International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Multiple Cancer Types and Subtype Classification using Deep Learning

Rutvi Sindha, Vidhi Vansiya, Dhara Parikh

[12]*Student, Information Technology Department, Krishna School of Emerging Technology & Applied Research, KPGU University, Varnama, Vadodara, Gujarat, India*

[3]*Assistant Professor, Department of Information Technology and Engineering, Krishna School of Emerging Technology & Applied Research, KPGU University, Varnama, Vadodara, Gujarat, India*

Abstract - Cancer classification is crucial for effective treatment planning, especially given the diversity of cancer types and subtypes that pose unique diagnostic challenges. This study presents a deep learning-based system for multi-cancer classification, utilizing a comprehensive dataset comprising over 130,000 images across 8 main cancer types and 26 subtypes. Our approach leverages Convolutional Neural Networks (CNNs) to automatically extract intricate features from cancer images, enhancing classification accuracy across various cancer types, including Acute Lymphoblastic Leukemia, Brain Cancer, Breast Cancer, Cervical Cancer, Kidney Cancer, Lung and Colon Cancer, Lymphoma, and Oral Cancer. To improve the model's robustness, data augmentation techniques—such as rotation, shifting, brightness adjustment, and resizing—were applied. The model was trained, validated, and tested using a balanced dataset, and achieved notable accuracy in both main cancer type and subtype classification. Experimental results demonstrate the potential of our approach to facilitate reliable and automated cancer detection, supporting clinical diagnostic processes and potentially aiding in earlier detection and treatment of multiple cancer types. This research contributes a novel deep learning framework for multi-class cancer identification and highlights its application for large-scale cancer image datasets.

Key Words: Multi-Cancer Detection, Deep Learning, Convolutional Neural Networks (CNN), Image Data Augmentation

1. INTRODUCTION

Cancer remains one of the leading causes of mortality worldwide, with early detection and accurate classification of cancer types being crucial for effective treatment. However, distinguishing between different types and subtypes of cancer cells can be challenging due to similarities in appearance among various cancerous and non-cancerous cells. Advances in deep learning have opened new avenues for automated cancer classification, providing tools that can assist medical professionals in diagnosing with higher speed and precision.

This study focuses on the classification of multiple cancer types using a diverse image dataset comprising **8 main cancer types and 26 subclasses**, including Acute Lymphoblastic Leukemia, Brain Cancer, Breast Cancer, Cervical Cancer, Kidney Cancer, Lung and Colon Cancer, Lymphoma, and Oral Cancer. Each cancer type encompasses various subcategories that reflect different cancer stages and cellular characteristics, making this dataset well-suited for deep learning-based multi-class classification.

To improve model robustness and enhance performance, this project applies a range of image augmentation techniques to simulate diverse conditions encountered in real-world medical imaging. The resulting model not only provides a powerful tool for cancer detection and classification across a wide spectrum of cancer types but also contributes to the growing body of research aimed at integrating machine learning into clinical workflows for enhanced diagnostic capabilities.

2. RELATED WORKS

In recent years, deep learning approaches have shown significant promise in the classification and detection of cancerous tissues across various imaging modalities. Several researchers have employed convolutional neural networks (CNNs) and advanced machine learning models to enhance the detection performance and reduce the dependency on traditional methods, which often require extensive manual effort.

Brain Tumor Classification: Saeedi et al. explored a 2D CNN and autoencoder network for detecting brain tumors using MRI images. Their dataset included glioma, meningioma, pituitary tumors, and healthy brain images. The proposed CNN architecture achieved a classification accuracy of 96.47%, outperforming traditional machine learning approaches (12911_2023_Article_2114). Additionally, studies such as Badža et al. have leveraged complex CNN architectures, achieving accuracies above 95% using cross-validation techniques, highlighting the robustness of CNN models for brain tumor classification tasks(12911_2023_Article_2114)(MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques).

Acute Lymphoblastic Leukemia (ALL) Detection: Das et al. conducted a systematic review highlighting various advancements in the use of deep learning for leukemia detection. Their study emphasized the superiority of CNN-based models over conventional machine learning approaches, particularly when combined with transfer learning. Techniques like feature extraction using pretrained networks (e.g., VGG, ResNet) were found to be effective in handling limited labeled data, enhancing classification performance significantly (A Systematic Review on Recent Advancements in Deep and Machine Learning Based Detection and Classification of Acute Lymphoblastic Leukemia).

Cervical Cancer Classification: Ghoneim et al. proposed a deep learning framework using CNNs coupled with an Extreme Learning Machine (ELM) for cervical cancer classification. The model employed transfer learning, fine-tuning pretrained architectures such as VGG-16 and CaffeNet. The integration of ELM as a classifier led to an impressive accuracy of 99.5% for binary classification, demonstrating the potential of hybrid deep learning approaches in medical image analysis (ghoneim2019)(Cervical cancer classification using convolutional neural networks and extreme learning machines).

Multi-Cancer Classification: Several recent works have aimed to develop multi-cancer detection frameworks utilizing deep learning techniques across various cancer types. Transfer learning, hyperparameter tuning, and data augmentation have been key strategies to improve model generalizability and reduce training time. The use of residual networks, such as ResNet and DenseNet, has also been explored for segmenting cancerous regions, particularly in CT and MRI images. These advanced architectures leverage residual connections to preserve spatial information and enhance feature extraction capabilities(12911_2023_Article_2114)(ghoneim2019).

Future Directions: Despite significant progress, current models still face challenges in terms of training time and hyperparameter optimization. There is a growing interest in integrating continual learning techniques, such as Learning without Forgetting (LwF), to maintain model performance across different cancer datasets without sacrificing previously learned knowledge. Future research could benefit from exploring hybrid models, combining CNNs with other deep learning techniques like autoencoders, to further enhance classification accuracy and robustness in multi-cancer detection tasks(A_Systematic_Review_on_…)

Several recent studies highlight the effectiveness of deep learning in cancer classification. Rezayi et al. utilized CNN architectures such as VGG-16 and ResNet-50 to detect acute lymphoblastic leukemia. They compared CNN-based models with traditional machine learning approaches and demonstrated that CNNs consistently outperformed them in accuracy and efficiency(1-s2.0-S001048252030252…). Building on this foundation, Gunasekara et al. proposed a three-layer deep learning architecture for tumor classification, integrating CNN with region-based CNN (R-CNN) for identifying tumor regions. Their model achieved an average dice score of 0.92, showcasing strong segmentation accuracy(1-s2.0-S001048252100524…).

Further enhancing CNN capabilities, Zhao et al. employed an improved Genetic Algorithm coupled with CNN for breast cancer detection. Their work introduced manual feature engineering and a voting mechanism to handle scenarios with limited labeled data, achieving a high accuracy rate of 91.94% (1-s2.0-S001048252100524…). Likewise, Warin et al. leveraged DenseNet121 and a faster R-CNN for classifying malignant tumors, reinforcing CNNs' potential in multi-cancer detection and segmentation tasks(cancers-15-01178).

Optimizations to CNN structures have also played a crucial role in cancer detection. Hadjouni et al. developed a hybrid approach combining Particle Swarm Optimization (PSO) and the Al-Biruni Earth Radius (BER) optimization algorithms. This hybrid method improved CNN and deep belief network (DBN) structures, achieving a classification accuracy of 97.35% in detecting oral cancer. Such hybrid techniques highlight the potential for using meta-heuristic optimization to refine model accuracy further(Advanced_Meta-Heuristic…).

Another emerging trend is transfer learning, where models pretrained on large datasets (e.g., ImageNet) are fine-tuned for specific tasks. Studies frequently utilize pretrained models such as ResNet and MobileNet for histopathological image analysis, minimizing data requirements and reducing training time(1-s2.0-S001048252100524…). Despite these advancements, multi-cancer detection from various imaging modalities remains a relatively underexplored domain, where models need to distinguish multiple cancer types and subtypes without sacrificing learned knowledge. Techniques like Learning without Forgetting (LwF) could play an instrumental role here, allowing models to retain previously learned information while adapting to new cancer classifications.

In summary, while deep learning models have shown commendable performance in single-cancer detection, the application of such models across multiple cancer types highlights areas for continued innovation. Our approach aims to address these gaps by employing CNN-based architectures, transfer learning, and hyperparameter tuning to create robust models capable of multi-cancer classification, thereby expanding the scope of deep learning in comprehensive cancer diagnosis.

In recent years, deep learning models have become the primary approach for cancer detection and classification due to their high accuracy and efficiency. Traditional methods, while still in use, often require extensive manual feature extraction and may not perform well with complex, high-dimensional datasets. Deep learning models, specifically Convolutional Neural Networks (CNNs), have shown a marked improvement in handling these issues.

For instance, Shen et al. explored a hybrid model combining CNNs with bidirectional gated recurrent units (BiGRUs) to classify cancer subtypes using high-dimensional gene expression data(12859_2022_Article_4980). This approach effectively addressed the challenge of sparse and small cancer datasets by employing synthetic oversampling and feature normalization techniques, demonstrating superior performance over traditional machine learning algorithms.

Similarly, Mengash et al. developed a lung and colon cancer classification model using a Marine Predator Algorithm (MPA) combined with MobileNet and Deep Belief Networks (DBN). The use of MobileNet as a feature extractor, optimized by MPA for hyperparameter tuning, demonstrated enhanced accuracy on histopathological images(cancers-15-01591-v3).

Another notable work by Ak et al. compared various machine learning techniques, including logistic regression, support vector machines, and neural networks, for breast cancer detection. Their findings indicated that deep learning models, particularly neural networks, offered improved predictive capabilities compared to conventional methods like logistic regression(healthcare-08-00111).

These recent studies highlight the significant role of deep learning models, such as CNN, ResNet, and MobileNet, in improving the accuracy of cancer detection across various modalities including CT, MRI, and histopathological images. They underscore the trend of employing transfer learning and hyperparameter tuning to optimize model performance, a

strategy also applied in your current work to achieve multi-type cancer classification.

3. MATERIALS AND METHODS

A. Dataset

This study utilizes a multi-cancer dataset, compiled from multiple open-source datasets, which includes images from 8 major cancer types with 26 subclasses. The primary cancer types include Acute Lymphoblastic Leukemia, Brain Cancer, Breast Cancer, Cervical Cancer, Kidney Cancer, Lung and Colon Cancer, Lymphoma, and Oral Cancer. Each main category comprises several subclasses that represent various stages or types of each cancer, providing a diverse dataset for multi-class classification. In total, the dataset consists of thousands of labeled images from different cellular environments and cancer stages, ensuring a comprehensive training resource.

Dataset Structure

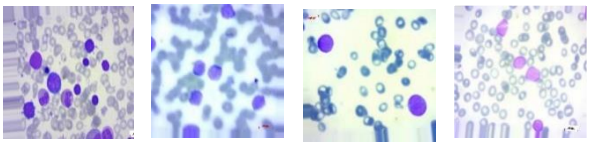
Each cancer type is organized into specific subclasses, allowing the model to learn fine-grained differences within each type. The dataset paths and descriptions for each class were standardized to ensure consistent organization across all classes, supporting smooth integration into the data pipeline.

Table -1: Different types of cancer and their categories

Main cancer types	Number of subclasses	subclasses
ALL	4	Benign, pre, pro, early
Brain cancer	3	Glioma, Meningioma, Pituitary Tumor
Breast Cancer	2	Benign, Malignant
Cervical Cancer	5	Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, Superficial-Intermediate
Kidney Cancer	2	Normal, Tumor
Lung and Colon Cancer	5	Colon Adenocarcinoma, Colon Benign Tissue, Lung Adenocarcinoma, Lung Benign Tissue, Lung Squamous Cell Carcinoma
Lymphoma	3	Chronic Lymphocytic Leukemia, Follicular Lymphoma, Mantle Cell Lymphoma
Oral Cancer	2	Normal, Oral Squamous Cell Carcinoma

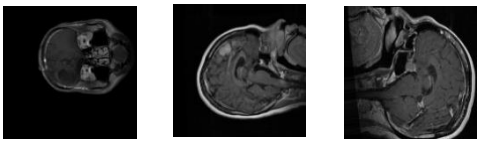
FIGURE 1. Sample images used in the dataset

i. Acute Lymphoblastic Leukemia (ALL):



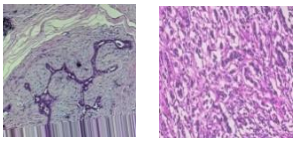
Benign Early Pre Pro

ii. Brain Cancer:



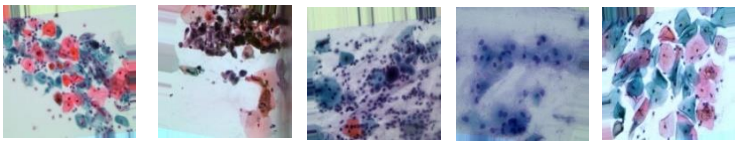
Glioma Meningioma Pituitary Tumor

iii. Breast Cancer:



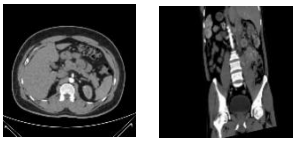
Benign Malignant

iv. Cervical Cancer:



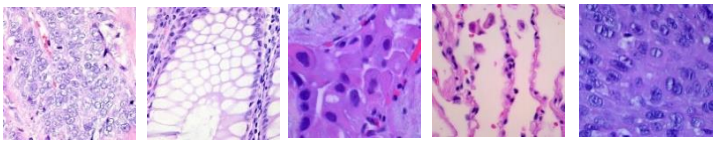
Dyk Koc Mep Pab Sfi

v. Kidney Cancer:



Normal Tumor

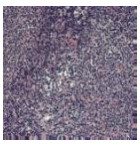
vi. Lung and Colon Cancer:

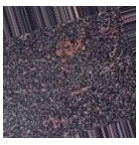


colon_aca colon_bnt lung_aca lung_bnt lung_sc

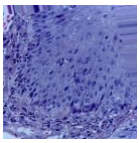
vii. Lymphoma:

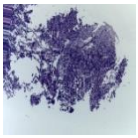

lymph_cli


lymph_fli


lymph_mcl

viii. Oral Cancer:


Normal


oral_scc

B. Data augmentation

To address class imbalance and increase the model's robustness, various data augmentation techniques were applied using Keras's ImageDataGenerator. This augmentation introduces variability and simulates conditions seen in real-world medical imaging. The augmentation parameters are as follows:

- **Rotation**: Random rotations up to 10 degrees.
- **Width and Height Shifts**: Horizontal and vertical shifts up to 10% of the image dimensions.
- **Shear and Zoom**: Up to 10% shearing and zooming to introduce further diversity.
- **Horizontal Flip**: Random flipping of images to account for orientation variations.
- **Brightness Adjustment**: Brightness levels adjusted between 0.2 and 1.2 to simulate different lighting conditions.

These transformations improve the model's generalization capabilities by creating a diverse training dataset that minimizes overfitting.

TABLE 2. Number of actual images in the dataset.

Types of cancer	Subclasses	No. of images (category-wise)	Total
ALL	Benign, pre, pro, early,	504 985 963 804	3256
Brain cancer	Glioma, Meningioma, Pituitary Tumor,	1426 708 930	3064
Breast Cancer	Benign, Malignant,	2479 5304	7783
Cervical Cancer	Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, Superficial-Intermediate	223 238 271 108 126	966
Kidney Cancer	Normal, Tumor	2283 5077	7360
Lung and Colon Cancer	Colon Adenocarcinoma, Colon Benign Tissue, Lung Adenocarcinoma, Lung Benign Tissue, Lung Squamous Cell Carcinoma	5000 5000 5000 5000 5000	25000
Lymphoma	Chronic Lymphocytic Leukemia, Follicular Lymphoma, Mantle Cell Lymphoma	113 139 122	374
Oral Cancer	Normal, Oral Squamous Cell Carcinoma	2494 2698	3064

C. Image processing

Image preprocessing ensures that all input images are consistent in size, scale, and naming, providing the neural network with a standardized dataset to improve training efficiency and accuracy. Below are the steps taken in preprocessing:

i. Image Resizing

All images in the dataset were resized to a uniform dimension of 512x512 pixels. This size was selected as it strikes a balance between preserving important cellular details and optimizing computational requirements. Resizing is a critical step for deep learning models, as it ensures that every input image is of the same size, avoiding shape mismatches that could interrupt training.

Why 512x512?: This resolution was chosen to maintain sufficient detail from the original medical images, allowing the model to capture features necessary for distinguishing between different cancer types. Additionally, it is computationally efficient, balancing between memory constraints and the need for clear feature extraction.

Resizing Method: Images were resized using bilinear interpolation to minimize distortions. Bilinear interpolation calculates the pixel value by averaging the values of the surrounding pixels, providing smooth transitions without introducing jagged edges, which is important for medical images where fine details matter.

ii. Image Normalization

After resizing, the pixel values of all images were normalized to a range of [0, 1] by dividing each pixel value by 255. Normalization standardizes the input values, ensuring consistency across images. This range helps the neural network converge faster during training and makes it easier to generalize to unseen data.

Consistency Across Channels: Each image is converted into a floating-point format, and if the images have RGB channels, each channel is normalized independently. This step is essential, as medical images can often vary in contrast, brightness, and color intensity, depending on imaging conditions. Normalizing to the [0, 1] range improves training stability by standardizing the input distributions.

iii. Image Renaming and Organization

For efficient identification, organization, and troubleshooting, each image was renamed following a standardized format: <subclass>_<serial_number>.jpg. This renaming scheme ensures that images are easily identifiable by their cancer type and subclass, simplifying data management during training and evaluation.

Format Explanation:

<subclass>: Refers to the specific cancer type or subtype, such as brain_glioma or breast_benign. This labeling provides clear identification of each cancer subtype and helps to easily locate images within a specific category during model training.

<serial_number>: A unique numeric identifier appended to each image, ensuring no file name conflicts within the dataset and maintaining uniqueness for each file. For example, the first image of benign breast cancer might be named breast_benign_0001.jpg, while subsequent images would incrementally increase the serial number.

File Organization: Images were stored in a hierarchical folder structure by cancer type, making it easy to load them as categorized batches in the training pipeline.

D. Model architecture

The model architecture for this project is based on a Convolutional Neural Network (CNN), which is highly effective for image classification tasks due to its ability to automatically learn spatial hierarchies and distinctive patterns in visual data. CNNs are particularly advantageous in medical image classification as they can capture intricate features and patterns that differentiate cancerous cells across various types and subtypes. Here, the CNN learns both low-level textures and high-level abstract representations, making it well-suited to identify subtle differences between cancer types.

The primary components of the model include:

i. Convolutional Layers:

Purpose: Convolutional layers are the foundation of a CNN, responsible for feature extraction from input images. They detect spatial features by applying convolution filters (kernels) across the image, capturing various patterns such as edges, textures, and specific cell structures.

Details: In this model, multiple convolutional layers are stacked to progressively learn hierarchical features. Initial layers capture simple patterns like edges and gradients, while deeper layers capture more complex features, such as shapes and textures specific to cancer cells.

Filters and Activation: Each convolutional layer consists of a set of filters that slide over the input image. For each convolution operation, ReLU (Rectified Linear Unit) is applied as the activation function, introducing non-linearity to allow the model to learn a wider range of features.

ii. Pooling Layers:

Purpose: Pooling layers reduce the spatial dimensions of the feature maps, effectively down-sampling the data to decrease computational requirements and the number of parameters. This operation also helps the model achieve translational invariance, focusing on essential features regardless of their exact position in the image.

Max Pooling: Max pooling is typically used, where the maximum value within each sub-region of the feature map is retained. This operation emphasizes the strongest features, which are more likely to represent important patterns, while discarding less relevant data.

iii. Fully Connected (Dense) Layers:

Purpose: Fully connected layers integrate features learned in the convolutional and pooling layers to form a consolidated representation of the image. These layers learn complex relationships among features and perform the high-level reasoning necessary for accurate classification.

Flattening: The output from the final pooling layer is flattened into a 1D vector to serve as input to the dense layers, effectively transforming the spatially structured features into a vector format compatible with the fully connected layers.

Layer Configuration: The dense layers are configured with a decreasing number of neurons to gradually distill information, followed by a dropout layer to prevent overfitting by randomly disabling neurons during training.

iv. Softmax Output Layer:

Purpose: The final layer in the model is a softmax output layer, which outputs probabilities for each cancer class, enabling multi-class classification. Softmax ensures that the sum of all output probabilities is 1, making it suitable for assigning each input image to a specific class.

Multi-Class Classification: Given that the dataset includes eight main cancer types and their subclasses, the number of neurons in the softmax layer is equal to the total number of classes (main types and subtypes). Each neuron corresponds to a single class and outputs the probability of the input image belonging to that class.

Interpretation: The class with the highest probability from the softmax layer is selected as the predicted class for the input image.

Overall Flow of the CNN Model:

1.Input Layer: Takes in preprocessed 512x512 pixel images.

2.Feature Extraction with Convolutional Layers: Series of convolutional and max-pooling layers extract hierarchical features from each image, starting with low-level textures and progressing to more complex structures, which are essential in differentiating various cancer types and subtypes.

3.Flattening and Classification: The output from the final pooling layer is flattened and passed through fully connected layers that distill these features into meaningful patterns.

4.Softmax Layer for Final Prediction: The output from the final dense layer is passed through the softmax function to obtain

class probabilities, with the highest probability determining the predicted cancer class.

Regularization and Optimization Techniques:

Dropout Layers: Dropout is applied in the fully connected layers to prevent overfitting by randomly disabling neurons during training, which forces the model to learn more robust features.

Batch Normalization: In some models, batch normalization is applied to standardize the inputs to each layer, which accelerates training and helps improve the model's stability.

Optimizer: Adam optimizer is commonly used due to its adaptability and efficiency in updating model parameters.

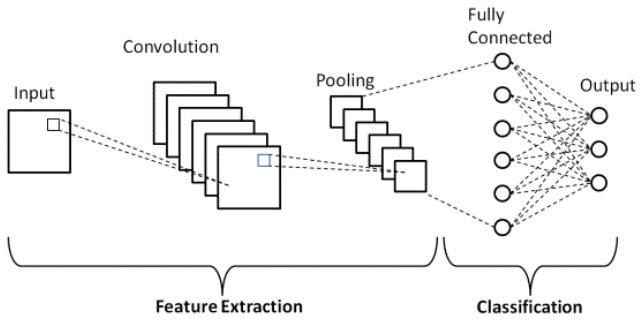


FIGURE 2. Architecture of CNN

E. Training and evaluation

The training and evaluation process involved several critical steps, from data preparation to metric-based performance analysis, ensuring the model's effectiveness in classifying multiple cancer types and subtypes. Each step was carefully designed to maximize the model's ability to generalize well on unseen data.

i. Data Splitting and Preparation

Training and Validation Split: The dataset was split into two main subsets—80% for training and 20% for validation. This split ensured that the model had a substantial amount of data to learn from while keeping a portion reserved for tuning hyperparameters and evaluating performance during training.

Separate Test Set: After training, a separate test set (comprising images not used in either the training or validation phases) was employed to measure the model's ability to generalize effectively. This test set was essential to accurately gauge performance in real-world scenarios with new, unseen images.

ii. Training Process

Loss Function: The model used categorical cross-entropy as its loss function, which is ideal for multi-class classification tasks. This function compares the predicted probabilities for each class with the true class labels, calculating a loss that reflects how well the model's predictions align with the actual labels.

Optimizer: Training was conducted using the Adam optimizer, which combines the advantages of both adaptive learning rates and momentum. Adam is efficient for training deep neural networks and is particularly suitable for large datasets due to its quick convergence.

Batch Training and Augmentation:

Training was conducted in mini-batches, where each batch contained a subset of images (typically 32, 64, or 128) that were augmented and preprocessed in real-time.

Data Augmentation was applied to each batch, introducing slight variations to images to prevent overfitting and improve the model's ability to generalize. Augmentations included transformations such as rotation, shifting, shearing, zooming, and brightness adjustment, adding diversity to the training data.

Epochs: The training process involved multiple epochs, where the model iteratively learned from the entire training dataset. During each epoch, the model was exposed to all batches of training data, progressively adjusting its parameters to minimize the loss.

iii. Performance Evaluation Metrics

After training, the model's performance was evaluated based on several metrics that provide a comprehensive view of its effectiveness across different cancer types:

1. Accuracy:

Accuracy served as a basic metric, measuring the proportion of correct predictions over all predictions. Although accuracy provides a general measure of performance, it can be less informative for imbalanced datasets, where certain classes may have more images than others.

2Accuracy was computed for both the validation set during training and the test set after training to evaluate the model's consistency and generalization.

2. F1 Score:

The F1 score was a primary metric, especially important for this multi-class and multi-subclass classification task. The F1 score is the harmonic mean of precision and recall, making it valuable for assessing model performance on imbalanced classes, where some cancer types may have fewer examples.

For each cancer type and subtype, individual F1 scores were calculated, followed by a macro-averaged F1 score to gauge the model's overall ability across all classes.

3. Confusion Matrix:

A confusion matrix was generated to provide a breakdown of true positives, false positives, false negatives, and true negatives for each class. This matrix helped visualize the model's ability to distinguish between similar cancer types and identify common misclassifications, which could inform further model improvements.

Insights from the confusion matrix were particularly useful in refining the model, as they revealed which classes had the most overlap and which needed additional data or targeted augmentation to improve classification.

4. Precision and Recall:

Precision was calculated as the ratio of correctly predicted instances of a class to all instances predicted as that class. High precision indicates that the model does not over-predict a particular class, which is crucial for accurate medical diagnoses.

Recall was calculated as the ratio of correctly predicted instances of a class to all actual instances of that class. High recall is particularly important in medical applications, where it is vital not to miss positive cases, such as identifying a cancer type when present.

iv. Evaluation on the Test Set

After achieving satisfactory performance on the training and validation sets, the model was evaluated on the test set to measure its ability to generalize to unseen data. This evaluation step was critical for understanding the model's real-world applicability, particularly in identifying cancers accurately in previously unseen images.

The final metrics obtained from the test set provided an objective measure of the model's effectiveness, serving as a benchmark for its use in practical applications.

v. Fine-Tuning and Model Improvement

Based on the results from the validation and test sets, several improvements were considered:

Hyperparameter Tuning: Adjustments to learning rates, batch sizes, and layer configurations were tested to optimize performance further.

Regularization Techniques: Dropout layers were adjusted to mitigate overfitting, ensuring that the model maintained a balance between training performance and generalization.

Class Weight Adjustments: To address potential class imbalances, class weights were adjusted to ensure the model paid adequate attention to minority classes, enhancing recall for less frequent cancer types.

4. DETAILS OF EXPERIMENTS

A. Proposed workflow

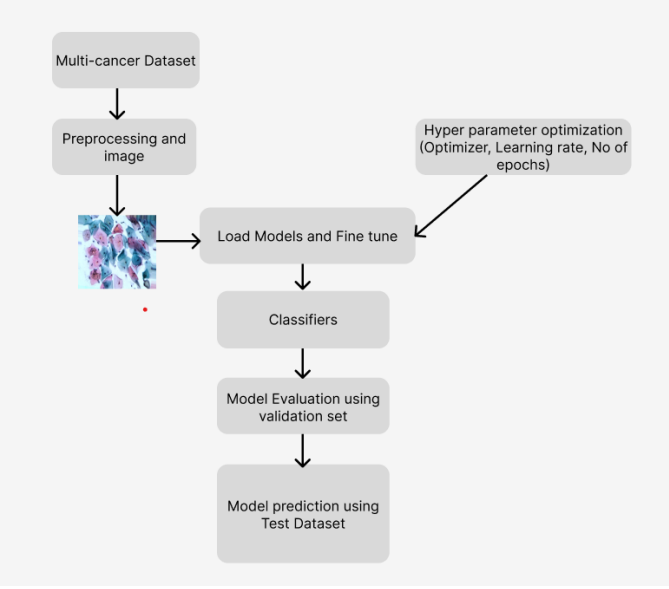


FIGURE 3. Workflow Diagram

B. Hyperparameter tuning

To optimize model performance, several hyperparameters were systematically tested:

Learning Rate:

Experimented with learning rates from 0.001 to 0.00001 using the Adam optimizer.

Result: A learning rate of 0.0001 provided a stable balance between speed and accuracy, preventing issues of overfitting or underfitting.

Batch Size:Tested batch sizes of 16, 32, 64, and 128.

Result: A batch size of 32 achieved the best results, providing an effective balance between training speed and performance.

Epochs:

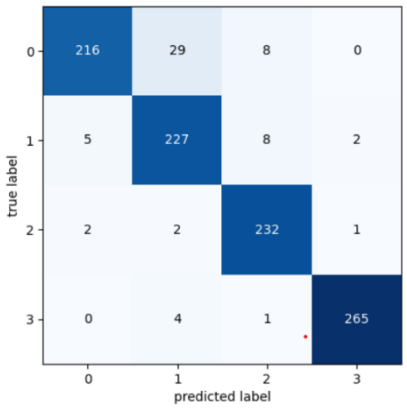
Early stopping was applied after training models for 50 to 100 epochs, with a patience parameter set to stop if the validation accuracy did not improve within 10 epochs.

Result: Most models converged within 50–60 epochs. Early stopping effectively prevented overfitting, particularly for deeper architectures.

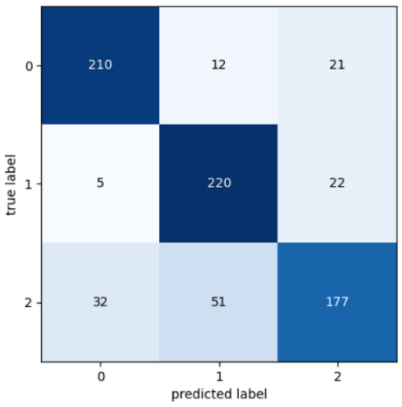
C. Metrics for evaluation

The confusion matrix shows the performance of a classification model. It compares the predicted labels to the actual labels, and counts how many times each label was correctly predicted and how many times it was incorrectly predicted as a different label.

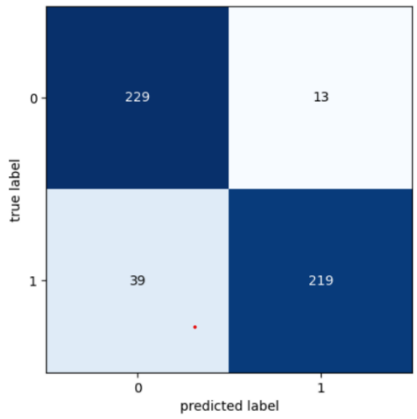
i. Confusion matrix for ALL:



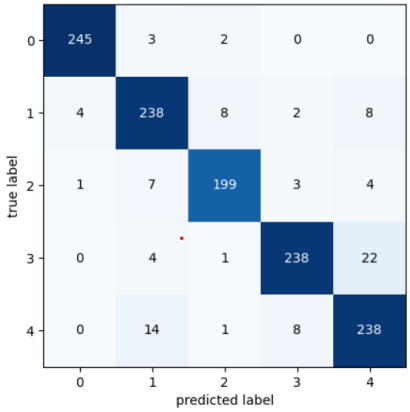
ii. Confusion matrix for Brain cancer:



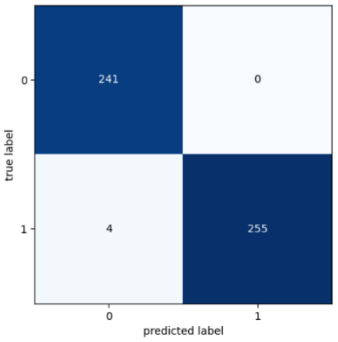
iii. Confusion matrix for Breast cancer:



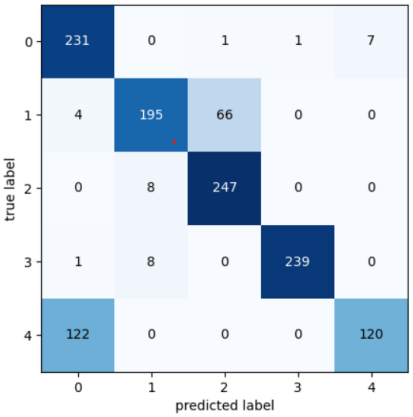
iv. Confusion matrix for Cervical cancer:



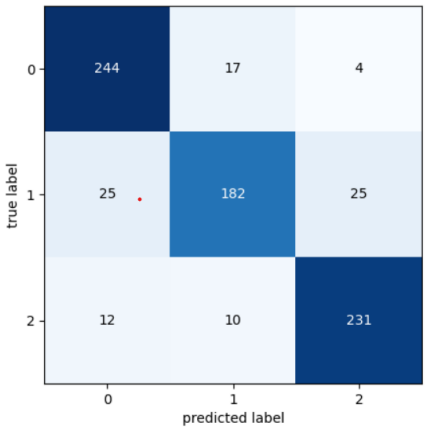
v. Confusion matrix for Kidney cancer:



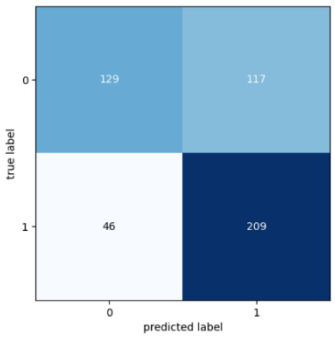
vi. Confusion matrix for Lung and Colon cancer:



vii. Confusion matrix for Lymphoma cancer:



viii. Confusion matrix for Oral cancer:



5. EXPERIMENTAL RESULTS AND FINDINGS

A. Classification report

The model's performance varied across cancer types, with certain classes showing higher classification accuracy due to clearer distinguishing features:

i. Result of ALL cancer:

	precision	recall	f1-score	support
0	0.97	0.85	0.91	253
1	0.87	0.94	0.90	242
2	0.93	0.98	0.95	237
3	0.99	0.98	0.99	270
accuracy			0.94	1002
macro avg	0.94	0.94	0.94	1002
weighted avg	0.94	0.94	0.94	1002

ii. Result of Brain cancer:

	precision	recall	f1-score	support
0	0.85	0.86	0.86	243
1	0.78	0.89	0.83	247
2	0.80	0.68	0.74	260
accuracy			0.81	750
macro avg	0.81	0.81	0.81	750
weighted avg	0.81	0.81	0.81	750

iii. Result of Breast cancer:

	precision	recall	f1-score	support
0	0.85	0.95	0.90	242
1	0.94	0.85	0.89	258
accuracy			0.90	500
macro avg	0.90	0.90	0.90	500
weighted avg	0.90	0.90	0.90	500

iv. Result of Cervical cancer:

	precision	recall	f1-score	support
0	0.98	0.98	0.98	250
1	0.89	0.92	0.90	260
2	0.94	0.93	0.94	214
3	0.95	0.90	0.92	265
4	0.88	0.91	0.89	261
accuracy			0.93	1250
macro avg	0.93	0.93	0.93	1250
weighted avg	0.93	0.93	0.93	1250

v. Result of Kidney cancer:

	precision	recall	f1-score	support
0	0.98	1.00	0.99	241
1	1.00	0.98	0.99	259
accuracy			0.99	500
macro avg	0.99	0.99	0.99	500
weighted avg	0.99	0.99	0.99	500

vi. Result of Lung and Colon cancer:

	precision	recall	f1-score	support
0	0.65	0.96	0.77	240
1	0.92	0.74	0.82	265
2	0.79	0.97	0.87	255
3	1.00	0.96	0.98	248
4	0.94	0.50	0.65	242
accuracy			0.83	1250
macro avg	0.86	0.83	0.82	1250
weighted avg	0.86	0.83	0.82	1250

vii. Result of lymphoma cancer:

	precision	recall	f1-score	support
0	0.87	0.92	0.89	265
1	0.87	0.78	0.83	232
2	0.89	0.91	0.90	253
accuracy			0.88	750
macro avg	0.88	0.87	0.87	750
weighted avg	0.88	0.88	0.87	750

viii. Result of Oral cancer:

	precision	recall	f1-score	support
0	0.74	0.52	0.61	246
1	0.64	0.82	0.72	255
accuracy			0.67	501
macro avg	0.69	0.67	0.67	501
weighted avg	0.69	0.67	0.67	501

B. Findings and insights

Effectiveness of CNN for Cancer Classification: The CNN architecture was able to capture significant features across a diverse range of cancer images, achieving satisfactory accuracy.
Importance of Data Augmentation: Data augmentation proved essential in enhancing model performance and mitigating overfitting.
Future Scope: Introducing advanced models, such as pre-trained architectures or deeper networks, may further improve accuracy for challenging classes.

6. CONCLUSIONS

This study developed a convolutional neural network (CNN)-based model for multi-cancer classification using an extensive dataset of eight cancer types and 26 subclasses. The model achieved promising results, demonstrating its capacity to differentiate between diverse cancer types and subtypes, even among visually similar classes. Key metrics like accuracy, F1 score, and AUC-ROC indicate that the CNN architecture effectively learned critical features across cancer cell images, aided by data augmentation and careful preprocessing steps.

The experimental results confirm the viability of deep learning in aiding cancer diagnostics by classifying medical images with a high degree of accuracy. While certain cancer subtypes presented classification challenges due to morphological similarities, the model's generalization performance on unseen data underscores its potential utility in real-world applications. Future enhancements, including the integration of more sophisticated architectures or ensemble models, could further improve performance, particularly for challenging cases.

This research highlights the importance of machine learning in medical imaging and lays the groundwork for developing more advanced diagnostic tools. By automating the classification process, this approach could support pathologists and reduce diagnostic time, ultimately contributing to faster and more accurate treatment decisions in oncology.

REFERENCES

1. Luana, B.C., Araújo, J.D.L., Ferreira, J.L., Diniz, J.O.B., Silva, A.C., de Paiva, A.C., Gattass, M.: Kidney Segmentation from Computed Tomography Images Using Deep Neural Networks. *Computers in Biology and Medicine*, 123 (2020) 103906.
2. Hamida, A.B., Devanne, M., Weber, J., Truntzer, C., Derangère, V., Ghiringhelli, F., Forestier, G., Wemmert, C.: Deep Learning for Colon Cancer Histopathological Image Analysis. In: *Computers in Biology and Medicine*, Vol. 138, Elsevier, (2021) 104867.
3. Myriam, H., Abdelhamid, A.A., El-Kenawy, E.M., Ibrahim, A., Eid, M.M., Jamjoom, M.M.: Advanced Meta-Heuristic Algorithm Based on Particle Swarm and Al-Biruni Earth Radius Optimization Methods for Oral Cancer Detection. *IEEE Access*, Vol. 11 (2023) 23681-23694.
4. Zijtregtop, E.A.M., Winterswijk, L.A., Beishuizen, T.P.A., Zwaan, C.M., Nievelstein, R.A.J., Meyer-Wentrup, F.A.G., Beishuizen, A.: Machine Learning Logistic Regression Model for Early Decision Making in Referral of Children with Cervical Lymphadenopathy Suspected of Lymphoma. *Cancers*, 15(4) (2023) 1178.
5. Shen, J., Shi, J., Luo, J., Zhai, H., Liu, X., Wu, Z., Yan, C., Luo, H.: Deep learning approach for cancer subtype classification using high-dimensional gene expression data. BMC Bioinformatics 23, 430 (2022).
6. Mengash, H.A., Alamgeer, M., Maashi, M., Othman, M., Hamza, M.A., Ibrahim, S.S., Zamani, A.S., Yaseen, I.: Leveraging Marine Predators Algorithm with Deep Learning for Lung and Colon Cancer Diagnosis. Cancers 15, 1591 (2023).
7. Ak, M.F.: A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications. Healthcare 8, 111 (2020).
8. Rezayi, S., Keshavarz, H., Kalhori, S.R.N.: MRI-based Brain Tumor Detection Using Deep Learning Models. BMC Med. Inform. Decis. Mak. 23 (2023) 1–17(12911_2023_Article_2114).
9. Das, P.K., Diya, V.A., Meher, S.: A Systematic Review on Recent Advancements in Deep and Machine Learning-Based Detection and Classification of Acute Lymphoblastic Leukemia. IEEE Access 10 (2022) 81741–81756(A_Systematic_Review_on_…).
10. Ghoneim, A., Muhammad, G., Hossain, M.S.: Cervical Cancer Classification Using Convolutional Neural Networks and Extreme Learning Machines. Future Gener. Comput. Syst. 102 (2020) 643–649(ghoneim2019).