**Machine Learning Project Documentation**

**Model Refinement**

**1. Overview**

The data preparation and feature engineering phase are crucial steps in any machine learning project, as it significantly impacts the model's performance and ability to generalize. This phase involves several key steps to transform raw data into a suitable format for the machine learning algorithm. These steps typically include resizing, dataset balancing, conversion to grayscale, normalization, equal distribution across training, validation, and testing sets, and data augmentation.

**2. Model Evaluation**

The evaluation process is essential for assessing the model's strength and efficiency. Precision, Recall, and F1-score were employed to understand the model's effectiveness comprehensively. Precision is the ratio of the true positive observation to the total optimistic prediction. Recall represents the ratio of predicted positive observation to all observations in the actual class, and the F1 score combines Precision and Recall.

(1)

(2)

(3)

TP represents True Positives, FP denotes False Positives, and FN signifies False Negatives.

The initial evaluation of the model shows a 96.4% accuracy on the testing dataset, which is considered satisfactory. However, there is room for improvement by employing additional preprocessing techniques and tuning the hyperparameters.

**3. Refinement Techniques**

Before feeding the dataset into the proposed model, several preprocessing steps were applied to the CT images to ensure optimal performance and minimize bias. Initially, the CT KIDNEY dataset images were resized to a standard dimension of 128x128 pixels, which helped in maintaining uniformity and reducing computational complexity. To enhance the model's ability to classify images accurately and to avoid any inherent bias, the dataset was balanced by ensuring an equal distribution of classes.To further simplify the model and concentrate on the essential features, the images were converted to grayscale, which eliminated unnecessary color information and allowed the model to focus on intensity values. These grayscale images were then normalized to a range of 0–1, a critical step for stabilizing the training process by ensuring that all pixel values were on a consistent scale.Additionally, to ensure the model trained effectively without bias, the dataset was divided so that each class had approximately equal representation in the training, validation, and testing sets. This careful distribution is crucial for building a robust model that performs well across different data subsets. Table 1 illustrates the datasets before and after the balancing process, highlighting the improvement in class distribution.To further enhance the dataset and improve the model's generalization capabilities, data augmentation techniques were applied. This involved creating variations of the existing images to artificially expand the dataset, leading to a more robust training process. Consequently, the dataset comprised 14,216 samples for training, 3,044 samples for validation, and 3,048 samples for testing, ensuring a comprehensive and well-balanced dataset for the model development.Table 2, shows the class distrbution across training, validation, testing datasets.

TABLE 1. CT images dataset before and after balancing.

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Before Balancing** | **After Balancing** | **After Augmentation** |
| **Normal** | 5,077 | 1,377 | 5,077 |
| **Cyst** | 3,709 | 1,377 | 5,077 |
| **Tumor** | 2,283 | 1,377 | 5,077 |
| **Stone** | 1,377 | 1,377 | 5,077 |

TABLE 2. CT images training , validation ,and testing class distrbution.

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Training Dataset** | **Validation Dataset** | **Testing Dataset** |
| **Normal** | 3,554 | 761 | 762 |
| **Cyst** | 3,554 | 761 | 762 |
| **Tumor** | 3,554 | 761 | 762 |
| **Stone** | 3,554 | 761 | 762 |

**4. Hyperparameter Tuning**

Grid search was utilized to systematically explore the combinations of various hyperparameters to identify the optimal settings. The hyperparameters adjusted included:

1. **Learning Rate**: The learning rate controls how much the model's weights are updated with respect to the loss gradient. Finding the right learning rate is crucial for training efficiency and model convergence.
2. **Batch Size**: This defines the number of samples that are processed before the model's internal parameters are updated. Different batch sizes can affect the stability and speed of the training process.
3. **Optimizer**: The optimization algorithm is responsible for updating the model's weights to minimize the loss function. Various optimizers such as Adam, SGD, and RMSprop were explored to find the one that best suited the model's needs.
4. **Number of Hidden Layers**: The depth of the neural network, determined by the number of hidden layers, impacts the model's ability to learn complex patterns. Grid search helped in determining the optimal number of hidden layers.
5. **Number of Filters**: In convolutional neural networks (CNNs), filters are used to detect features in the input images. The grid search method helped in selecting the right number of filters for each convolutional layer, balancing between model complexity and performance.

Table 3 shows all the hyperparamters adjusted using grid search.

Table 3 : The adjusted hyperparameters

|  |  |
| --- | --- |
| **Hyperparameter** | **Value** |
| Batch size | 32 |
| Optimizer | Adam |
| Learning rate | 0.001 (with decay ) |
| Epochs | 20 |
| Number of hidden layers | 7 |
| Number of filters | 32, 64, 128 , 128, 256 |
| Kernal size | 3 x 3 |
| Activation functions | ReLU (hidden layers) – Softmax (output layer) |

**Test Submission**

**1. Overview**

After applying refinement techniques, a substantial improvement in accuracy was observed. This study introduces a novel Hybrid CNN-Transformer model designed for detecting kidney abnormalities in CT images. Remarkably, the model achieved an exceptional overall accuracy of 100%. It demonstrates outstanding performance across various classes, with precision, recall, and F1 scores all nearing 1.00. These results highlight the efficacy of integrating explainable AI techniques in medical diagnostics, significantly enhancing the accuracy and reliability of automated systems. This advancement not only underscores the model's robustness in detecting intricate patterns in medical images but also its potential to assist healthcare professionals in making more informed decisions swiftly and accurately.

**2. Data Preparation for Testing**

The testing dataset was split from the main dataset to ensure that the model would be evaluated using an unseen dataset, providing a robust assessment of its performance. This testing dataset underwent extensive preprocessing steps to enhance the quality and relevance of the data. These steps included resizing to maintain consistency in dimensions, scaling to normalize the data, data augmentation to increase the diversity of the dataset and reduce overfitting, and binarization masks to convert the data into a binary format suitable for certain types of analysis. This thorough preprocessing ensures that the model's evaluation is based on well-prepared and representative data.

**3. Model Application**

The testing dataset was preprocessed (as mentioned the previous section) before being applied to the model to evaluate its performance using a wide range of test metrics such as accuracy, F1-score, recall, sensitivity, RMSE, MAE, and confusion matrix.

**4. Test Metrics**

In this section, we discuss the results of our proposed method. The proposed model employs a hybrid approach, combining a CNN and a Transformer, integrated with explainable AI techniques, specifically Grad-CAM. Grad-CAM is applied to the final layer of the CNN component to enhance interpretability. The model performance was evaluated using the metrics outlined in the Model Evaluation section. The training and validation accuracy and loss graphs in Fig. 1 demonstrate rapid convergence, achieving over 90% accuracy within the first three epochs and stabilizing of 100% validation accuracy. Both training and validation losses decrease sharply and stabilize at low values, indicating effective learning and minimal overfitting. The alignment of training and validation metrics confirms the model's robustness and generalization capabilities, underscoring its potential utility in medical diagnostic applications.

A graph of a training and training performance

Description automatically generated with medium confidence Figure 1: Losses and accuracies curves of training and validation datasets

The analysis of the hybrid CNN-Transformer model reveals exceptional performance across all evaluated metrics, as summarized in Table 4. The model achieved near-perfect precision, recall, and F1 scores for the Tumor, Normal, and Cyst classes, each scoring 1.00 in all metrics. While performing exceptionally well, the Stone class achieved a precision of 1.00, a recall of 1.00, and an F1-score of 1.00. This indicates perfect classification process. The model achieved an overall accuracy of 100%, highlighting its robustness in generalizing to unseen data. This accuracy represents the proportion of correctly predicted instances out of the total number of cases evaluated (3048), reaffirming the reliability of the model's predictions across the entire dataset. Additionally, the model achieved impressive performance metrics, with a RMSE (Root Mean Squared Error) score of 0.0049 and a MAE (Mean Absolute Error) of 0.00019. These scores indicate minimal prediction errors, highlighting the model's precise estimation capabilities. Such low error rates underscore its robustness and accuracy in predicting outcomes, further validating its potential utility in clinical settings. This underscores the significant strides made in leveraging advanced AI techniques to enhance diagnostic accuracy and improve patient care in medical imaging applications.

Table 4: Performance Metrics of the Hybrid CNN-Transformer Model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | Precision | Recall | F1-score | Samples |
| Tumor | 1.00 | 1.00 | 1.00 | 762 |
| Normal | 1.00 | 1.00 | 1.00 | 762 |
| Stone | 1.00 | 1.00 | 1.00 | 762 |
| Cyst | 1.00 | 1.00 | 1.00 | 762 |
| Accuracy |  |  | 1.00 | 3048 |

The confusion matrix in Fig. 2 further illustrates the model's performance. It shows perfect classification for the Tumor, Normal, Stone and Cyst categories.

A diagram of a confusion matrix

Description automatically generated

Figure 3: Confusion matrix of the CNN-Transformer model over the testing dataset

The confusion matrix demonstrates the performance of a hybrid CNN-Transformer model in classifying medical imaging data into four categories: Tumor, Normal, Stone, and Cyst. The model achieved perfect of all classifications (Tumor, Normal, Stone, and Cyst categories), with all 762 instances in each category correctly identified. This high level of performance highlights the potential of the CNN-Transformer hybrid model in medical diagnostics. Additionally, Fig. 4 shows the applied XAI (Grad-CAM) in two test samples.

A close-up of a infrared image

Description automatically generatedA close-up of a mri scan

Description automatically generated

**Tumor Stone**

Fig. 4. XAI for two test samples

The results achieved with high accuracy without reliance on pre-trained models underscore the efficacy of integrating Transformers alongside Convolutional Neural Networks (CNNs) for medical image classification. Transformers capture intricate spatial relationships and global context within images through their attention mechanisms, enhancing the model's ability to discern subtle patterns critical for accurate diagnosis. By leveraging the Transformer's capacity for end-to-end learning and sequence modeling, the proposed architecture not only learns hierarchical representations directly from the data but also adapts effectively to the complexities of medical imaging datasets. This approach facilitates robust generalization and achieves perfect precision, recall, and F1-score metrics across diverse classes, demonstrating the transformative impact of Transformers in advancing image analysis tasks.

**5. Model Deployment**

Our application leverages Flask, HTML, and JavaScript to deploy a machine learning model for classifying CT images. Users can upload a CT image, which will be preprocessed in the backend before classification. Upon pressing the "Predict Image" button, the model classifies the image into one of four conditions: “Normal,” “Cyst,” “Tumor,” or “Stone.” The result is then displayed to the user. Additionally, a heatmap is generated and displayed using the Grad-CAM algorithm to provide visual insights into the model's decision-making process.

The images below show the application user interface , when the user upload a CT image ,and the classification result along with the heatmap.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated A screenshot of a computer screen

Description automatically generated

**6. Code Implementation**

**A screen shot of a computer program

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screen shot of a computer program

Description automatically generated**

**A screen shot of a computer program

Description automatically generated**

**A screenshot of a computer program

Description automatically generated**

**Conclusion**

Integrating CNNs and transformers with explainable AI is a significant advancement in medical imaging, particularly for detecting kidney abnormalities in CT scans. The TransConvNet model achieves 100% overall performance and combines CNNs' local feature extraction with transformers' global context understanding for a robust diagnostic tool. Using explainable AI mechanisms like Grad-CAM improves the model's ability to explain its predictions. This helps healthcare providers understand and trust AI-driven diagnostic systems, ultimately improving patient care and outcomes.

The study shows that TransConvNet has the potential to revolutionize medical diagnostics by accurately detecting kidney abnormalities. Future work will refine the model and conduct clinical evaluations to validate its effectiveness in real-world scenarios. This research paves the way for the broader adoption of AI technologies in healthcare, enhancing medical professionals' capabilities and improving patient outcomes.