Journal of Future Artificial Intelligence and Technologies

E-ISSN: 3048-3719

Research Article

Deep Learning-Based Cross-Cancer Morphological Analysis: Identifying Histopathological Patterns in Breast and Lung Cancer

Mezgebu Birhan Teferi [1] and Lateef Adesola Akinyemi [2,3,*]

[1] School of Electrical Engineering and Computing, Adama Science and Technology University, Adama, Ethiopia; e-mail: birhanmezgebu74@gmail.com
[2] Centre for Augmented Intelligence and Data Science, School of Computing, CSET, University of South Africa, Johannesburg, South Africa; e-mail: akinyla@unisa.ac.za
[3] Department of Electronic and Computer Engineering, Faculty of Engineering, Lagos State University, Epe, Lagos, Nigeria; e-mail: lateef.akinyemi@lasu.edu.ng
* Corresponding Author: Lateef Adesola Akinyemi

Abstract: The efficacy of Cancer treatment often varies across different types of cancers. This study aims to investigate any pattern relationship in histopathological images of different cancer types to find any potential correlation between those patterns. Using deep image analysis techniques and artificial intelligence (AI), we extract, analyze, and compare the morphological parameters of cancer images to identify potential indicators of that treatment effective for one type that might be applicable for its correspondence. This research applied advanced image analysis, artificial intelligence (AI), machine learning, and more sophisticated statistical analysis to find the required pattern relationship for those parameters. The study answers the question regarding the correlation of different measurement parameters across different varieties of cancer cells. The model achieved an impressive ROC-AUC score of 0.967, an F1-score of 0.805, and Cohen's kappa coefficient of 0.767, indicating a high level of agreement and predictive performance. The overall accuracy of the model was 81%, with both macro and weighted averages also at 81%. These results provide strong evidence of meaningful pattern relationships across different cancer types, potentially enhancing treatments' applicability across various cancers.

Keywords: Cellpose; Efficacy morphological parameters; Correlation; Histopathology; Image analysis; Machine learning; Oncology.

Received: September, 25th 2024
Revised: October, 18th 2024
Accepted: October, 23rd 2024
Published: October, 24th 2024
Curr. Ver.: October, 24th 2024

1. Introduction

Cancer remains one of the leading killer diseases worldwide, requiring the development of different treatment strategies and approaches. Currently, Ki-67 is the main proliferation marker that has been used widely for a certain type of cancer cells. One of the methods for assessing the proliferative activity of cells is the immunohistochemical detection of cell cycle-specific antigens [1]. Ki-67 is a marker that primarily indicates the proliferation rate of cells within a given sample. However, while Ki-67 is useful in identifying cell proliferation, it lacks specificity regarding different cell types and their unique responses to treatment, which limits its potential to guide precise therapeutic strategies. To overcome this limitation, analysis of additional cell-specific markers, such as CK8/18, can offer a more comprehensive understanding of the tumor microenvironment. This multi-marker approach would enable a deeper understanding of the cellular environment and treatment responses, allowing for more personalized and effective interventions. It also helps to understand the cell cycle dynamics and assess the growth rate of tumors. CK8/18 are specific markers for epithelial cells. These are used to identify the presence of epithelial cells and to diagnose epithelial tumor cells, and they basically provide the origin, cell type, and differentiation status. However, these markers lack

more issues in creating effective treatment like Ki-67 does since those have no proper correlation like Ki-67's. The main objective of this paper is to find out the relationship between these profilers. Cytokeratin (CK), an intermediate filament observed mainly in epithelial cells, is an essential cytoskeletal component involved in the fixation of the nucleus and maintenance of cell morphology[2]. CK8/18 rabbit monoclonal may be useful as a staining mask in a multiplex stain with mouse monoclonal antibody Ki-67[3].

In recent years, digital pathology and artificial intelligence (AI) have revolutionized the analysis of histopathological images, enabling faster, more accurate diagnoses. Rapid advances in digital histopathology have also allowed the extraction of clinically relevant information embedded in tumor slides by applying machine learning and artificial intelligence methods, capitalizing on recent advancements in image analysis via deep learning[4]. Histopathological images are considered an appropriate tool for diagnosing cancer, which is tedious and error-prone if done manually[5]. With feature extraction, deep learning, and digital pathology, disease diagnosis and prognosis could be more quickly and accurately determined, potentially revolutionizing modern healthcare and pathology practice[6]. Digital pathology is vital in advancing and optimizing histology processes to improve efficiency and reduce turnaround times[7]. Cellular segmentation has good out-of-the-box performance on various image types[8]. Multiple studies showed that luminal A tumors are usually smaller in size, while basal-like type tumors are mostly larger and more likely to have a regular shape[9]. Cases that showed the proper tumor morphology and were positive for P63 and CK 5/6 immunostains were considered squamous cell carcinoma[10].

This study aims to discover the possibilities of using deep learning to analyze histopathological images and explore the relationship between CK8/18 images. By developing an AI-based image analysis framework, we seek to quantify similarities between CK8/18 image types and evaluate their potential for predicting treatment outcomes. This approach would give a personalized expedite the time for a treatment. The contributions of this study are twofold:

1. Developing an AI-based image analysis methodology to identify and quantify similarities between CK8/18 image types.
2. Providing a framework for leveraging these similarities to predict the potential efficacy of cross-cancer treatments, offering new insights into personalized medicine and oncology research.

Given the growing incidence of cancer globally, especially in regions like sub-Saharan Africa, where access to healthcare resources is limited, this research also highlights the importance of utilizing data-driven solutions to address disparities in cancer diagnosis and treatment. This study highlights the need for both parties to accelerate their efforts to save lives currently lost due to the absence of these critical resources. At 128.2 per 100,000 people, cancer incidence in the 46 mostly low-income sub-Saharan African countries exceeds the average of 115.7 per 100,000 people in low and 108.5 per 100,000 people in medium Human Development Index (HDI) regions globally[11]. This alarming rate indicates that integrating advanced technologies like AI into the healthcare industries, especially for regions lacking many resources. The use of machine learning models in solving data-driven problems is growing significantly, resulting in higher prediction performance expectations[12]. This paper has applied different analysis methods to rectify and prepare the dataset for training using the ML method.

The paper is organized as follows: Section 1 talks about the introduction regarding the subject matter alongside the study's contributions. Section 2 discusses the related works. Section 3 presents the methodology employed in this study. Section 4 presents and discusses the numerical results and interpretation of the results. Section 5 concludes the paper alongside future studies.

2. Related works

Statistical calculations were performed with JMP 14 software (SAS Institute Inc., NC, USA). Contingency tables and the chi2-test were performed to search for associations between CK18 and tumor phenotype[13]. A total of 11,952 (82.0%) of 14,579 tumor samples were interpretable in our TMA analysis. The remaining 2627 (18.0%) samples were not analyzable due to the lack of unequivocal tumor cells or loss of the tissue spots during the technical procedures[13]. The triple test includes clinical examination, imaging (mammogram or ultrasonogram), and tissue sampling (fine needle aspiration cytology (FNAC) and core needle

biopsy). These tests happen sequentially, and their findings must support the final diagnosis for accurate patient management [14]. Cancers originating from different organs and cell types are known, with the most common ones being breast, lung, colorectal, prostate, and stomach[15]. That was with exception as there are a lot of places where cancer can originate and can't be detected as easily as possible using the previous methods, which have been developed mainly for known types. The significant correlation between prognosis and stable expression in biopsy specimens suggests the usefulness of CK18 in selecting treatment strategies for OSCC[2]. That was the most recent approach, which has done significant work in this field using a different methodology with an accuracy of 56% correlation. Color segmentation processes were applied to the original image based on the foreground image's detected hue and intensity distributions [16]. The potential to use quantitative image analysis and artificial intelligence is one of the driving forces behind digital pathology[17]. Deep learning models face several challenges. One major issue is the need for large annotated datasets, often difficult to obtain in biological research[18]. Additionally, the interpretability of deep learning models remains a concern, as their "Blackbox" nature can obscure the understanding of model decisions[19]. Recent efforts have focused on developing explainable AI methods to address these challenges[20].

One key area of focus has been using convolutional neural networks (CNNs) to analyze medical images, such as histopathological slides and radiological scans. CNNs have demonstrated the ability to accurately detect and classify various types of cancer, including breast, lung, and prostate cancer[21]. Beyond individual cancer types, researchers have also investigated the potential of AI-based image analysis to identify cross-cancer patterns. For example, a study by [22] utilized a CNN to analyze lung cancer histopathological images and found that the model could distinguish between different subtypes of lung cancer and identify shared features across various cancer types. This suggests that AI-based image analysis may be able to uncover common molecular and morphological characteristics that cut across different cancer diagnoses. Similarly, [23] explored the use of deep learning for the analysis of brain tumor MRI scans, demonstrating the ability of their model to differentiate between glioblastoma and lower-grade gliomas while also identifying shared imaging features among different brain tumor types. This work highlights the potential of AI-based image analysis to provide insights into the underlying biology and cross-cutting patterns of various cancers. As AI-powered image analysis continues to evolve, researchers are also exploring the development of interpretable and explainable models, which can shed light on the algorithms' specific features and decision-making processes [24].

Although much work has been done in this field, there is still a research gap specifically addressing the correlation analysis of CK8/18 markers. Most studies have focused on analyzing pathological images to treat cancer using the Ki-67 marker, widely known and used globally. While there has been some research on CK8/18 markers, these studies often lack accuracy and fail to provide a user-friendly modeling framework for non-technical workers, an area that requires special attention. According to[2], the highest correlation accuracy achieved so far is 56%, which is too low to make reliable decisions based on the data. While AI models often rely on probabilities, even a high probability can leave room for errors.

3. Proposed Method

This research employs a straightforward method that begins with data collection (images), then feature extraction and measurement of relevant parameters. Various statistical analyses are then applied using AI models, analysis software, and other statistical tools. Figure 1 illustrates the overall flow of the methodologies used in this paper.

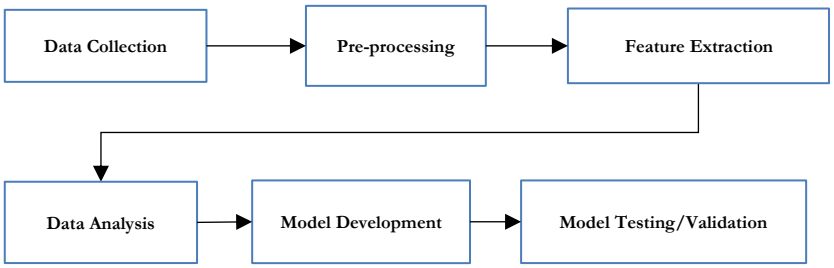

Figure 1. Workflow of Data-Driven Model Development

While the main steps align with standard practices, there are some key deviations from traditional machine learning models. The critical focus here is on feature engineering, which introduces a slight modification to the usual approach. Instead of relying solely on primary features, detailed pattern detection combines features and uncovers relationships through mathematical formulas. This unique aspect sets this research apart from previous studies in the field. Moreover, the methodology is clear and accessible, making it easy for those interested in learning more about this domain.

3.1. Data Collection

Histopathological images from lung and breast cancers have been collected. These images are categorized into two groups, each being divided into three groups. These groups contain three distinct individual images, a total of 18 images. These images are raw images used for analysis throughout the whole process in this study. The images are collected from different medical institutions on open-source Git Hub(https://github.com/Masklom/data). These images will be classified into six categories, and from those 18 images, around six features will be extracted to get the required correlations. After going through some preprocessing stages, we can easily find any pattern relationship from those features.

Table 1. Categorization of Histopathological Images by CK/Ki Markers and Treatment Status.

Main Group	Subgroup	Number of Images	Features to be Extracted
CNT (Control)	CK+Ki+_CNT	3	Brightness, Area, Circularity, Centroids, Amorphism
	CK+Ki-_CNT	3	Brightness, Area, Circularity, Centroids, Amorphism
	CK-Ki-_CNT	3	Brightness, Area, Circularity, Centroids, Amorphism
Treat (Treated)	CK+Ki+_Treat	3	Brightness, Area, Circularity, Centroids, Amorphism
	CK-Ki+_Treat	3	Brightness, Area, Circularity, Centroids, Amorphism
	CK-Ki-_Treat	3	Brightness, Area, Circularity, Centroids, Amorphism

Table 1 shows that the first group is controlled, which refers to drug-free. In addition, the second group is treated, which in this case refers to adding chemicals. The subgroups are separated according to whether or not CK and KI markers are present. The marker is shown as present when the plus sign is present and absent when the minus sign is present. Since we are interested in examining any potential correlation between these features from all images, the features that were extracted from each image are identical. As explained in the subsequent sections, all features have their description.

3.2. Preprocessing

This is the process of collecting images to be processed and analyzed. This stage includes normalization and noise reduction to ensure consistency and quality for analysis. Maintaining the highest resolution, proper aspect ratio and better depth of the images makes it easier for analysis and, more importantly, for visualization of the image with clear understanding. We can use the OpenCV library to implement these steps. Figure 2 shows how the whole step is being implemented.

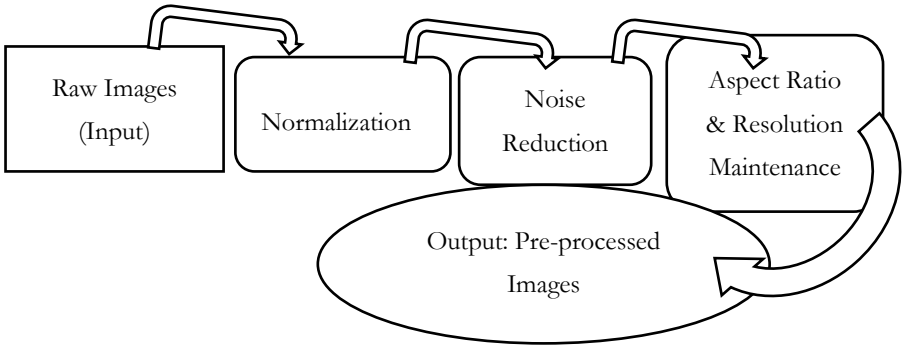

Figure 2. Image preprocessing flow diagram

The raw image is loaded from a specified file path in the image preprocessing stage. The pixel intensities of the image are then normalized into the range [0, 1]. To reduce noise in the image, a Gaussian filter is applied, which helps improve the image's smoothness. After that, the image's aspect ratio is calculated to ensure that the proportions are maintained when the image is resized based on the desired output size. The processed image will be saved in the specified directory if the save option is enabled. Finally, this preprocessed image is ready for further analysis.

3.3. Segmentation

Convolutional Neural Networks (CNNs), widely used for image processing tasks, play a crucial role in advanced models like Cellpose, which we use here for cell segmentation**.** This process is particularly crucial in analyzing medical images like X-rays, where extracting tissues, organs, and pathological structures is imperative[25]. This is the process where the processed images will be segmented to identify each cell. Segmenting is key here because that is where the feature extraction starts. It was time to pinpoint the required region after preprocessing the image, such as scaling, normalizing, and resizing. Segmentation is one of the most advanced deep learning fields where pre-trained models can be tuned to meet our objectives. Different segmentation models have been developed specifically for this purpose, but in this case; to get the best out, we have tuned some parameters. In this case, we use Cellpose, which gives us a range of parameters to tune according to our needs. All parameters used to perform image segmentation can be adjusted based on our needs. Parameter settings are presented in Table 2.

Table 2. Cellopse parameters.

Parameter	Value
What to segment	cytoplasm
Channel to segment	blue
Nuclear Channel	none
Avg cell diameter	24
Model	cyto3

This process is important for the subsequent steps as it provides a quantitative output of each cell. This process needs a very highly calibrated way of getting the highest accuracy or detecting with much higher precision. It has been completed using the Cellpose library with the cyto3 model. As this is the most advanced model at the time it gave a very good result in terms of getting better accuracy. Sample segmentation results are presented in Figure 3.

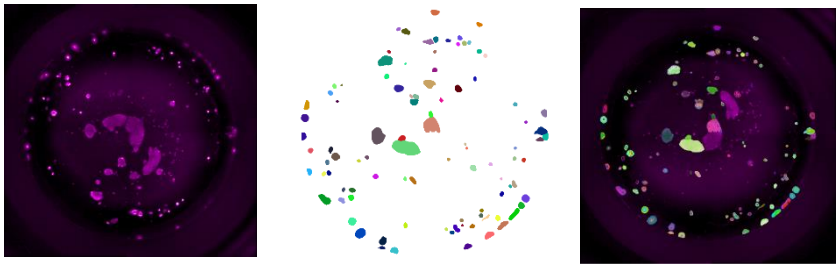

Figure 3. Original, segmented, and merged images of CK8/18 images

3.4. Feature Extraction

The segmented image will then go for the extraction of important features that are used for analysis and the planning of appropriate models. By taking different structures into account, different parameters were extracted. These are the six main parameters extracted from the segmented image. The skimage library which is integrated with Python, has been used for extracting these parameters. This is the most crucial step for calibrating the goal of the result. This means that if we have properly extracted all features appropriately, then all the rest of the tasks will be easier to analyze. Many features can be extracted either directly or indirectly.

We can extract as many parameters from each cell as we want, but for this paper, we will only use those that are very important for demonstration purpose. For example, the segmented image has extracted area, perimeter, centroid_x, centroid_y, brightness, amorphism, and circularity features. The first four features are extracted directly from those features, while the last two can be extracted indirectly. All these parameters can be expressed as follows:

- Brightness: The mean intensity value of each segmented object. For a segmented object, brightness can be calculated by summing all pixels' intensity values and then dividing by the total number of pixels. Equation (1) shows how to compute the brightness of each segment.

$$\text{Brightness} = \frac{\sum_{i=1}^{N} I_i}{N} \tag{1}$$

  where I_i is the intensity of pixel i, and N is the total number of pixels in the object.
- Area: The area of each segmented object.
- Centroid X, Centroid Y: The x,y-coordinate of the centroid of each segmented object is computed as follows in Equations (2) and (3).

$$X - \text{coordinate: } X = \frac{\sum_{i=1}^{N} x_i}{N} \tag{2}$$

$$Y - \text{coordinate: } Y = \frac{\sum_{i=1}^{N} y_i}{N} \tag{3}$$

  where x_i and y_i are the coordinates of pixel i, and N is the total number of pixels in the object.
- Amorphism: A measure of the irregularity of each segmented object, calculated as the standard deviation of distances from the centroid to the vertices of a pentagon. It can be computed using an indirect method with the following steps: First, identify or estimate the vertices of the inscribed pentagon. Then, measure the distance from the centroid to each vertex. Finally, calculate the standard deviation of these distances. A higher standard deviation indicates a more irregular shape. See sample in Figure 4.

Figure 4. Amorphism Assessment Using Standard Deviation of Centroid-to-Vertex Distances

- Circularity: A measure of how circular each segmented object is, In the same way as amorphism, circularity can be calculated as Equation 4.

$$\text{Circularity} = \frac{4\pi * \text{Perimeter}^2}{\text{Area}} \tag{1}$$

  where the perimeter is the total length of the boundary of the object. A circularity of 1 indicates a perfect circle, while values less than 1 indicate more elongated or irregular shapes.

3.5. Statistical Analysis

3.5.1. Correlation

Finding a pattern that correlates one group with another is the main objective here, and this has been done by using the extracted features above and finding different patterns. This was the main concern for this research paper as it will clearly distinguish extracted parameters and show a clear relationship between different variables, extracted by using segmentation. This analysis shows us a relationship that will help us infer or project some generalizations for our problems. Beyond correlation analysis, other methods have been used, such as regression, ANOVA, and cluster analysis. As this is the last step in developing a model, it has gone

through lots of analysis methods. As it is shown in the graphs in Figure 5, we can observe their clear, distinct relationship for different groups and parameters.

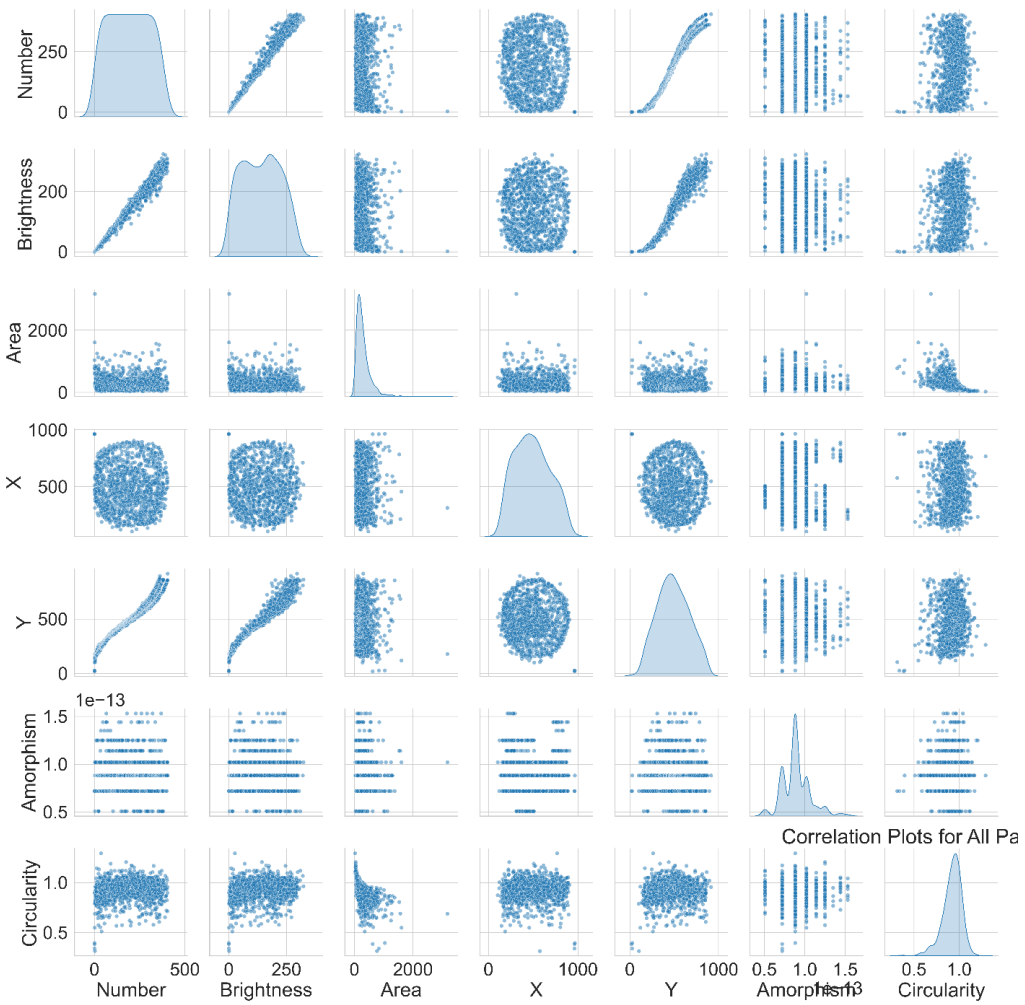

Figure 5. Correlation analysis graph

3.5.2. Cluster Analysis

Clustering is one of the most important analysis methods for this purpose as the cell is spread evenly, so it is one technique that helps us cluster into different groups. As most biological-related things fall under classification science, this was the best approach to deal with such kinds of analysis. There are six groups in this category, each of which has unique features based on the above parameters. We can easily observe how they exhibit this nature in Figure 6

3.6 Deep Learning Analysis

After testing using traditional analysis methods to find if there is any pattern relationship in each group, we got results that did not satisfy our expectations. For that reason, CNNs have been employed for deeper analysis. The model has been developed as follows and summarized in Table 3. Even though our main target is finding the pattern relationship between each group, the paper has gone through more advanced ways to develop a generic model that will help classify with minimum time and cost resources. The model development starts with adjusting the data and ends with the final training using different methods.

Data preprocessing is an important step for getting the data ready to train. The data was not in the proper way to fit the training model, and more importantly, there was an issue with balance where the data was not equal for all groups. Synthetically generated data has been employed to correct this. These are the steps that have been employed to rectify the dataset or to make corrections.

- Feature Encoding: The data was simply divided only into six groups, and it was not numerical data so it has been appropriately labeled for training purposes.
- Balancing the dataset: The dataset for each group is not distributed evenly, so it has been adjusted using synthetically generated data.
- Parameters selection: In this model development, six features or parameters were used to train the model. Among those parameters, some are more important than others, and here is the rank for each variable. Figure 7 shows how important each parameter is for the model to be trained.

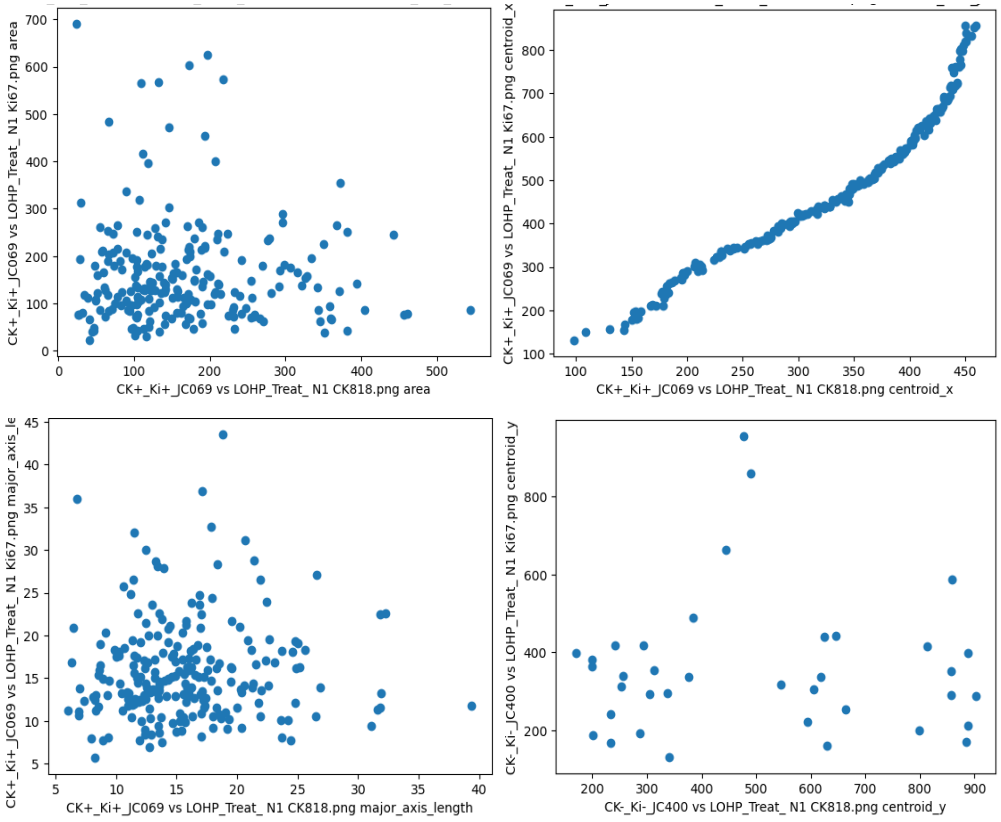

Figure 6. Scatter plots for different parameters

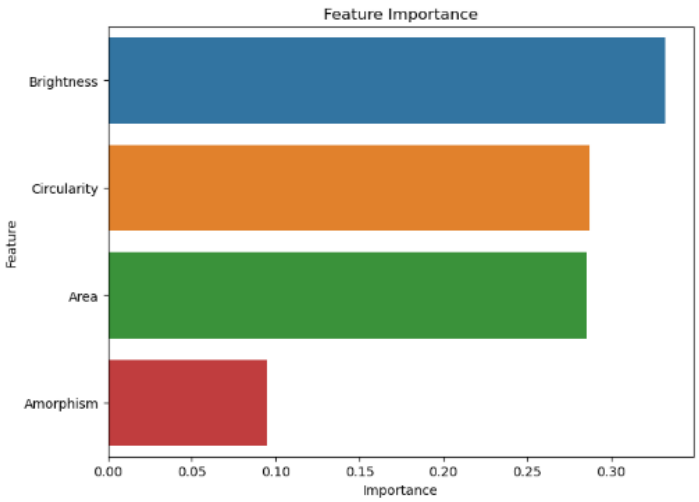

Figure 7. Feature importance diagram

This simplified model, implemented using Keras, is primarily designed for analysis purposes, aiming to identify potential relationships that traditional methods may have overlooked. The CNN architecture is specifically used for extracting spatial features from the

input images, such as textures, brightness, and shape, which are crucial for image-based tasks like segmentation. While a slight improvement in accuracy was observed, the results did not fully meet expectations. As a result, further analysis is required, particularly in exploring deeper relationships among the extracted features, to enhance the model's ability to identify key patterns. Efforts to develop a more accurate model and improve performance are discussed in the following sections.

Table 3. CNN Model Configuration.

Layer (Type)	Output Shape	Parameters	Description
Input Layer (Image Input)	(None, image_height, image_width, 3)	0	Input image with 3 color channels (RGB)
Conv2D (32 filters)	(None, new_height_1, new_width_1, 32)	896	Convolutional layer with 32 filters, kernel size (3, 3)
MaxPooling2D	(None, pooled_height_1, pooled_width_1, 32)	0	Pooling layer to reduce spatial dimensions
Conv2D (64 filters)	(None, new_height_2, new_width_2, 64)	18,496	Convolutional layer with 64 filters, kernel size (3, 3)
MaxPooling2D	(None, pooled_height_2, pooled_width_2, 64)	0	Pooling layer to further reduce spatial dimensions
Flatten	(None, flattened_size)	0	Flattening the 3D output to feed into the dense layer
Input Layer (Features Input)	(None, 4)	0	Input for extracted features (e.g., area, centroid)
Dense (32 units)	(None, 32)	160	Fully connected dense layer for manually extracted features
Concatenate	(None, concatenated_size)	0	Concatenation of CNN features and manual feature vectors
Dense (Output Layer)	(None, 6)	concatenated_size * 6 + 6	Final output layer with softmax activation, 6 categories

3.7. Pattern detection and Model building

This model was developed using the XGB method. It has important features as it handles a lot of decision trees and is more suitable for large groups like this, and it has become the most important in identifying relationships. Other models have been utilized to get the best result, but none of them were better than the XGB model. The nature of the dataset, where there are many groups to be classified, makes this model more effective than any other available model. By employing this method, the model can only predict 56%.

After going through a deeper analysis using the above model and the first extracted features, I found low accuracy. As a better solution, further re-engineering was done using the previously extracted features. The features were engineered in this way: Brightness_to_Area_Ratio(£), Distance_to_Center(Ƴ), and Amorphism_Circularity(π). Each of them can be calculated as follows in Equations (5), (6), and (7).

$$£ = \frac{\alpha}{A} \tag{5}$$

where α is the brightness of the region calculated as the mean of intensity from (1), A is the area of the region and $£$ is the ratio of brightness to area.

$$Ƴ = \sqrt{(X - \mu_x)^2 + (Y - \mu_y)^2} \tag{6}$$

where X and Y are x, y-coordinates of the centroid (geometric center) of the region calculated using Equations (2) and (3) in section 3.4, while, μ_x, μ_y represent the mean coordinates and $Ƴ$ The Euclidean distance from the region's centroid to a reference point.

$$\pi = \text{Amorphism} * \text{Circularity} \tag{7}$$

Where Amorphism and Circularity can be calculated from section 3.4, while π is the product of amorphism and circularity.

The new features have been shown to reveal more distinct pattern relationships compared to the originally extracted features, enhancing the model's ability to identify significant correlations between variables. In this analysis, the data was split into training and testing sets, with 80% of the data allocated for training and 20% for testing, ensuring that the model can generalize well. After the dataset was divided, both Random Forest and XGBoost models were trained, resulting in a notable improvement in accuracy and prediction performance. The Random Forest classifier provided insights into feature importance, while XGBoost helped handle more complex data structures, leading to more refined classification outcomes. The model development is summarized in Table 4, and a detailed XGB model summary is provided in Table 5.

Table 4. Model Development Summary.

Step	Description
Feature Engineering	Added Brightness_to_Area_Ratio, Distance_to_Center, and Amorphism_Circularity_Product. Dropped unused columns.
Train-Test Split	Split data into 80% training and 20% testing with train_test_split.
SMOTE	Applied SMOTE to balance the training data (over-sampling minority classes).
Scaling	Standardized the features using StandardScaler.
Model Used	RandomForestClassifier with default parameters.
Evaluation	Calculated metrics: ROC-AUC, F1-score, Cohen's Kappa, Confusion Matrix.

Table 5. XGB Model Summary.

Attribute	Details
Number of Trees	100
Max Depth	Unlimited
Feature Importance	Random Forest naturally ranks feature importance.
Handling Imbalance	Handled by SMOTE before training the model.
Accuracy	Measured using precision, recall, and f1-score for classification performance.
Feature Scaling	Standardized using StandardScaler.

3.8. Model Testing/Validation

This is the final step for the model to validate whether it is acting as expected. Some metrics were used to validate this. In this research, the Confusion Matrix and ROC-AUC is used. Confusion Matrix clearly shows true positive, false negative, true negative, and false positive values for all of the six groups. All of them help to understand the predicted and actual values of the model. All these metrics are True Positive (TP), which indicates correctly predicted positive cases, False Negative (FN) which shows incorrectly predicted negative cases, True Negative (TN) which tells correctly predicted negative cases, and False Positive (FP) which shows Incorrectly predicted positive cases. These metrics show the model's performance in tabular form. Whereas ROC-AUC is a graphical representation that directly shows what is happening in the overall metrics that have been discussed clearly. It just uses the visual form to signify how each group's values are performing well

4. Results and Discussion

The numerical results section will present the findings of the statistical analyses, including:

4.1. Confusion Matrix

The confusion matrix provides a detailed breakdown of the model's performance for each class. The results in Figure 8 show that the model performs well across all classes, with a high number of true positives and a low number of false positives and false negatives.

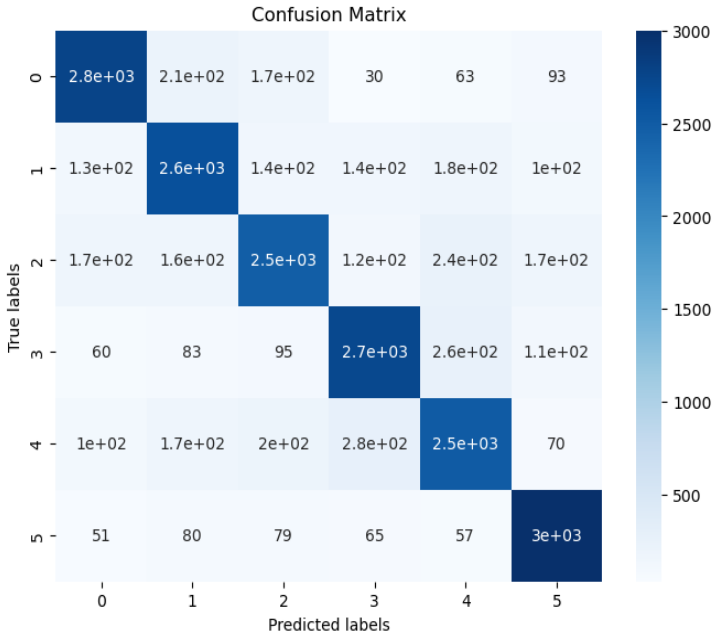

Figure 8. Confusion Matrix

4.2. Evaluation Metrics

The classification report provides a summary of the model's performance for each class. The metrics used are:

- Accuracy: This is another evaluation metric that measures how accurate our model is performing in classifying those individual groups. It is also different from precision in which it includes both the positive and negative portion It can also be classified mathematically as follows in Equation (8).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FT + FN} \tag{8}$$

- Precision: The ratio of true positives to the sum of true positives and false positives. It measures the model's ability to correctly identify instances of a particular class. Equation (9) shows how to compute the precision.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{9}$$

- Recall: The ratio of true positives to the sum of true positives and false negatives. It measures the model's ability to correctly identify all instances of a particular class. Equation (10) shows how to calculate the Recall

$$\text{Recall} = \frac{TP}{TP + FN} \tag{10}$$

- F1-score: The harmonic mean of precision and recall. It provides a balanced measure of both precision and recall. F1-score calculated using Equation (11).

$$\text{F1}_{\text{score}} = 2 * \frac{(\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}} \tag{11}$$

The results show in Table 6 that the model performs well across all classes, with precision, recall, and F1-score values ranging from 0.74 to 0.90. The support values indicate that the classes are relatively balanced, with approximately 3330 instances in each class.

Accuracy and Macro Average: The model's accuracy is 0.81, indicating that it correctly classifies approximately 81% of the instances. The macro average values for precision, recall, and F1-score are also 0.81, indicating that the model performs consistently across all classes.

The F1-score is 0.805 for macro-average case, which is a good indicator of the model's performance. It considers both precision and recall, providing a balanced measure of the model's ability to classify instances correctly.

Table 6. Evaluation Metrics for each group.

Groups	Precision	Recall	F1-Score	Support
CK+Ki+_CNT	0.85	0.83	0.84	3334
CK+Ki+_Treat	0.79	0.79	0.79	3329
CK-Ki+_CNT	0.78	0.74	0.76	3334
CK-Ki+_Treat	0.81	0.82	0.81	3338
CK-Ki-_CNT	0.76	0.75	0.76	3331
CK-Ki-_Treat	0.84	0.90	0.87	3334

Table 7. Summarizes both the macro-average and weighted-average results for the model.

Metric	Macro-Average	Weighted-Average
Precision	0.805	0.818
Recall	0.805	0.820
F1-Score	0.805	0.810
Accuracy	-	0.810

The ROC-AUC score is 0.967, which is very high. This indicates that the model can distinguish between the classes very effectively. The ROC-AUC value is calculated using Equation (12), and the graph is presented in Figure 9.

$$\mathrm{ROC}-\mathrm{AUC}\ =\int_0^1 \mathrm{TPR}(T)d\big(\mathrm{FPR}(T)\big) \tag{12}$$

where TPR is the true positive rate and FPR is the false positive rate.

This study also includes an additional matrix, Cohen's Kappa Coefficient, calculated using Equation (13). Cohen's kappa coefficient result is 0.767, indicating a high agreement level between the predicted and actual classes.

$$K=\frac{(p_o-p_e)}{1-p_e} \tag{13}$$

where p_o is the observed agreement and p_e is the expected agreement.

4.3 Ablation Study Results

The ablation study provides important insights into the relative importance of various components in our pipeline, particularly feature extraction and pattern detection. By systematically removing or altering these components, we observed significant changes in performance, underscoring their critical role. The results of the ablation study tests are presented in Table 8.

Table 8. Ablation Study Results.

Component	Accuracy	F1-Score	ROC-AUC
No Pattern Detection	0.56	0.61	0.87
Simplified Segmentation	0.65	0.68	0.80
No Statistical Analysis	0.76	0.78	0.92
Proposed Model/All component	0.81	0.805	0.967

The analysis reveals that the newly derived features, such as Brightness-to-Area Ratio, Distance-to-Center, and Amorphism-Circularity, exhibit strong positive correlations with CK8/18, suggesting that CK8/18 could be a more reliable marker for these parameters compared to Ki67. These features are crucial for distinguishing between different morphological

patterns in the dataset, as indicated by their high correlation with CK8/18. Meanwhile, Mean Intensity and Eccentricity demonstrate moderate correlations, showing some level of agreement between CK8/18 and Ki67, though these features do not have as strong an influence as the former group. This suggests that while these features contribute to the model's predictive power, they may not be as critical in distinguishing between different cancer types.

In contrast, most of the other primarily extracted parameters show weak correlations, indicating that CK8/18 may not be a reliable marker for these features compared to Ki67. This implies that these features contribute less to the model's overall performance and require further refinement. Overall, the model achieves strong performance, reflected in an ROC-AUC score of 0.9673, an F1-score of 0.8053, and a Cohen's kappa coefficient of 0.767. The model also achieved an accuracy of 0.81, with both macro and weighted averages confirming this consistency.

The ablation study results in Table 8 further underscore the importance of these features. When pattern detection is removed, model performance is substantially declined (accuracy drops to 0.56), demonstrating that the strong correlations derived from features like Brightness-to-Area Ratio and Distance-to-Center rely heavily on pattern detection. Similarly, when segmentation is simplified, the performance drops (accuracy falls to 0.65), suggesting that precise segmentation is vital for features like Mean Intensity and Eccentricity, which show moderate correlations. When statistical analysis is omitted (accuracy of 0.76), the relatively smaller drop in performance implies that while statistical analysis refines the model's ability to interpret weakly correlated features, it is less crucial than pattern detection or segmentation in the overall model pipeline. This shows that the ablation study results align well with the observed correlations of the features, highlighting the critical role of pattern detection and segmentation in achieving the model's strong performance metrics.

6. Conclusions

To sum up, this research has shown how deep image analysis and artificial intelligence can be used to find and measure morphological similarities between various cancer types. Using sophisticated artificial intelligence models and statistical techniques to analyze histopathological images, we have found that re-engineered parameters like brightness-to-area ratio, distance to center, amorphism, and circularity have strong positive correlations. Suggesting the potential for cross-cancer treatment insights. These results suggest that treatments that work for one type of cancer may work for other types as well, laying the groundwork for more individualized and effective cancer treatments. In addition to speeding up the analysis process, AI in this study has revealed patterns and relationships that might not have been obvious using more conventional techniques. Some morphological parameters showed strong correlations between each group of CK8/18 staining, indicating that these markers could be useful in cross-comparison research. Future research should enlarge the dataset to include a wider range of cancer types and more comprehensive image samples to validate the results further. Additionally, experimenting with various image processing methods and adding more AI models could improve the analysis's robustness and accuracy.

Translating this research into useful oncology solutions will require examining the clinical applicability of these findings in real-world treatment scenarios. Overall, this work adds to the expanding corpus of research in artificial intelligence-assisted medical image analysis and emphasizes how cross-cancer morphological analysis may be used to develop novel therapeutic strategies for cancer patients. We can get closer to developing more efficient, individualized cancer treatments that enhance patient outcomes by carrying out more research and improvement of these methods.

Author Contributions: Conceptualities: M.B.T and L.A.A.; Methodology: M.B.T.; Software: M.B.T Validation: M.B.T. and L.A.A.; Formal analysis: M.B.T.; Investigation: M.B.T and L.A.A.; Resources: M.B.T and L.A.A.; Data curation: M.B.T; Writing original draft preparation: M.B.T.; Writing-review and editing: L.A.A.; Visualization: M.B.T.; Supervision: L.A.A.; Project administration: L.A.A.

Funding: No funding body is responsible for this research.

Data Availability Statement: The data supporting the findings of this study were collected privately and are not publicly available due to privacy and ethical restrictions. Access to the data may be available from the corresponding author upon reasonable request and with appropriate permissions. More importantly, the datasets will be uploaded on GitHub shortly.

Acknowledgments: The authors acknowledge Dr Kasu Belay for providing the datasets for this study and the support received from colleagues who offer technical advice.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] M. Juríková, Ľ. Danihel, Š. Polák, and I. Varga, "Ki67, PCNA, and MCM proteins: Markers of proliferation in the diagnosis of breast cancer," *Acta Histochem.*, vol. 118, no. 5, pp. 544–552, Jun. 2016, doi: 10.1016/j.acthis.2016.05.002.

[2] T. Makino *et al.*, "Cytokeratins 18 and 8 are poor prognostic markers in patients with squamous cell carcinoma of the oesophagus," *Br. J. Cancer*, vol. 101, no. 8, pp. 1298–1306, Oct. 2009, doi: 10.1038/sj.bjc.6605313.

[3] Biocare Medical, "Ki-67 + CK7/8/18," *Biocare*, 2023. https://biocare.net/product/ki-67-ck7818/ (accessed Aug. 23, 2024).

[4] J. Rosenthal *et al.*, "Building Tools for Machine Learning and Artificial Intelligence in Cancer Research: Best Practices and a Case Study with the PathML Toolkit for Computational Pathology," *Mol. Cancer Res.*, vol. 20, no. 2, pp. 202–206, Feb. 2022, doi: 10.1158/1541-7786.MCR-21-0665.

[5] I. Chhillar and A. Singh, "A feature engineering-based machine learning technique to detect and classify lung and colon cancer from histopathological images," *Med. Biol. Eng. Comput.*, vol. 62, no. 3, pp. 913–924, Mar. 2024, doi: 10.1007/s11517-023-02984-y.

[6] M. Jagannath, "Editorial: Feature extraction and deep learning for digital pathology images," *Front. Signal Process.*, vol. 3, Oct. 2023, doi: 10.3389/frsip.2023.1296745.

[7] M. Boktor, J. E. D. Tweel, B. R. Ecclestone, J. A. Ye, P. Fieguth, and P. Haji Reza, "Multi-channel feature extraction for virtual histological staining of photon absorption remote sensing images," *Sci. Rep.*, vol. 14, no. 1, p. 2009, Jan. 2024, doi: 10.1038/s41598-024-52588-1.

[8] O. M. Burciu, I. Sas, T.-A. Popoiu, A.-G. Merce, L. Moleriu, and I. M. Cobec, "Correlations of Imaging and Therapy in Breast Cancer Based on Molecular Patterns: An Important Issue in the Diagnosis of Breast Cancer," *Int. J. Mol. Sci.*, vol. 25, no. 15, p. 8506, Aug. 2024, doi: 10.3390/ijms25158506.

[9] G. Cserni, R. Bori, É. Ambrózay, and O. Serfőző, "Histological Patterns and Mammographic Presentation of Invasive Lobular Carcinoma Show No Obvious Associations," *Cancers (Basel).*, vol. 16, no. 9, p. 1640, Apr. 2024, doi: 10.3390/cancers16091640.

[10] M. A. ALQudah *et al.*, "Epidemiology and histopathological classification of lung cancer: A study from Jordan, retrospective observational study," *Ann. Med. Surg.*, vol. 65, May 2021, doi: 10.1016/j.amsu.2021.102330.

[11] H. D. Larkin, "Cancer Deaths May Double by 2030 in Sub-Saharan Africa," *JAMA*, vol. 327, no. 23, p. 2280, Jun. 2022, doi: 10.1001/jama.2022.10019.

[12] M. Banda, E. K. Ngassam, and E. Mnkandla, "Enhancing Classification and Prediction through the Application of Hybrid Machine Learning Models," in *2024 IST-Africa Conference (IST-Africa)*, May 2024, pp. 1–12. doi: 10.23919/IST-Africa63983.2024.10569590.

[13] A. Menz *et al.*, "Diagnostic and prognostic impact of cytokeratin 18 expression in human tumors: a tissue microarray study on 11,952 tumors," *Mol. Med.*, vol. 27, no. 1, p. 16, Dec. 2021, doi: 10.1186/s10020-021-00274-7.

[14] R. Mohan *et al.*, "Correlation of Histopathology and Radiological Findings Among the Diverse Breast Lesions in a Tertiary Care Centre," *Cureus*, Jan. 2024, doi: 10.7759/cureus.52097.

[15] A. Menon, P. Singh, P. K. Vinod, and C. V. Jawahar, "Exploring Histological Similarities Across Cancers From a Deep Learning Perspective," *Front. Oncol.*, vol. 12, Mar. 2022, doi: 10.3389/fonc.2022.842759.

[16] P.-W. Huang *et al.*, "Deep-learning based breast cancer detection for cross-staining histopathology images," *Heliyon*, vol. 9, no. 2, p. e13171, Feb. 2023, doi: 10.1016/j.heliyon.2023.e13171.

[17] P. Bankhead, "Developing image analysis methods for digital pathology," *J. Pathol.*, vol. 257, no. 4, pp. 391–402, Jul. 2022, doi: 10.1002/path.5921.

[18] V. Gulshan *et al.*, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA*, vol. 316, no. 22, p. 2402, Dec. 2016, doi: 10.1001/jama.2016.17216.

[19] C. McQuin *et al.*, "CellProfiler 3.0: Next-generation image processing for biology," *PLOS Biol.*, vol. 16, no. 7, p. e2005970, Jul. 2018, doi: 10.1371/journal.pbio.2005970.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[21] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017, doi: 10.1038/nature21056.

[22] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, no. 23, Jun. 2000, doi: 10.1161/01.CIR.101.23.e215.

[23] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic Brain Tumor Segmentation Using Convolutional Neural Networks with Test-Time Augmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 2019, pp. 61–72. doi: 10.1007/978-3-030-11726-9_6.

[24] W. Yang *et al.*, "Survey on Explainable AI: From Approaches, Limitations and Applications Aspects," *Human-Centric Intell. Syst.*, vol. 3, no. 3, pp. 161–188, Aug. 2023, doi: 10.1007/s44230-023-00038-y.

[25] K. Pyar, "Segmentation Performance Analysis of Transfer Learning Models on X-Ray Pneumonia Images," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 64–74, Jun. 2024, doi: 10.62411/faith.2024-10.