**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 15 March 2024 |
| Team ID | 740050 |
| Project Title | YOLOChemDetect safeguarding with Automated Drug Name Detection |
| Maximum Marks | 6 Marks |

**Preprocessing Template**

The images will be preprocessed by resizing, normalizing, augmenting, denoising, adjusting contrast, detecting edges, converting color space, cropping, batch normalizing, and whitening data. These steps will enhance data quality, promote model generalization, and improve convergence during neural network training, ensuring robust and efficient performance across various computer vision tasks.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | Medical text for drug detection. |
| Resizing | Adjust document size. |
| Normalization | Standardize text. |
| Data Augmentation | Add synthetic data. |
| Denoising | Remove noise from documents to improve accuracy. |
| Edge Detection | Highlight key areas for drug name identification. |
| Color Space Conversion | Convert document images to standard format for better analysis. |
| Image Cropping | Crop irrelevant sections for documents for focused processing. |
| Batch Normalization | Normalize document data in batches to improve model performance. |
| **Data Preprocessing Code Screenshots** | |
| Loading Data | Import medical documents for drug name detection. |
| Resizing | Resize images to standard dimensions for processing. |
| Normalization | Standardize document data for consistent analysis. |
| Data Augmentation | Enhance the dataset with synthetic samples for robustness. |
| Denoising | Remove noise from documents to improve detection accuracy. |
| Edge Detection | Identify key areas in images for drug name recognition. |
| Color Space Conversion | Covert images to suitable color formats for analysis. |
| Image Cropping | Crop irrelevant parts of images to focus on key information. |
| Batch Normalization | Normalize data in batches to optimize model performance. |