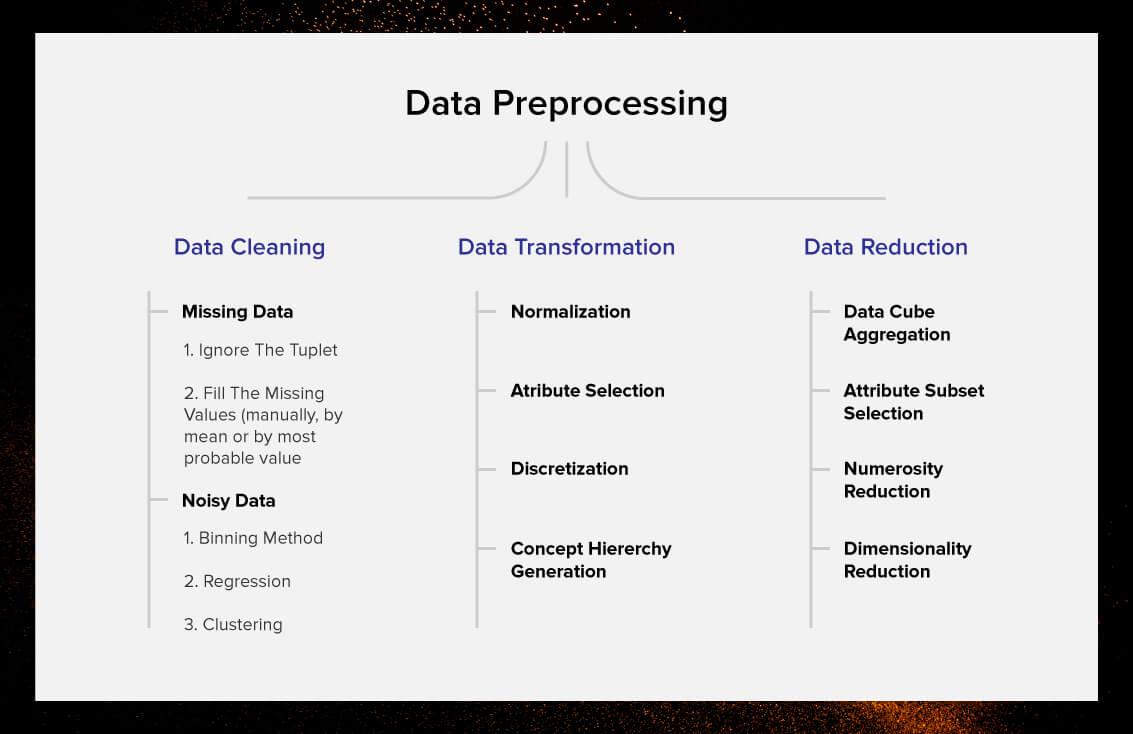
**STEP1:** 

Pre-processing of data to remove noisy data from the dataset is an essential step in data analysis and machine learning. Noise refers to the presence of irrelevant or erroneous data in the dataset that can affect the accuracy of the model's predictions. Here are some common techniques used for pre-processing of data to remove noisy data from the dataset:

**1. Outlier detection and removal**: Outliers are data points that lie far away from the normal range of values in the dataset. Outliers can be detected using various statistical methods and can be removed from the dataset to reduce their impact on the model's predictions.

**2.Data normalization:** Data normalization is the process of transforming the data into a standard format to remove variations and inconsistencies in the data. This can be achieved by scaling the data to a common range, such as between 0 and 1, or by using techniques such as Z-score normalization.

**3.Data smoothing:** Data smoothing involves removing noise from the dataset by averaging adjacent data points or using filters such as moving averages or median filters.

**4.Feature selection:** Feature selection involves identifying the most relevant features in the dataset and removing irrelevant or redundant features. This can help reduce noise and improve the accuracy of the model's predictions.

**5.Data augmentation:** Data augmentation involves generating new data points by applying transformations to the existing data. This can help increase the size of the dataset and reduce noise.

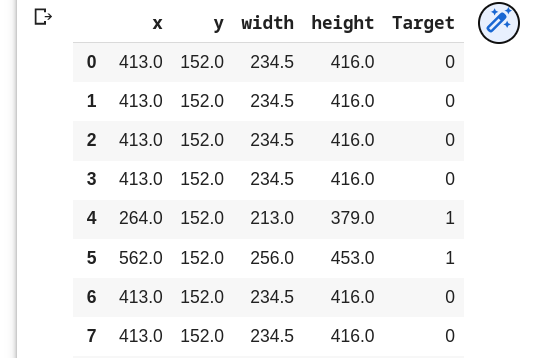
Overall, pre-processing of data to remove noisy data from the dataset is an important step in ensuring the accuracy and reliability of the model's predictions.

1. The dataset contains a lot of NaN values, which need to be removed so we can perform operations on the data.
2. The NaN values are filled by using KNN so that the data will contain all numeric values.
3. The data is segregated based on the target, i.e., target = 1 is for having lung opacity.

Table

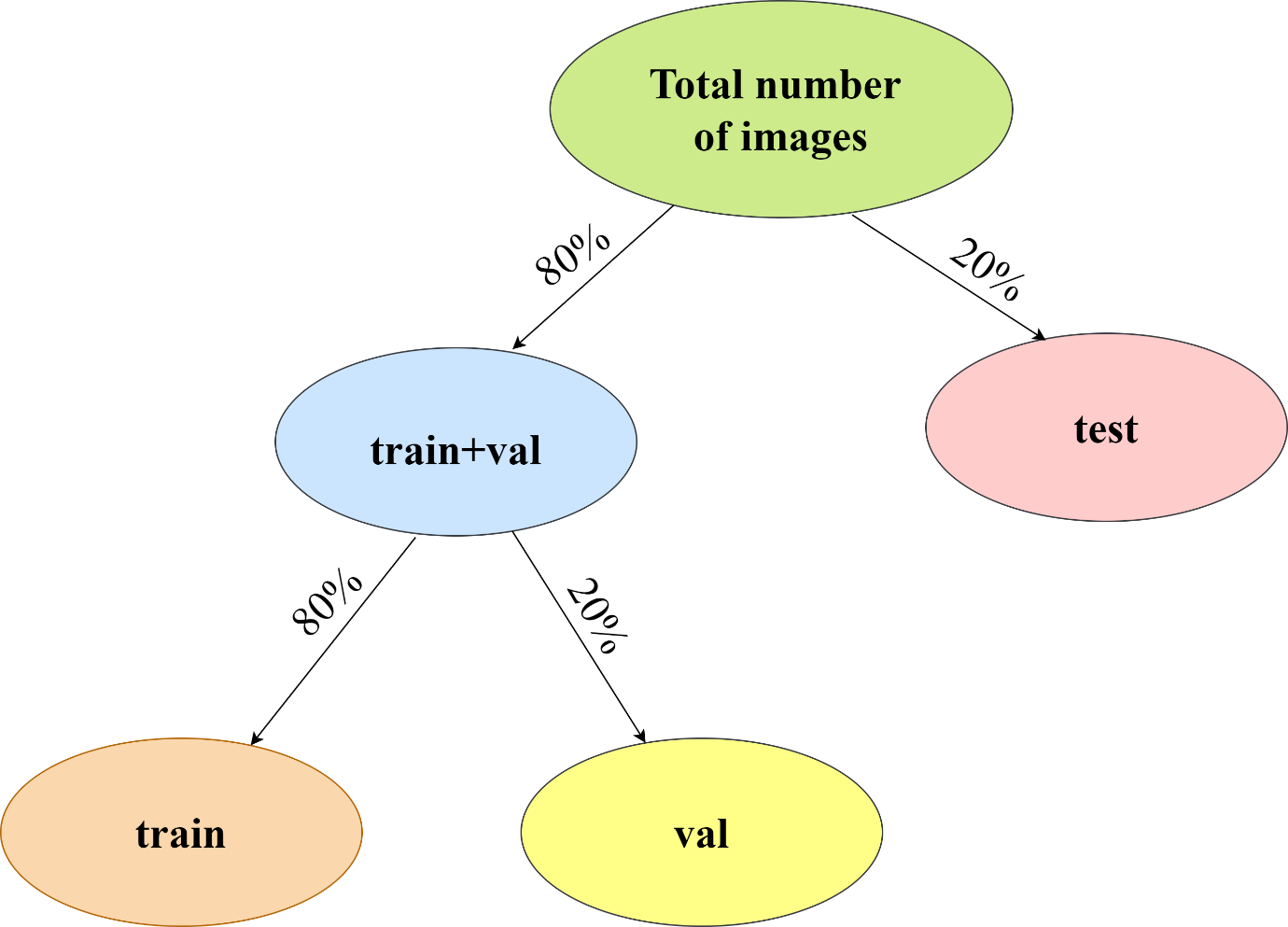
Description automatically generated

**Before processing the data**



**After processing the data using KNN**

**STEP2:**



Classifying images as labels and analyzing the data to train a model is a common task in computer vision and machine learning. Here are some of the key steps involved in this process:

**1.Data collection:** The first step in training a model to classify images is to collect a dataset of labeled images. The dataset should be representative of the problem you are trying to solve and should include a range of different images that are relevant to the task.

**2.Data preprocessing:** Once you have collected a dataset of labeled images, the next step is to preprocess the data. This might involve resizing the images to a standard size, normalizing the pixel values, and performing other operations to clean and standardize the data.

**3.Data augmentation:** Data augmentation involves generating additional training data by applying transformations such as rotation, scaling, and flipping to the existing images. This can help improve the robustness of the model and prevent overfitting.

**4.Model selection:** There are many different models that can be used for image classification, such as convolutional neural networks (CNNs) and deep neural networks (DNNs). The choice of model will depend on the size and complexity of the dataset, as well as the computational resources available.

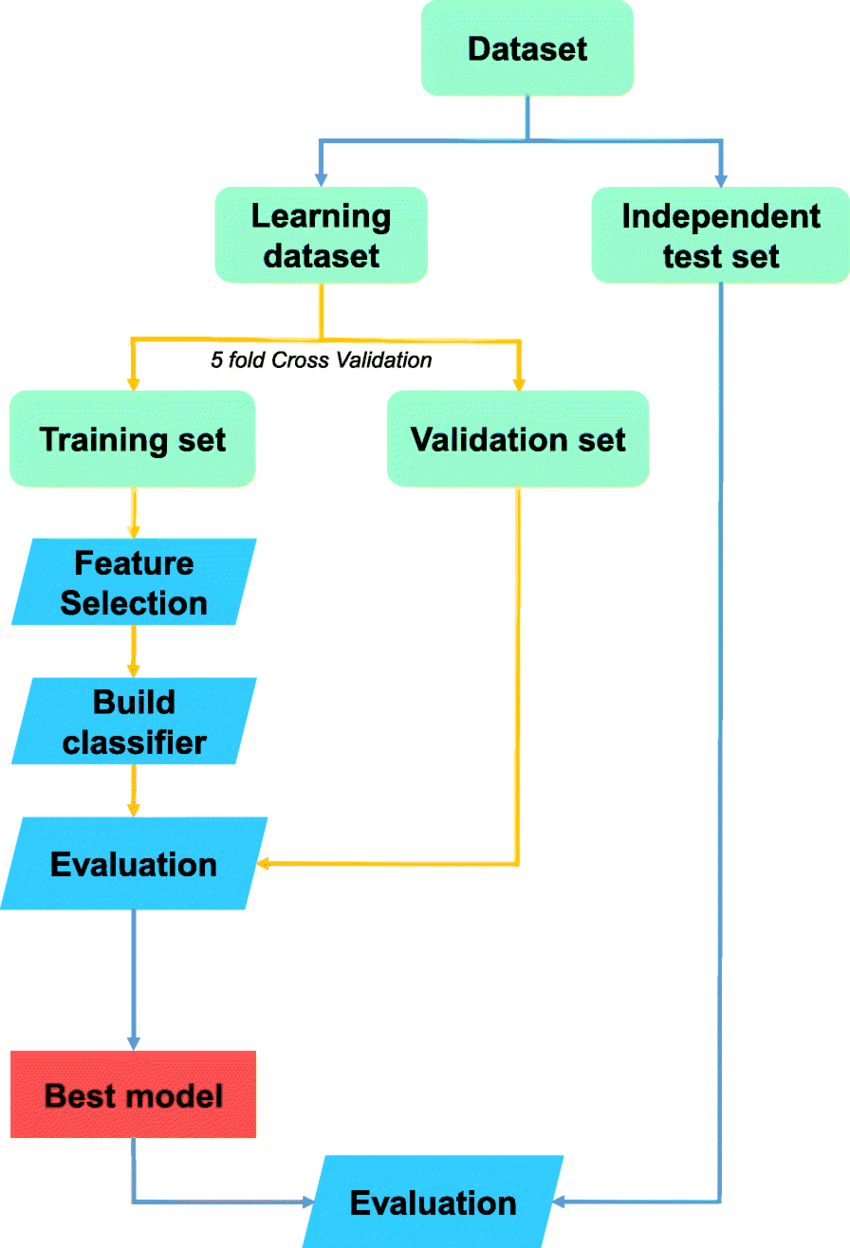
**5.Model training:** Once you have selected a model, the next step is to train it on the labeled dataset. This involves optimizing the model parameters to minimize the loss function, using techniques such as stochastic gradient descent (SGD).

**6.Model evaluation: After training the model, you will need to evaluate its performance on a validation set of images that were not used during training. This will give you an estimate of how well the model is likely to perform on new, unseen images.**

**7.Model deployment: Finally, once you have trained and evaluated the model, you can deploy it to classify new images in real-world applications.**

**Overall, classifying images as labels and analyzing the data to train a model is a complex and iterative process that requires careful consideration at each step. With the right techniques and tools, however, it is possible to build highly accurate and effective image classification models.**

**STEP3:**



Model selection and training is an important step in the process of developing a machine learning model. The goal of this step is to select the best model architecture, and train it on the available dataset. Here are some key considerations for model selection and training:

**1.Model selection**: There are many different types of machine learning models, each with its own strengths and weaknesses. The first step in model selection is to determine which type of model is best suited to your particular problem. For example, for image classification tasks, convolutional neural networks (CNNs) are often used, while for natural language processing tasks, recurrent neural networks (RNNs) or transformer-based models may be more appropriate.

**2.Model architecture:** Once you have selected the type of model, you need to determine the specific architecture to use. This involves selecting the number and type of layers, as well as the activation functions and other hyperparameters.

**3.Training data:** The performance of a machine learning model is heavily dependent on the quality and quantity of training data. It is important to use a representative dataset, and to ensure that the data is labeled correctly.

**4.Data preprocessing:** Preprocessing the data can improve the performance of a machine learning model by reducing noise and normalizing the input data. Common preprocessing steps include normalization, feature scaling, and dimensionality reduction.

**5.Hyperparameter tuning:** Many machine learning models have hyperparameters that can be tuned to improve performance. These might include the learning rate, the regularization term, or the number of hidden layers in the model. Hyperparameter tuning involves selecting the optimal values for these hyperparameters.

**6.Model training:** Once you have selected the model architecture and hyperparameters, it is time to train the model. This involves feeding the training data into the model and adjusting the weights and biases of the model to minimize the loss function.

**7.Model evaluation**: After training the model, it is important to evaluate its performance on a validation dataset. This can help identify overfitting and determine whether the model is generalizing well to new data.

Overall, model selection and training is a complex process that requires careful consideration of many different factors. With the right techniques and tools, however, it is possible to develop highly accurate and effective machine learning models.

**The model used in this project is ResNet-50.**

