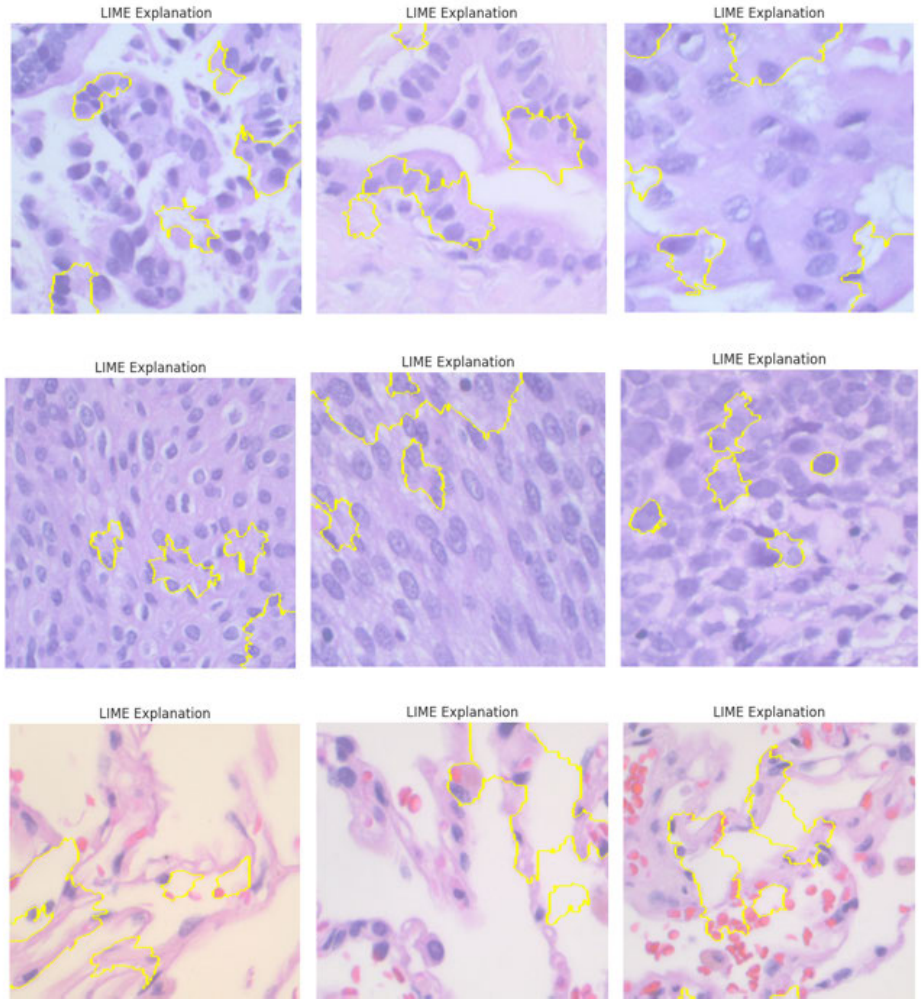

FIGURE 4. LIME interpretation.

In Table 3, the classification metrics of accuracy, sensitivity, specificity, precision, AUC, and F-score were utilized. These metrics facilitate an exhaustive evaluation of the model's performance across various dimensions, also adhering to the comparison of the model's performance subjected to the various models after data processing (Table 4). The

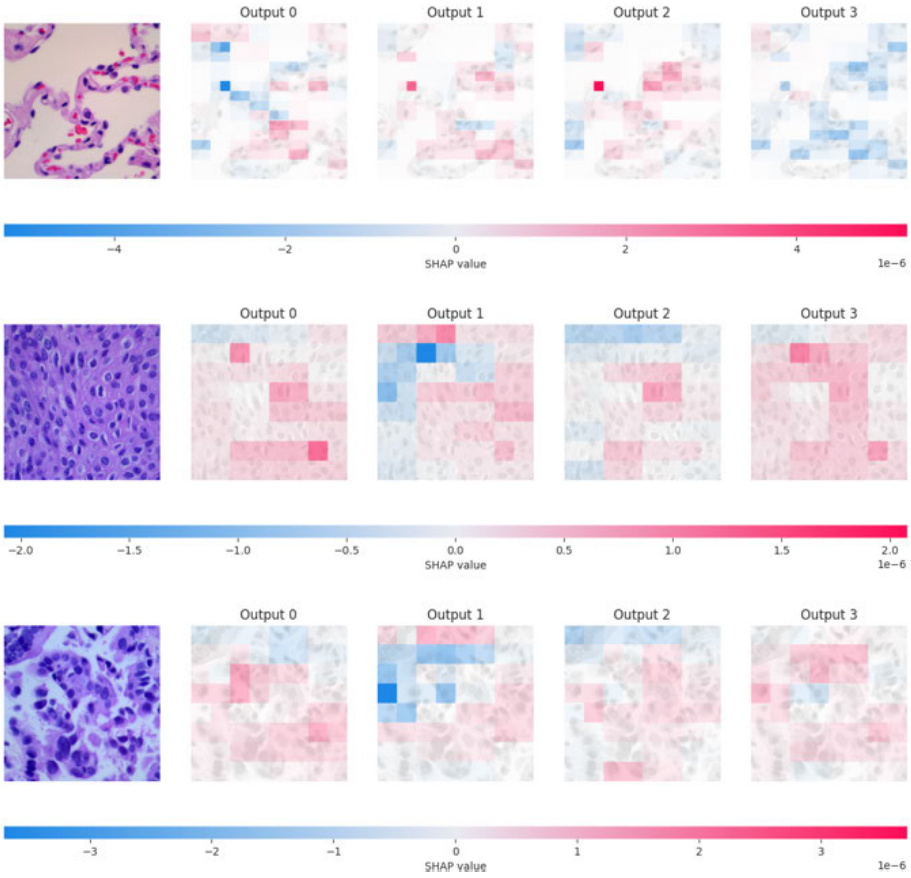

FIGURE 5. SHAP interpretation.

comparison of the model with existing literature can be found in Figure 3, showing the prowess of the KAFPN model as compared to its counterparts across the majority of the metrics. The model was able to exhibit a higher value of accuracy, underscoring its superior capability in detecting and classifying lung cancer histopathological images. The comparison between Table 3 and Table 4 reveals significant improvements in the performance of all models after data preprocessing. In terms of accuracy, the KAFPN model showed the most substantial improvement, increasing from 73.31% to 99.63%, a 26.32% jump. The ResNet 50 model also demonstrated a notable improvement, rising from 79.60% to 97.74%, an 18.14% increase.

The sensitivity of the models also saw considerable enhancements. The KAFPN model's sensitivity increased from 59.43% to 99.11%, a 39.68% improvement. The ResNet 50 model's sensitivity rose from 64.58% to 98.06%, a 33.48% increase. Similarly, the Inception V3 model's sensitivity jumped from 62.65% to 97.06%, a 34.41% improvement.

The Area Under the Curve (AUC) values also exhibited notable improvements. The KAFPN model's AUC increased from 0.74 to 0.99, a 25% jump. The ResNet 50 model's AUC rose from 0.77 to 0.91, a 14% increase. The Inception V3 model's AUC improved from 0.77 to 0.95, an 18% increase. These enhancements demonstrate the significant impact of data preprocessing on the performance of the models.

B. LIME AND SHAP INTERPRETABILITY

Proceeding to the subsequent component of the proposed model, the focus shifted towards the augmentation of the explainability of results for the various decisions made. To further this and dissect the opaqueness of the KAFPN model, LIME and SHAP techniques were used. As already discussed, LIME gives local explanations and identifies the regions that contribute to the final predictions; an example of the same can be found in Figure 4. In this figure, the LIME explanation of the original image highlights the regions that are most influential in the decision-making process by the yellow color. This allows the clinicians and the medical practitioners to pinpoint the exact region that the model utilized to validate the CAD diagnostic suggestion. These visual cues from the LIME algorithm facilitate the medical professionals to interpret the complex outputs in a conducive manner.

Conversely, the SHAP algorithm was also used to elucidate the individual predictions, which makes use of game theory. The compiled SHAP values allow us to quantify the

TABLE 3. Models' classification performances before data preprocessing.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	AUC	F-score (%)
Base CNN	74.98	69.54	55.84	80.74	0.78	63.74
Inception V3	68.58	62.65	63.00	74.55	0.77	64.11
VGG16	76.74	64.44	59.28	76.45	0.75	63.02
ResNet 50	79.60	64.58	58.94	76.05	0.77	64.60
EffecientNet-B7	78.29	69.25	62.39	79.22	0.71	62.39
ResNet-101	69.33	65.71	57.34	75.49	0.76	65.22
DenseNet-121	77.59	58.55	63.89	71.86	0.72	62.78
KAFPN	73.31	59.43	67.12	74.68	0.74	66.34

TABLE 4. Models' classification performances after data preprocessing.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	AUC	F-score (%)
Base CNN	96.58	98.51	98.95	97.72	0.93	96.53
Inception V3	98.89	97.06	97.23	98.72	0.95	96.35
VGG16	96.56	98.23	97.01	99.18	0.94	97.78
ResNet 50	97.74	98.06	97.69	98.09	0.91	97.64
EffeicientNet-B7	98.11	97.31	97.14	98.37	0.92	94.11
ResNet-101	95.43	92.13	91.21	95.19	0.95	98.26
DenseNet-121	96.77	95.18	95.32	97.16	0.91	97.16
KAFPN	99.63	99.11	98.78	99.23	0.99	98.36

TABLE 5. Models' classification performances per different classes.

Model	Class	Accuracy (%)	Sensitivity	Specificity	Precision
Base CNN	Adenocarcinoma	97.34	0.979	0.957	0.978
	Benign	97.66	0.955	0.959	0.974
	Squamous	95.70	0.952	0.973	0.979
Inception V3	Adenocarcinoma	96.95	0.980	0.978	0.963
	Benign	96.66	0.971	0.962	0.968
	Squamous	97.55	0.976	0.961	0.953
VGG16	Adenocarcinoma	97.64	0.979	0.969	0.967
	Benign	96.52	0.956	0.953	0.960
	Squamous	96.35	0.973	0.951	0.966
ResNet 50	Adenocarcinoma	95.92	0.957	0.970	0.970
	Benign	96.31	0.953	0.966	0.969
	Squamous	96.66	0.951	0.976	0.959
ResNet 50	Adenocarcinoma	95.92	0.957	0.970	0.970
	Benign	96.31	0.953	0.966	0.969
	Squamous	96.66	0.951	0.976	0.959
EffecientNet-B7	Adenocarcinoma	97.95	0.981	0.98	0.980
	Benign	98.10	0.983	0.981	0.982
	Squamous	98.40	0.984	0.983	0.982
ResNet-101	Adenocarcinoma	97.85	0.982	0.979	0.978
	Benign	98.02	0.982	0.980	0.981
	Squamous	98.35	0.983	0.982	0.981
DenseNet-121	Adenocarcinoma	98.10	0.982	0.981	0.981
	Benign	98.20	0.983	0.981	0.982
	Squamous	98.55	0.984	0.983	0.982
KAFPN	Adenocarcinoma	98.78	0.989	0.986	0.990
	Benign	99.13	0.990	0.989	0.987
	Squamous	99.74	0.994	0.992	0.993

contribution of each pixel to the model's output. Figure 5 provides a visualization of these contributions of the individual pixels with the help of a heatmap. The colors of the heatmap indicate the directions and magnitude of the influence of these pixel areas (red denotes the positive contribution and blue vice versa). This allows the practitioner to not only understand which features are important but also how the different features interact with each other, thereby influencing the model prediction. Both the techniques used in conjunction provide distinct advantages this includes LIME provides an overview of the immediate, instance-specific insights that are crucial to the decision-making process, while on the other hand, SHAP provides global consistent patterns across the dataset to ensure that the model is overall reliable and fair. The usage of LIME and SHAP allows for an essential aspect of interpretability as explored by the proposed structure to improve the implementation of deep learning models in the medical domain.

In Figure 4, it is seen that LIME highlights the specific localized areas with the help of the yellow boundaries in the given histopathological image. These are the regions that contribute significantly to the model's predictions and thus are highlighted. The presence of LIME's superpixel-based segmentation allows clinicians to incorporate them in their workflow, where clinicians focus on cellular and tissue-level features such as nuclear pleomorphism, hyperchromasia, or abnormal mitotic activity. Examining the highlighted areas, it is seen that these are the regions with hypercellularity or disrupted architecture, which validates the capacity of the model to make such decisions and are also in line with the clinician's diagnostic process. The practical utility here lies in leveraging such visual cues to expedite manual reviews, ensuring that high-risk areas are scrutinized further.

Analyzing the aspect of SHAP values from Figure 5, the visualization shows that there are multiple heatmaps generated against the outputs. The additive explanations of SHAP allow an advantage to quantify the feature contribution of each region, thereby proving to be beneficial in complex histopathological assessments. The red and blue regions indicate positive and negative contributions, respectively, in terms of contributing to the model's decision-making process. The visualizations allow for a better assessment that is more in line with pathological indicators such as tumor heterogeneity, necrosis, or infiltrative margins. As seen, the red regions highlight the presence of dense tumor cells or necrotic zones, which present themselves as a reliable signal to highlight the instance to be malignant. Therefore, with the help of the LIME and SHAP explanations it helps to enhance trust in AI systems, particularly in environments where decisions have life-altering consequences. By aligning the outputs of LIME and SHAP with domain knowledge, such as patterns in adenocarcinoma or squamous cell carcinoma, the models can reinforce decision-making processes.

The incorporation of both LIME and SHAP into the evaluation process of KAFPN not only improves the model's transparency but also helps in fostering a trustworthy AI in the field of medical CAD. Analyzing Figures 4 and 5 from the perspective of a medical explanation, it is seen that there is greater clinical relevance, which helps to aid healthcare professionals. This dual interpretation was done to ensure that the model is performing accurately, with the results being justifiable and comprehensible. This integration allows the proposed model to distinguish itself by providing reliable and actionable insights.

C. CLASSIFICATION PERFORMANCE ACROSS ADENOCARCINOMA, BENIGN, AND SQUAMOUS CLASSES

Table 5 presents the classification performance of the various models as per the lung cancer classes (Adenocarcinoma, Benign, and Squamous). A thorough examination of the table reveals that the KAFPN model consistently outperforms its counterparts, achieving superior accuracy rates of 98.78%, 99.13%, and 99.74% for Adenocarcinoma, Benign, and Squamous classes, respectively. Furthermore, the KAFPN model demonstrates exceptional sensitivity, specificity, and precision values, indicating its robustness and reliability in classifying these cancer types. The model's outstanding performance can be attributed to its ability to effectively capture and represent complex patterns and features inherent in the data. In comparison, the other models, including Base CNN, Inception V3, VGG16, ResNet 50, EfficientNet-B7, ResNet-101, and DenseNet-121, also exhibit satisfactory performance, albeit with slightly lower accuracy rates and other metrics compared to the KAFPN model. Notably, the EfficientNet-B7 and DenseNet-121 models demonstrate impressive accuracy rates, often exceeding 98%, indicating their strong potential in cancer classification tasks. These findings suggest that the KAFPN model is the top-performing model in this study, followed closely by other models that also demonstrate robust performance in classifying different types of cancer. The results of this study have significant implications for the development of accurate and reliable cancer classification systems, and highlight the importance of careful model selection and evaluation in achieving optimal performance.

D. IMPACT OF AUGMENTATION

The proposed study uses various techniques to highlight the aspect of data augmentation that has been utilized in the study. The detailed experiments about the same included different techniques, such as the usage of random rotations, which augmented the images up to $\pm 15°$ to be able to simulate the various aspects of different viewing angles. Augmentations also included various viewing angles with the help of horizontal and vertical flipping and the usage of zooming between 90%-110% for making varied distances for image acquisition. Other aspects of data augmentation included shifting to have translations of up to 10% of the image dimensions while incorporating the aspect of various lighting conditions with the implementation of brightness adjustments by $\pm 20\%$. These augmentations were implemented in the dataset with the help of the Keras ImageDataGenerator class and applied to the training process. To understand the performance of these augmentations, the evaluation of the proposed model was done under three conditions, including without augmentation, where there was no modification to the raw dataset, with basic augmentation, which included augmentations of rotations, flips, and zoom. The third category applied to the same was advanced augmentation techniques, where the combination of all the various as discussed was applied.

The model's performance was assessed using accuracy, precision, recall, and F1-score. The results are listed in Table 6 as follows. From the results, it can be seen that the presence of data augmentation significantly improves the results of the proposed algorithm in terms of more generalizability and, thereby, the aspect of accuracy. Table 6 presents a comprehensive evaluation of the impact of different data augmentation strategies on the performance of classification

TABLE 6. Performance metrics of the classification models with different data augmentation strategies.

		Accuracy (%)	Sensitivity	Specificity	Precision
Without Augmentation	Adenocarcinoma	87.32	0.874	0.872	0.875
	Benign	87.63	0.875	0.874	0.873
	Squamous	88.17	0.879	0.877	0.878
With Basic Augmentation	Adenocarcinoma	91.513	0.916	0.913	0.917
	Benign	91.837	0.917	0.916	0.914
	Squamous	92.402	0.921	0.919	0.920
Advanced Augmentation	Adenocarcinoma	98.78	0.989	0.986	0.990
	Benign	99.13	0.990	0.989	0.987
	Squamous	99.74	0.994	0.992	0.993

TABLE 7. Performance metrics of the classification models on balanced and imbalanced datasets.

		Accuracy (%)	Sensitivity	Specificity	Precision
Balanced Dataset	Adenocarcinoma	98.78	0.989	0.986	0.990
	Benign	99.13	0.990	0.989	0.987
	Squamous	99.74	0.994	0.992	0.993
Imbalanced Dataset (70-20-10)	Adenocarcinoma	96.11	0.813	0.991	0.986
	Benign	99.44	0.974	0.995	0.991
	Squamous	98.33	0.991	0.981	0.996

models in distinguishing between Adenocarcinoma, Benign, and Squamous cancer types.

The results indicate that the absence of data augmentation (Without Augmentation) yields accuracy rates of 87.32%, 87.63%, and 88.17% for Adenocarcinoma, Benign, and Squamous classes, respectively. These findings suggest that the models' performance is suboptimal without data augmentation, highlighting the importance of this technique in enhancing model robustness. The implementation of basic data augmentation strategies (With Basic Augmentation) leads to a notable improvement in model performance, with accuracy rates increasing to 91.51%, 91.84%, and 92.40% for Adenocarcinoma, Benign, and Squamous classes, respectively. This enhancement underscores the effectiveness of basic data augmentation in improving model performance. However, the most significant improvement is observed when advanced data augmentation strategies (Advanced Augmentation) are employed. The accuracy rates surge to 98.78%, 99.13%, and 99.74% for Adenocarcinoma, Benign, and Squamous classes, respectively. This substantial enhancement demonstrates the profound impact of advanced data augmentation on model performance, highlighting its crucial role in achieving state-of-the-art results.

As visible in Table 6, the model trained without augmentation performed well in the generalization aspect, as reflected in the aspect of model accuracy and F1-score, which was the lowest compared to the counterparts, likely attributed to the presence of overfitting. Compared to the usage of basic augmentation, it is seen that the model was able to perform better with an increase of 4.8% compared to its counterpart in the aspect of accuracy. The presence of such augmented data allowed the model to perform better generalization and be less prone to overfitting by being able to learn from more diverse features. Moving on, it is seen that the advanced augmentation techniques provided similar results, with the model's performance being increased by a factor of 11.6% compared to the baseline aspect. The findings presented in Table 6 provide compelling evidence for the importance of data augmentation in enhancing the performance of classification models in cancer diagnosis. The results demonstrate that advanced data augmentation strategies can significantly improve model accuracy, sensitivity, specificity, and precision, ultimately leading to more reliable and accurate cancer diagnosis.

E. IMPACT OF THE UNBALANCED DATASET

While the dataset used in this study is balanced, real-world medical datasets often suffer from class imbalance, which can significantly impact model performance. To address this, we artificially induced imbalances in the dataset and evaluated the model's performance under these conditions. To simulate imbalance, we applied random undersampling to reduce the number of samples in certain classes. Specifically, for a three-class dataset, we created an imbalance by setting the class distributions to:

- Class 1 (Majority): 70% of total samples
- Class 2 (Moderate): 20% of total samples
- Class 3 (Minority): 10% of total samples

This setup mimics real-world scenarios where certain conditions are significantly more prevalent than others. The performance of the proposed model was compared under balanced and imbalanced settings. Key metrics, including accuracy, precision, recall, and F1-score, are summarized below in Table 7. Table 7 presents a comparative analysis

of the performance metrics of classification models on balanced and imbalanced datasets. The results are categorized into two distinct datasets: a balanced dataset and an imbalanced dataset with a class distribution of 70-20-10. The balanced dataset yields exceptional performance metrics, with accuracy rates of 98.78%, 99.13%, and 99.74% for Adenocarcinoma, Benign, and Squamous classes, respectively. The sensitivity, specificity, and precision values are also remarkably high, indicating the models' ability to effectively distinguish between the classes. In contrast, the imbalanced dataset presents a more challenging scenario. The accuracy rates for Adenocarcinoma, Benign, and Squamous classes are 96.11%, 99.44%, and 98.33%, respectively. Notably, the sensitivity value for Adenocarcinoma is significantly lower (0.813) compared to the balanced dataset, indicating the model's difficulty in detecting this class in the presence of class imbalance. The results suggest that the classification models are more robust and effective when trained on balanced datasets. However, when faced with imbalanced datasets, the models' performance may be compromised, particularly for the minority class (Adenocarcinoma). These findings underscore the importance of addressing the class imbalance in dataset construction to ensure the development of reliable and accurate classification models.

These results highlight the importance of addressing class imbalances in real-world applications. Techniques like Synthetic Minority Over-sampling Technique (SMOTE) or cost-sensitive learning could be incorporated to mitigate performance degradation in future work.

V. CONCLUSION

The KAFPN model suggested for improving lung cancer detection performed exceptionally well in terms of efficacy and accuracy while classifying the images in the histopathological image dataset. The proposed architecture builds upon integrating the inherent capabilities of CNNs with the theoretical support of the KAN-based FPN and CAM mechanism to produce a model that achieved an accuracy of 99.63% for testing purposes and 99.60% for validation purposes. These metrics show that the model performed well, and the same is corroborated by the various tables, showing that the model outperformed the base models in the different metrics used. The innovative incorporation of Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive explanations (SHAP) further augmented the model's interpretability, elucidating the visual regions of the images that most significantly influenced the model's predictive decisions. The increased interpretability that the two provide allows medical professionals to place their trust in the system and, eventually, in the CADs. The comparison of the model with the existing literature and the baseline models allows us to understand that KAFPN provides a better accuracy and is more fit to understand the complex patterns of the given dataset. The presence of LIME and SHAP also contributes to the increased interpretability of the study but also adheres to the fact that there could be improvements in the future. The study, therefore, contributes to the literature by presenting a novel model that enhances the aspects of lung cancer detection. Future research in this area calls for better modalities and preprocessing techniques to broaden the impact on the healthcare sector.

REFERENCES

[1] R. Nooreldeen and H. Bach, "Current and future development in lung cancer diagnosis," *Int. J. Mol. Sci.*, vol. 22, no. 16, p. 8661, Aug. 2021.
[2] A. Leiter, R. R. Veluswamy, and J. P. Wisnivesky, "The global burden of lung cancer: Current status and future trends," *Nature Rev. Clin. Oncol.*, vol. 20, no. 9, pp. 624–639, Sep. 2023.
[3] Y. Li, B. Yan, and S. He, "Advances and challenges in the treatment of lung cancer," *Biomed. Pharmacotherapy*, vol. 169, Dec. 2023, Art. no. 115891.
[4] V. Buhrmester, D. Münch, and M. Arens, "Analysis of explainers of black box deep neural networks for computer vision: A survey," *Mach. Learn. Knowl. Extraction*, vol. 3, no. 4, pp. 966–989, Dec. 2021.
[5] A. Das and P. Rad, "Opportunities and challenges in explainable artificial intelligence (XAI): A survey," 2020, *arXiv:2006.11371*.
[6] A. B. Arrieta, N. Díaz-Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.
[7] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4793–4813, Nov. 2021.
[8] X. Man and E. P. Chan, "The best way to select features? Comparing MDA, LIME, and SHAP," *J. Financial Data Sci.*, vol. 3, no. 1, pp. 127–139, Jan. 2021.
[9] B. Aldughayfiq, F. Ashfaq, N. Z. Jhanjhi, and M. Humayun, "Explainable AI for retinoblastoma diagnosis: Interpreting deep learning models with LIME and SHAP," *Diagnostics*, vol. 13, no. 11, p. 1932, Jun. 2023.
[10] H. T. T. Nguyen, H. Q. Cao, K. V. T. Nguyen, and N. D. K. Pham, "Evaluation of explainable artificial intelligence: Shap, lime, and cam," in *Proc. FPT AI Conf.*, 2021, pp. 1–6.
[11] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, "KAN: Kolmogorov–Arnold networks," 2024, *arXiv:2404.19756*.
[12] J. Schmidt-Hieber, "The Kolmogorov–Arnold representation theorem revisited," *Neural Netw.*, vol. 137, pp. 119–126, May 2021.
[13] A. D. Bodner, A. S. Tepsich, J. N. Spolski, and S. Pourteau, "Convolutional Kolmogorov–Arnold networks," 2024, *arXiv:2406.13155*.
[14] V. Dhiman, "Kan: Kolmogorov–Arnold networks: A review," Tech. Rep., 2024.
[15] I. Drokin, "Kolmogorov-arnold convolutions: Design principles and empirical studies," 2024, *arXiv:2407.01092*.
[16] A. Jamali, S. K. Roy, D. Hong, B. Lu, and P. Ghamisi, "How to learn more? Exploring Kolmogorov–Arnold networks for hyperspectral image classification," 2024, *arXiv:2406.15719*.
[17] S. Reddy, "Explainability and artificial intelligence in medicine," *Lancet Digit. Health*, vol. 4, no. 4, pp. e214–e215, Apr. 2022.
[18] W. Ding, M. Abdel-Basset, H. Hawash, and A. M. Ali, "Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey," *Inf. Sci.*, vol. 615, pp. 238–292, Nov. 2022.
[19] J. Amann, D. Vetter, S. N. Blomberg, H. C. Christensen, M. Coffee, S. Gerke, T. K. Gilbert, T. Hagendorff, S. Holm, M. Livne, A. Spezzatti, I. Strümke, R. V. Zicari, and V. I. Madai, "To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems," *PLOS Digit. Health*, vol. 1, no. 2, Feb. 2022, Art. no. e0000016.
[20] J. A. McDermid, Y. Jia, Z. Porter, and I. Habli, "Artificial intelligence explainability: The technical and ethical dimensions," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 379, no. 2207, Oct. 2021, Art. no. 20200363.
[21] L. Weber, S. Lapuschkin, A. Binder, and W. Samek, "Beyond explaining: Opportunities and challenges of XAI-based model improvement," *Inf. Fusion*, vol. 92, pp. 154–176, Apr. 2023.
[22] S. Makaju, P. W. C. Prasad, A. Alsadoon, A. K. Singh, and A. Elchouemi, "Lung cancer detection using CT scan images," *Proc. Comput. Sci.*, vol. 125, pp. 107–114, Jan. 2018.

[23] S. K. Thakur, D. P. Singh, and J. Choudhary, "Lung cancer identification: A review on detection and classification," *Cancer Metastasis Rev.*, vol. 39, no. 3, pp. 989–998, Sep. 2020.

[24] M. Toğaçar, B. Ergen, and Z. Cömert, "Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks," *Biocybernetics Biomed. Eng.*, vol. 40, no. 1, pp. 23–39, Jan. 2020.

[25] P. M. Shakeel, M. A. Burhanuddin, and M. I. Desa, "Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier," *Neural Comput. Appl.*, vol. 34, no. 12, pp. 9579–9592, Jun. 2022.

[26] B. K. Hatuwal and H. C. Thapa, "Lung cancer detection using convolutional neural network on histopathological images," *Int. J. Comput. Trends Technol.*, vol. 68, no. 10, pp. 21–24, Oct. 2020.

[27] A. Elnakib, H. M. Amer, and F. E. Z. Abou-Chadi, "Early lung cancer detection using deep learning optimization," *Int. J. Online Biomed. Eng.*, vol. 16, no. 6, pp. 82–94, May 2020.

[28] S. Garg and S. Garg, "Prediction of lung and colon cancer through analysis of histopathological images by utilizing pre-trained CNN models with visualization of class activation and saliency maps," in *Proc. 3rd Artif. Intell. Cloud Comput. Conf.*, Dec. 2020, pp. 38–45.

[29] A. Ben Hamida, M. Devanne, J. Weber, C. Truntzer, V. Derangère, F. Ghiringhelli, G. Forestier, and C. Wemmert, "Deep learning for colon cancer histopathological images analysis," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104730.

[30] M. A. Talukder, M. M. Islam, M. A. Uddin, A. Akhter, K. F. Hasan, and M. A. Moni, "Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning," *Expert Syst. Appl.*, vol. 205, Nov. 2022, Art. no. 117695.

[31] A. Riasatian et al., "Fine-tuning and training of densenet for histopathology image representation using TCGA diagnostic slides," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 102032.

[32] M. Masud, N. Sikder, A.-A. Nahid, A. K. Bairagi, and M. A. AlZain, "A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework," *Sensors*, vol. 21, no. 3, p. 748, Jan. 2021.

[33] S. Dzhenzher and A. Skopenkov, "A structured proof of Kolmogorov's superposition theorem," 2021, *arXiv:2105.00408*.

[34] M.-J. Lai and Z. Shen, "The Kolmogorov superposition theorem can break the curse of dimensionality when approximating high dimensional functions," 2021, *arXiv:2112.09963*.

[35] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, pp. 1–74, Mar. 2021.

[36] D. Bhatt, C. Patel, H. Talsania, J. Patel, R. Vaghela, S. Pandya, K. Modi, and H. Ghayvat, "CNN variants for computer vision: History, architecture, application, challenges and future scope," *Electronics*, vol. 10, no. 20, p. 2470, Oct. 2021.

[37] Z. J. Wang, R. Turko, O. Shaikh, H. Park, N. Das, F. Hohman, M. Kahng, and D. H. Polo Chau, "CNN explainer: Learning convolutional neural networks with interactive visualization," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 2, pp. 1396–1406, Feb. 2021.

[38] A. W. Salehi, S. Khan, G. Gupta, B. I. Alabduallah, A. Almjally, H. Alsolai, T. Siddiqui, and A. Mellit, "A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope," *Sustainability*, vol. 15, no. 7, p. 5930, Mar. 2023.

[39] M. R. Zafar and N. Khan, "Deterministic local interpretable model-agnostic explanations for stable explainability," *Mach. Learn. Knowl. Extraction*, vol. 3, no. 3, pp. 525–541, Jun. 2021.

[40] N. B. Kumarakulasinghe, T. Blomberg, J. Liu, A. S. Leao, and P. Papapetrou, "Evaluating local interpretable model-agnostic explanations on clinical machine learning classification models," in *Proc. IEEE 33rd Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jul. 2020, pp. 7–12.

[41] G. Plumb, D. Molitor, and A. Talwalkar, "Model agnostic supervised local explanations," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2018, pp. 1–10.

[42] Y. Nohara, K. Matsumoto, H. Soejima, and N. Nakashima, "Explanation of machine learning models using Shapley additive explanation and application for real data in hospital," *Comput. Methods Programs Biomed.*, vol. 214, Feb. 2022, Art. no. 106584.

[43] S. Bordt and U. V. Luxburg, "From Shapley values to generalized additive models and back," in *Proc. Int. Conf. Artif. Intell. Statist.*, Jan. 2022, pp. 709–745.

[44] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.

[45] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, and F. Shen, "Image data augmentation for deep learning: A survey," 2022, *arXiv:2204.08610*.

[46] J. Tan, J. Yang, S. Wu, G. Chen, and J. Zhao, "A critical look at the current train/test split in machine learning," 2021, *arXiv:2106.04525*.

[47] K. Sonawane and A. Rai, "LungStat: Improving lung cancer diagnostic accuracy through computer vision," in *Proc. IEMS*, Mar. 2021, pp. 64–69.

[48] N. Faruqui, M. A. Yousuf, M. Whaiduzzaman, A. Azad, A. Barros, and M. A. Moni, "LungNet: A hybrid deep-CNN model for lung cancer diagnosis using CT and wearable sensor-based medical IoT data," *Comput. Biol. Med.*, vol. 139, Oct. 2021, Art. no. 104961.

[49] A. Y. Saleh, C. K. Chin, V. Penshie, and H. R. H. Al-Absi, "Lung cancer medical images classification using hybrid CNN-SVM," *Int. J. Adv. Intell. Informat.*, vol. 7, no. 2, p. 151, Jul. 2021.

YAJNASENI DASH received the Ph.D. degree from Indian Institute of Technology, New Delhi, in 2022. She currently holds the position of an Assistant Professor with the School of Artificial Intelligence (SoAI), Bennett University. Additionally, she has been honored with the National Science Foundation (NSF) Travel Grant and the World Bank Trust Fund for Statistical Capacity Building (WB TFSCB) Grant. She has published several research papers in peer-reviewed journals and conference proceedings. Her research interests include many topics, encompassing artificial intelligence, healthcare, climate change, and computational real-world applications. She has received recognition for her contributions, including the Austin Student Travel Grant by the American Geophysical Union, the Research Scholar Travel Award from the Indian Institute of Technology Delhi, and the Best Paper Award from an IEEE Conference.

SUDHIR C. SARANGI received the M.B.B.S. degree (Hons.) from the MKCG Medical College, Berhampur, Odisha, and the M.D. degree in pharmacology and the D.M. degree in clinical pharmacology from All-India Institute of Medical Sciences (AIIMS), New Delhi. He completed three years of a Junior and Senior Residency from AIIMS, and an additional six months of a Senior Residency from the Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh. He is currently an Additional Professor with the Department of Pharmacology, AIIMS. He has authored many papers in indexed peer-reviewed journals, conference proceedings, and book chapters. He has undertaken 23 funded research projects from government agencies and institutes. He has presented his research at several international forums, several international forums (International Epilepsy Congress, in 2023, 2021, 2019, 2017, European Epilepsy Congress, in 2022, AGU 2019, Safety Pharmacology Society Annual meetings, in 2012 and 2014). His research interests include clinical pharmacology, neuropharmacology, regulatory pharmacology, pharmacogenetics, artificial intelligence in pharmacology, pharmacoeconomics, and pharmacovigilance. He has received numerous awards, including membership in the National Academy of Medical Sciences, in 2017, the best paper awards in the Indian Epilepsy Society annual conferences, in 2019 and 2023, and the Certificate of Achievement for Poster Presentations in various International Epilepsy Congresses. His Reaction Monitoring Center (AMC) under PvPI, Government of India, AIIMS.

VINAYAK GUPTA received the M.Tech. degree from the School of Computer Science and Technology, Bennett University, where he honed his expertise in artificial intelligence and machine learning. He is currently an Assistant Professor with the Department of Computer Science and Engineering, ITS Engineering College, Greater Noida. His academic journey is centered around developing ethical AI systems, ensuring transparency, and building trust-based AI models that prioritize human welfare. Through his research and projects, he aspires to contribute meaningfully to advancements in technology while fostering innovation in a collaborative environment. As an active student member of IEEE, he is deeply engaged in innovative research with a special focus on AI's transformative impact in healthcare, particularly through computer vision and explainable AI.

NAWEEN KUMAR received the Ph.D. degree in computer science and engineering from the National Institute of Technology Patna, India, in 2021. He is currently an Assistant Professor with the School of Computer Science, Engineering, and Technology, Bennett University, Greater Noida, India. His expertise spans WSN, MANET, optimization and its applications, the Internet of Things, and cloud computing. He has contributed more than 50 papers in various refereed journals and international conferences, and contributed many book chapters and books. His research interests include wireless sensor networks, the IoT, cloud computing, and optimization. He is a member of the Computer Science Teacher Association (CSTA) and the International Association of Engineers (IAENG). He has been serving as a reviewer for several prestigious journals published by IEEE, Elsevier, and Springer.

AJITH ABRAHAM received the B.Tech. degree (Hons.) in electrical and electronic engineering from the University of Calicut, in 1990, the M.S. degree from Nanyang Technological University, Singapore, in 1998, and the Ph.D. degree in computer science (artificial intelligence) from Monash University, Melbourne, Australia, in 2001. He is currently the Vice Chancellor of Sai University, Chennai, India. He was the Vice Chancellor and the Founding Dean of the School of Artificial Intelligence, Bennett University, Greater Noida; the Dean of the Faculty of Computing and Mathematical Sciences, FLAME University, Pune; and the Founding Director of the Machine Intelligence Research Laboratories (MIR Labs), USA. During the last 25 years, he has authored/co-authored more than 1,500 research publications with more than 65,000 academic citations, according to Google Scholar. He was the Chair of the IEEE Systems, Man, and Cybernetics Society Technical Committee on Soft Computing (which has more than 200 members), from 2008 to 2021. He was the Editor-in-Chief of the Engineering Applications of Artificial Intelligence (EAAI), from 2016 to 2021. He serves/served on the editorial board for more than 15 international journals indexed by Thomson ISI. He served as a Distinguished Lecturer for the IEEE Computer Society representing Europe, from 2011 to 2013.