# **Abstract**

Gastric cancer or stomach cancer is one of the most common causes of cancer mortality worldwide. As a result of its biological diversity, the disease develops because of a complicated combination of diverse environmental and hereditary factors, dietary patterns, and Helicobacter pylori infection. The prognosis for treating this disease is extremely poor due to the lack of early warning signs which often results in late stage diagnosis. Significant progress in the understanding of the disease's molecular and genetic development has enhanced the precision of diagnostic markers, targeted treatments, and even advanced bowel cancer immunotherapy. This paper seeks to discuss in detail about the epidemiology, associated risk factors, the pathophysiology, and the diagnosis and treatment methods of gastric cancer while also focusing on the recent advancements made in the study. Prevention and treatment methods need to be refined to effectively deal with the disease's biological structure, thus more research needs to be conducted to understand the biological framework.

# **Introduction**

GC or gastric cancer is the 3 rd most frequently diagnosed form of cancer in the world according to the WHO and gastric ulcers are responsible for over a 1/3 of deaths in the world counting it a global disadvantage. Regardless of the death toll, the popularity of stomach cancer among East Asian and Eastern European countries alongside parts of South America, has maintained its firm footing even after the global advances in healthcare, technology and hygiene. Unlike the advancements made in surgical methods and chemotherapy treatments, the 5-year survival rate for gastric cancer remains low, particularly for patients who are diagnosed in later stages. The early-stage disease is asymptomatic which makes it very difficult to detect and treat on time. On the other hand, recent progress in genomic and proteomic research, along with the field of personalized medicine, provide prospective possibilities for early stage diagnosis and tailored treatments. This paper aims to provide current knowledge with respect to gastric cancer by analyzing its epidemiology, disease mechanisms, clinical features, and treatment approaches. Moreover, it discusses recent advances in molecular studies which may enhance the strategies aimed at improving the clinical outcomes of patients, while also alleviating the burden of the disease at a population level. Among all forms of proven cancers, Gastric Cancer remains one the most painful one and utmost agony doesn’t just cease with the painful procedures it calls for, in fact it takes multi factorial approaches encompassing the use of medicines, dietary manipulations and lifestyle adjustments at the same time extending to genetics.With stomach cancer proving to be a mainstay in the medically minded over world, diffuse and intestinal type remain the two distinguished types with underlying physiological factors accompanying them.

# **Related Work**

M.M.Poly [[1](https://www.mdpi.com/2072-6694/13/21/5253)] have made a convolutional neural network (CNN)-based model that can detect gastric cancer from endoscopic pictures. That model helps doctors diagnose the disease quickly and accurately. They analyzed about 13,584 endoscopic pictures using different CNN architectures , like- VGG and ResNet. The data were collected from different hospitals and included a variety of cancer and non-cancer pictures. Although the accuracy of the model is remarkable ; which is 92.5%, in some cases it is confused in differentiating gastritis and primary cancer.The performance was poor in low-resolution or blurred images. The model can be further strengthened with clinical integration and multisource data;confused on gastritis vs cancer. The effectiveness can be improved by adding more diverse images and real-time trials. Also they did not use any custom cnn.

Lee [[2](https://www.mdpi.com/2072-6694/13/15/3811)] has made an automated deep learning system for subclassification of gastric carcinoma using histopathology images. This method is very useful for detecting heterogeneous subtypes of gastric cancer. The WSI was evaluated using EfficientNet and DenseNet architectures;high accuracy. The dataset consisted of approximately 1200 H&E-stained WSIs collected from South Korean businesses. The model was able to identify poorly-differentiated adenocarcinoma with high accuracy. The limitations are the heavy computational requirements and the number of target artifacts. The results can be further improved by adding stain normalization and attention processing. This method has a large potential for digital pathology and unmet diagnosis.

Hirasawa [[3](https://onlinelibrary.wiley.com/doi/full/10.1111/den.13688)] made a CNN model and compared it with experienced endoscopists in detecting gastric cancer using the ResNet-based model achieved an accuracy of about 92%; where the average physician had an accuracy of 80–85%. The model was trained on 2,296 endoscopic pictures of early gastric cancer. The pictures were collected from multiple hospitals in Japan. The model saves time and is helpful in detecting early cancer. Even so, it shows weakness in identifying complex cases or rare lesions. The performance can be improved by incorporating clinical features or patient history.

Ueyema and Kato [[4](https://onlinelibrary.wiley.com/doi/full/10.1111/jgh.15190)] have made a CNN-based model for gastric cancer detection using magnifying endoscopy and Narrow-Band Imaging (NBI). The model uses the GoogLeNet architecture fitted for deep image analysis, aiming to distinguish early-stage gastric cancer from ulcers and other lesions. The study used NBI images from 386 patients collected at Tokyo Medical University with manual annotations. The model achieved about 89% accuracy, helping clinicians make early and accurate decisions. Main benefit is its ability to detect cancer at an early stage when visual symptoms may not be clear. Even so due to dependence on image quality and single-center data, future improvements include adding multi-center data, diverse patients, enhanced image pre-processing, and varied endoscopic images.

Horiuchi [[5](https://link.springer.com/article/10.1007/s10620-019-05862-6)] has made a CNN model to distinguish between gastric cancer and gastritis using magnifying endoscopy and Narrow Band Imaging (NBI) images. The model uses VGG and ResNet-based structures because they are good at detecting subtle differences in images. They used almost 1,500 high-resolution endoscopic images collected from Tokyo Medical University Hospital, Japan, as a dataset. This model can distinguish gastritis and cancer with about 85% accuracy.The model helps in rapid diagnosis and shortens treatment time ;mainly in areas with high population density. Even so, due to the low image diversity, the model may perform poorly in patients with different cameras or different ethnicities. For improvement, more diverse images, data collected from different hospitals, and stain normalization can be added.

Wei [[6](https://onlinelibrary.wiley.com/doi/full/10.1111/cas.15592)] has made a multimodal deep learning model to predict survival in gastric cancer patients by combining histopathology images and gene expression data. The model uses a convolutional neural network to extract image features and an autoencoder-based encoder for analyzing gene expression, achieving high accuracy. Used Whole Slide Images (WSI) and gene expression data from about 500 gastric cancer patients collected from The Cancer Genome Atlas (TCGA). The model accurately classified patients into high-risk and low-risk groups, supporting personalized treatment planning by predicting clinical outcomes in advance. Although limited stain variability and lack of model interpretability made it difficult to gain physician trust. These issues can be addressed by adding some SHAP or LIME for interpretability and including diverse ethnic and geographic data for broader applicability.

Rahaman [[7](https://ieeexplore.ieee.org/abstract/document/10635260)] developed an advanced deep learning model for gastric histopathology image classification using supervised contrastive learning. The key novelty is using contrastive loss, which pulls embeddings of the same class close while pushing different classes apart, ensuring clearer separation in feature space. The model includes a ResNet-50 encoder, projection head, and classification head, with contrastive pretraining followed by supervised fine-tuning. The dataset carries over 4,000 gastric histopathology images, with various cancer grades. The results show that this unlike learning approach outperforms traditional cross-entropy loss, especially in handling imbalanced datasets. However, the model has high computational cost and long training time. Moreover, the contrastive method’s complexity may challenge clinician interpretation. Future improvements include model pruning or knowledge distillation for optimization.

Loddo [[8](https://www.mdpi.com/2313-433X/10/8/195?utm_source=chatgpt.com)] developed different deep learning models and feature fusion techniques for gastric cancer image classification. They manage a comparative analysis by combining traditional methods like CNN, ResNet-50, and handcrafted texture features. By these features within the models, they achieved higher classification accuracy compared to single-source feature models. The study used the Gastric-Histopath dataset, which contains 10,000 histological images with various tissue annotations. The paper shows that combining texture and deep features enhances model sensitivity in microstructure analysis using feature fusion models. Even so, handcrafted feature extraction is time-consuming and prone to manual errors. Deriving automatic feature learning and lightweight models could improve performance in real-world applications. Moreover incorporating various data augmentation techniques and color normalization would increase model robustness, supporting semi-automated clinical tools

Hung and Tian [[9](https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(21)00424-2/fulltext)] applied deep learning to digital pathology image analysis in a multicenter retrospective study to develop two separate models for gastric cancer. One model focused on disease diagnosis and the second other predicted patient survival and therapy response. They used approximately 2,500 Whole Slide Images (WSIs) and related clinical outcome data from multiple hospitals in China. AI models achieved over 90% accuracy sensitivity and significantly outperformed traditional pathological evaluations. This value in both diagnosis and treatment planning. However, performance was affected by stain inconsistency and scanning quality variations. To improve, the study suggests stain normalization and federated learning. More model interpretability can increase physician confidence and encourage clinical use, pushing personalized gastric cancer treatment closer to reality.

Tian.M [[10](https://link.springer.com/article/10.1186/s12967-023-04838-5)], a specialized AI model developed to analyze whole-slide images of gastric cancer for predicting digital pathology signatures and therapy responses. It uses attention-based Multiple Instance Learning (MIL) and weakly supervised learning, allowing the analysis of data without manual annotations. The model was trained using about 1,600 histopathology images and clinical outcome data collected from several cancer centers in China. Deep-Risk outperformed traditional deep learning models, gaining significantly higher AUC scores and better therapy predictions. The study shows AI’s role in gastric cancer, not only for diagnosis but also for treatment decision-making. Even so its reliance on single-country data limits global use. To modify this suggest using multi-regional datasets, transfer learning, and adding molecular and radiological features.

Zubair [[11](https://www.nature.com/articles/s41598-025-97256-0)] have made an interpretable framework for gastric cancer detection using multi-channel attention mechanism.An interpretable framework for gastric cancer classification using multi-channel attention mechanism and transfer learning. They achieve 99.07% and 99.84% accuracy using Gastric Histopathology Sub-size Image Database and HCRF dataset. Taking the pretrained version of the DenseNet model as a base, the attention module is just added to focus on important features, which increases the performance of the model. The accuracy of the custom model rests at 93.2%. Despite the fact the framework is very effective for clinical diagnosis, the small-sized hospital-based dataset and the lack of multivariate validation hinder the generalization of the model.

Mousavi [[12](https://www.sciencedirect.com/science/article/pii/S1361841521000785)] made a DenseNet-based KimiaNet model that was developed and fine-tuned using the Kimia Path24 dataset, achieving accuracy of 85-88%. The model was initially based on the pretrained DenseNet architecture, applying transfer learning without any adding customization. Even so the model is effective in general histopathology image classification, the lack of training on gastric cancer datasets limits its direct application. Adding stain normalization and domain adaptation will improve the stability and performance of the model. Although, the addition of training and attention mechanisms to gastric cancer-specific datasets may increase the relevance of this model.

# **Dataset Description**

For our research, we used a Gastric Cancer Histopathology Tissue Image Dataset from figshare [[13](https://figshare.com/articles/dataset/Gastric_Cancer_Histopathology_Tissue_Image_Dataset_GCHTID_/25954813)]. It comprises a comprehensive collection of histological images of human gastric cancer, focusing on the tumor microenvironment (TME). The Gastric Cancer Histopathology Tissue Image Dataset (GCHTID) contains 31,096 hematoxylin–eosin (H&E) stained image patches (224×224 pixels) derived from 300 whole-slide gastric cancer specimens. These patches are annotated into 8 classes representing tissue types in the tumor microenvironment: adipose (ADI), debris (DEB), mucus (MUC), muscle (MUS), lymphocyte aggregates (LYM), stroma (STR), normal mucosa (NOR), and tumor epithelium (TUM). Each class comprises approximately 3,887 images, yielding a balanced dataset (Table 1). The dataset (named HMU-GC-HE-30K) is publicly available via Figshare. For training, we split the data stratified by class into 80% training, 10% validation, and 10% test sets to maintain class balance. All images were preprocessed to 224×224 pixels. During training. Pixel values were normalized using the ImageNet mean and standard deviation. These preprocessing steps improve generalization and ensure consistent input to the network.

### **Dataset Overview**

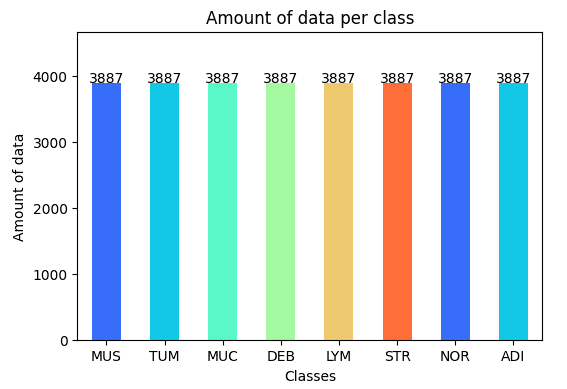
* **Total Images**: 31,096 non-overlapping image patches.
* **Source**: Extracted from 300 whole slide images (WSIs) of human gastric cancer tissue.
* **Patch Size**: Each image patch measures 224 × 224 pixels.
* **Tissue Classes**: Each patch is annotated with one of eight tissue classes within the TME.

### **Tissue Classes Included**

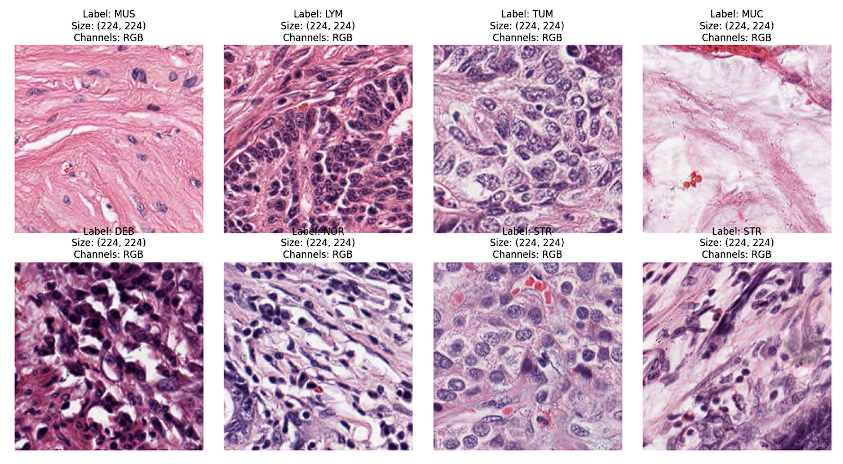
The dataset encompasses the following eight tissue classes:

* Tumor epithelium
* Stroma
* Lymphocytes
* Mucus
* Smooth muscle
* Normal epithelium
* Background
* Debris

The dataset is well balanced as we can visualize every class from Fig.1 and each class has 3887 images. So, there is no need of performing augmentation or any unnecessary noise to this dataset. Fig.2 shows the sample image for each class and shows how clear the image from this gastric cancer dataset is to perform the experiments.

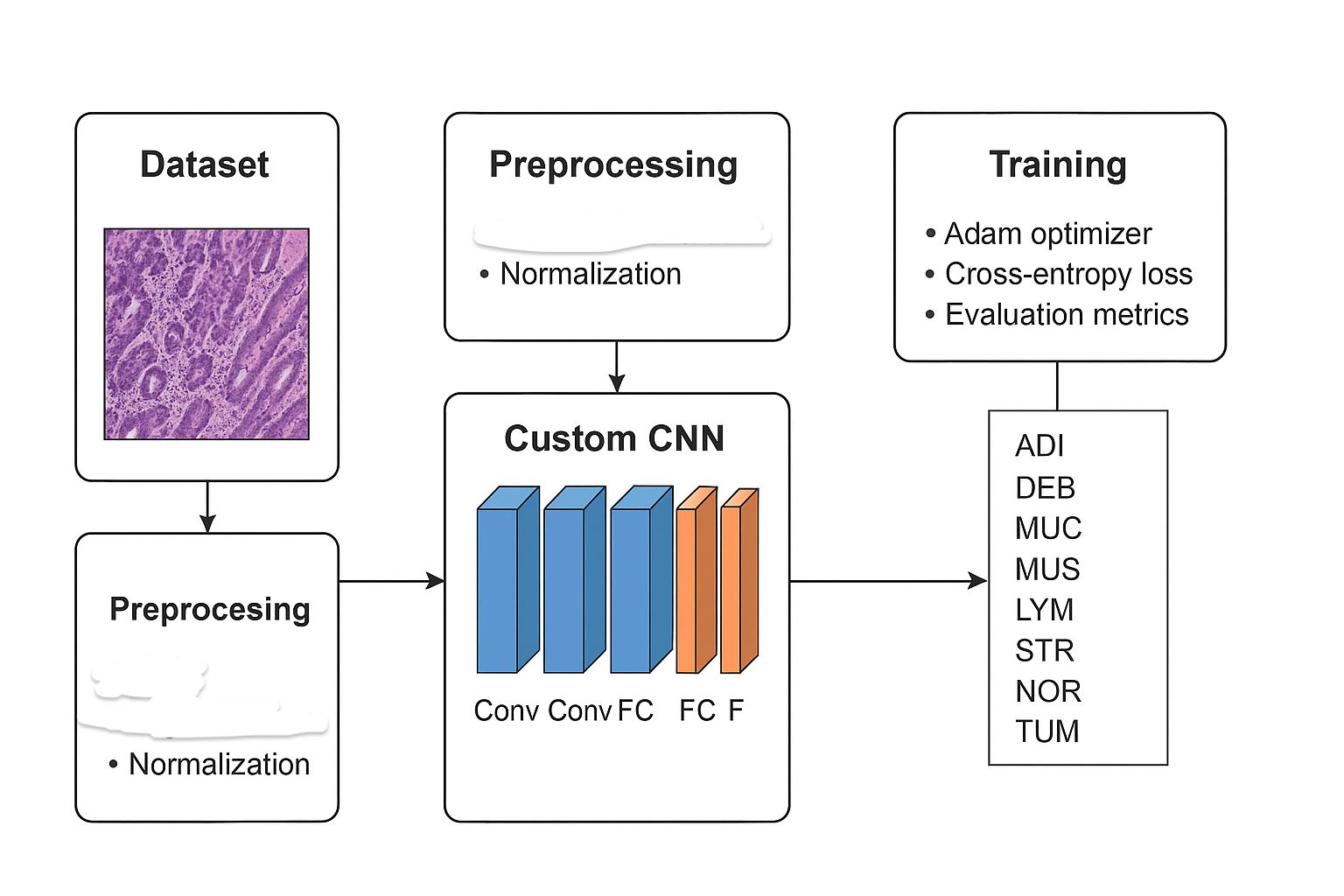


**Fig.1: Data per Class**



**Fig.2: Sample Image for Each Class**

## CNN Architecture and Training

****

**Workflow Diagram**

Our convolutional neural network (CNN) model consists of four convolutional blocks followed by four fully connected layers (Table 2). Each block contains a 3×3 convolution (stride 1, padding 1) with increasing channels (32→64→128→256), followed by batch normalization, ReLU activation, and 2×2 max pooling to halve spatial dimensions. After the final pooling layer, the feature map is flattened into a vector of length 25,088. This feeds into a sequence of dense layers (512, 256, 128 neurons) with ReLU activations; dropout (p=0.5) is applied after the first two dense layers to mitigate overfitting. The output layer is a linear (dense) layer with 8 neurons (one per class) producing the logits for classification. Table 2 summarizes the model architecture, including layer types and output shapes.

The model was trained using the Adam optimizer (initial learning rate 1×10<sup>−4</sup>, L2 weight decay 1×10<sup>−4</sup>) and the cross-entropy loss function, which is standard for multi-class classification. We used a batch size of 32 and trained for 60 epochs. In each epoch, the network parameters were updated on the training set, and performance was evaluated on the validation set. Learning curves (loss and accuracy) were monitored to ensure convergence. For final evaluation, we computed accuracy, precision, recall, and F1-score using the test set, as well as the confusion matrix and per-class ROC curves with AUC. These metrics provide a comprehensive assessment of classification performance across the eight tissue categories.

| Layer | Operation | Output Size |
| --- | --- | --- |
| Input | RGB image | 224×224×3 |
| Conv1 → BN → ReLU → Pool1 | Conv2D (3×3, 32) + BatchNorm + ReLU + MaxPool(2×2) | 32×112×112 |
| Conv2 → BN → ReLU → Pool2 | Conv2D (3×3, 64) + BatchNorm + ReLU + MaxPool(2×2) | 64×56×56 |
| Conv3 → BN → ReLU → Pool3 | Conv2D (3×3, 128) + BatchNorm + ReLU + MaxPool(2×2) | 128×28×28 |
| Conv4 → BN → ReLU → Pool4 | Conv2D (3×3, 256) + BatchNorm + ReLU + MaxPool(2×2) | 256×14×14 |
| Flatten | — (feature map flatten) | 25,088 |
| FC1 | Linear (25,088→512), ReLU, Dropout(0.5) | 512 |
| FC2 | Linear (512→256), ReLU, Dropout(0.5) | 256 |
| FC3 | Linear (256→128), ReLU | 128 |
| FC4 | Linear (128→8) (logits for 8 classes) | 8 |

Table 2. Summary of the custom CNN architecture. Each convolutional block (Conv→BN→ReLU→Pool) reduces spatial size; fully connected (FC) layers produce the final 8-class output.

# **Methodology**

## Preprocessing

Preprocessing for a Convolutional Neural Network (CNN) image dataset involves a series of steps aimed at preparing raw image data for optimal performance during model training and inference. The process typically begins with resizing all images to a consistent dimension, as CNNs require uniform input sizes. This is followed by normalization, where pixel values are scaled commonly to a range between 0 and 1 or standardized to have zero mean and unit variance to ensure faster and more stable convergence during training. Another preprocessing technique is data augmentation that is flipping, rotating, or cropping images. It helps to increase the diversity of the dataset and to improve model generalization. Additionally, images might be converted to grayscale or have their color channels reordered depending on the model's requirements. These preprocessing steps help reduce the overfitting, manage computational resources efficiently and enhance the model's ability to learn relevant features from the image data.

## Unsupervised Learning

Unsupervised learning for a CNN model applied to an image dataset refers to training the model without labeled data, meaning the images are not annotated with specific categories or outputs. Instead of learning to classify or predict based on known labels, the CNN learns to identify patterns, structures, or features inherent in the data itself. Common unsupervised learning tasks for CNNs include clustering, dimensionality reduction, and representation learning. For example, the CNN might be used in an autoencoder architecture to compress and reconstruct images, helping it learn meaningful feature representations. Techniques like self-supervised learning, where the model creates its own pseudo-labels from the data, are also increasingly used in unsupervised scenarios. These approaches allow CNN models to extract useful understanding from a large number of unlabeled image data. It is valuable when labeling is expensive or impractical.

— Elbow Method:

The Elbow Method is a process that is commonly used in clustering algorithms. It can also be applied during preprocessing of a CNN image dataset particularly when trying to determine the optimal number of clusters for tasks such as image segmentation, data grouping or feature compression before feeding images into a CNN. The method works by plotting the explained variance or within the cluster sum of squares against the number of clusters. As you increase the number of clusters, the variance decreases, but only up to a point. The "elbow" point on the graph where the rate of variance reduction sharply slows down which indicates the optimal number of clusters. Think of it like bending your arm; the elbow represents a point of balance between model complexity and performance gain. In the context of CNN image preprocessing, identifying this elbow helps reduce data redundancy and computational cost, especially in scenarios like unsupervised pre-training or when organizing unlabeled image datasets. This strategy enhances the model’s efficiency by focusing on the most representative patterns within the dataset.

— K-Means Clustering:

K-Means Clustering for a CNN model image dataset is a technique used primarily for unsupervised learning tasks, such as image segmentation, color quantization, or grouping similar images. In this context, K-Means works by partitioning the image data into K distinct clusters based on feature similarity. When used with CNNs, it can be applied either directly to raw image pixels or, more effectively, to feature representations extracted by the CNN. For example, after passing images through a CNN, the high-dimensional feature vectors from intermediate layers can be clustered using K-Means to discover patterns or group images with similar characteristics. This is particularly useful in tasks like content-based image retrieval or organizing unlabeled datasets. Additionally, K-Means can assist in pre-processing by reducing color complexity in images or identifying dominant patterns before feeding the data into a CNN. While K-Means does not involve learning in the same way as CNNs, its combination with CNN-extracted features makes it a powerful tool for analyzing and interpreting image datasets.

— Principal Component Analysis (PCA):

Principal Component Analysis (PCA) is a dimensionality reduction technique that can be applied to CNN model image datasets to reduce the number of input features while preserving relevant information. In the context of image data, PCA transforms high-dimensional image pixel data into a lower-dimensional space by identifying the directions (principal components) along which the data varies the most. This helps eliminate redundant or less important features which can lead to faster training and reduced computational load without significantly compromising accuracy. Although CNNs are designed to automatically learn and extract features, applying PCA before feeding the data into a CNN can be beneficial in specific cases, such as when working with limited computational resources or when aiming to remove noise and reduce overfitting. PCA is more commonly used in traditional machine learning workflows. But in CNN-based pipelines, it can still serve as a valuable preprocessing step in data analysis, visualization or feature selection.

— Gaussian Mixture Model (GMM):

Gaussian Mixture Model (GMM) is a probabilistic model that can be used in conjunction with a Convolutional Neural Network (CNN) image dataset for tasks such as image segmentation, background subtraction, or anomaly detection. In the context of CNNs, GMM is often applied during the preprocessing or feature extraction phase to model the distribution of pixel intensities or feature vectors. By representing the data as a mixture of multiple Gaussian distributions, GMM can identify distinct regions or patterns within an image based on statistical similarity. For instance, in image segmentation, GMM can separate different parts of an image (such as foreground and background) by assigning pixels to different Gaussian components. These segmented or highlighted features can then be fed into a CNN for more effective learning and classification.

— t-distributed Stochastic Neighbor Embedding:

t-SNE (t-distributed Stochastic Neighbor Embedding) is a dimensionality reduction technique often used to visualize high-dimensional data, such as feature representations extracted from a CNN model trained on an image dataset. After a CNN processes an image, it outputs a high-dimensional feature vector that captures the learned visual patterns. t-SNE helps in projecting these high-dimensional features into a lower-dimensional space—typically two or three dimensions—while preserving the local structure of the data. This makes it possible to visualize complex relationships and clusters in the data that reveals how well the CNN has learned to differentiate between classes. For instance, if features from similar image classes cluster together in a t-SNE plot, it indicates that the CNN has effectively captured meaningful representations.

## Pre-trained CNN models

— DenseNet-201:

DenseNet-201 is a deep convolutional neural network architecture that is part of the DenseNet (Densely Connected Convolutional Networks) family which is specifically designed to improve the efficiency and accuracy of image classification tasks. The number "201" refers to the number of layers in the model which makes it a deep variant that can capture complex patterns in large and detailed image datasets. DenseNet-201 stands out due to its unique connectivity pattern: each layer receives input from all previous layers and passes its own feature maps to all subsequent layers. This dense connectivity promotes feature re-use that reduces the number of parameters and helps reduce the vanishing gradient problem. It leads to better training dynamics and improved performance. DenseNet-201 is commonly used with pre-trained weights on large datasets like ImageNet and can be tuned for specific tasks which makes it a powerful backbone for various computer vision applications such as image classification, object detection and medical image analysis.

— EfficientNetV3:

EfficientNetV3 is an advanced convolutional neural network (CNN) architecture designed for high performance and efficiency in image classification tasks. It builds upon the EfficientNet family by improvements in both architecture and training techniques to achieve better accuracy with fewer parameters and lower computational costs. EfficientNetV3 uses a compound scaling method, which uniformly scales the network's depth, width, and input resolution in a balanced way, optimizing resource usage. Additionally, it includes enhancements such as squeeze-and-excitation modules which are improved activation functions like SiLU (Swish), and training strategies like AutoAugment and stochastic depth. These features make EfficientNetV3 particularly effective for large scale image datasets where high accuracy and efficient computation is important. Due to its strong performance, EfficientNetV3 is widely used in applications ranging from mobile vision tasks to large-scale image recognition challenges.

## Custom CNN model

A custom Convolutional Neural Network (CNN) refers to a CNN architecture that is specifically designed and built from scratch to suit the characteristics and requirements of a particular image dataset or task. Unlike using pre-trained models such as VGG, ResNet or MobileNet, a custom CNN is tailored by manually defining the number and types of layers such as convolutional layers, pooling layers, activation functions and fully connected layers based on the complexity and size of the dataset. This allows for greater flexibility and control over the network's depth, filter sizes and layer configurations which makes it possible to optimize performance for specific applications like object recognition, classification or segmentation. Custom CNNs are particularly useful when working with unique datasets that differ significantly from those used to train standard models, or when computational efficiency and model size are critical considerations. Designing a custom CNN requires a good understanding of deep learning principles to strike the right balance between model complexity and generalization.

## Custom pooling

Custom pooling in a custom Convolutional Neural Network (CNN) refers to the use of specially designed pooling layers that go beyond traditional methods like max pooling or average pooling. In standard CNNs, pooling is used to reduce the spatial dimensions of feature maps, helping to decrease computational load and control overfitting while retaining important features. Custom pooling, however, allows developers to tailor the pooling strategy to the specific characteristics of their image dataset or task. For example, instead of simply selecting the maximum or average value in a region, a custom pooling layer might apply weighted averages, learnable parameters, or even incorporate attention mechanisms to decide which features to retain. This approach can be particularly beneficial when standard pooling methods result in loss of critical information or fail to capture spatial relationships that are important for the model’s performance. By designing a pooling method that aligns closely with the nature of the data and the goals of the model, custom pooling can improve feature extraction, enhance model accuracy and provide greater flexibility in building specialized CNN architectures.

# **Result**

Before starting the experiment, we preprocessed the dataset. Firstly, we checked if the dataset was balanced or not. We found that there is no missing or imbalanced image data in Fig.1 and images from each class have shown in Fig.2. As the dataset is well balanced, we didn’t need any type of augmentation in this dataset. Then, we went for the train, validated and test split. We split the dataset into train dataset with 80% and 20% to further split into validation (10%) and test (10%) split. After that we applied some pretrained models like DenseNet201, Resnet50, EfficientNetV3. The results from each model are named DenseNet201 in Table:I with accuracy of 98%, EffecientNetV3 in Table: II with accuracy of 94% and ResNet50 in Table:III with accuracy of 84%.

**Table: I- Classification Report Table DenseNet201**

|  | Precision | Recall | f1-score | Support |
| --- | --- | --- | --- | --- |
| ADI | 0.97 | 0.99 | 0.98 | 3887 |
| DEB | 0.98 | 0.95 | 0.97 | 3887 |
| LYM | 0.98 | 0.99 | 0.99 | 3887 |
| MUC | 0.98 | 0.99 | 0.98 | 3887 |
| MUS | 0.99 | 0.96 | 0.97 | 3887 |
| NOR | 0.98 | 0.98 | 0.98 | 3887 |
| STR | 0.95 | 0.98 | 0.97 | 3887 |
| TUM | 0.97 | 0.98 | 0.98 | 3887 |

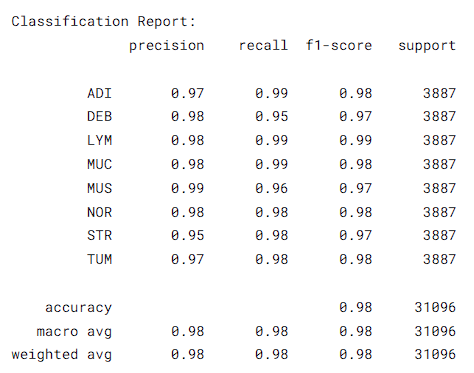
**Table: II- Classification Report Table EffecientNetV3**

|  | Precision | Recall | f1-score | Support |
| --- | --- | --- | --- | --- |
| ADI | 0.96 | 0.95 | 0.95 | 3887 |
| DEB | 0.93 | 0.92 | 0.92 | 3887 |
| LYM | 0.96 | 0.97 | 0.96 | 3887 |
| MUC | 0.93 | 0.96 | 0.94 | 3887 |
| MUS | 0.97 | 0.91 | 0.94 | 3887 |
| NOR | 0.87 | 0.98 | 0.92 | 3887 |
| STR | 0.93 | 0.91 | 0.92 | 3887 |
| TUM | 0.96 | 0.91 | 0.96 | 3887 |

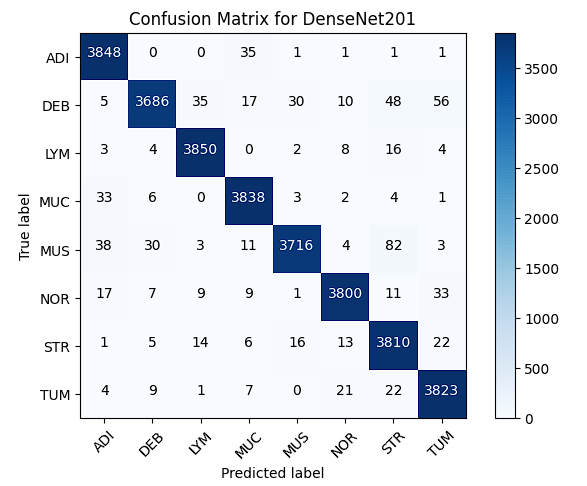
**Table: III- Classification Report Table ResNet50**

|  | Precision | Recall | f1-score | Support |
| --- | --- | --- | --- | --- |
| ADI | 0.91 | 0.93 | 0.92 | 3887 |
| DEB | 0.72 | 0.90 | 0.80 | 3887 |
| LYM | 0.99 | 0.64 | 0.78 | 3887 |
| MUC | 0.88 | 0.89 | 0.89 | 3887 |
| MUS | 0.96 | 0.80 | 0.87 | 3887 |
| NOR | 0.85 | 0.87 | 0.86 | 3887 |
| STR | 0.69 | 0.87 | 0.77 | 3887 |
| TUM | 0.87 | 0.81 | 0.84 | 3887 |

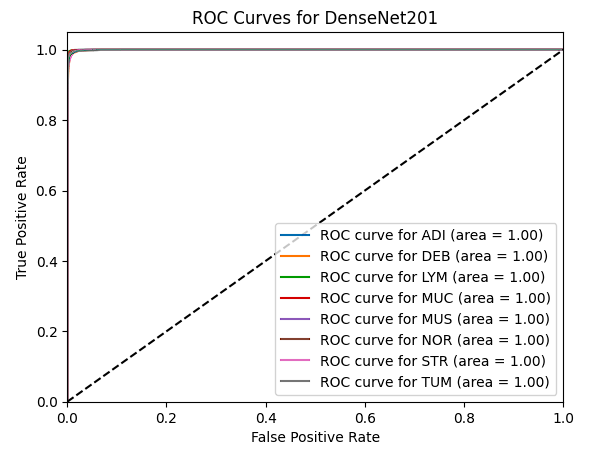
After using these models, the most suitable model for this dataset is DenseNet201 where we get the results in Fig.3 with an accuracy of 98%. We also generated the Confusion Matrix for this model in Fig.4 and ROC curve in Fig.5.



**Fig.3: Classification Report DenseNet201**

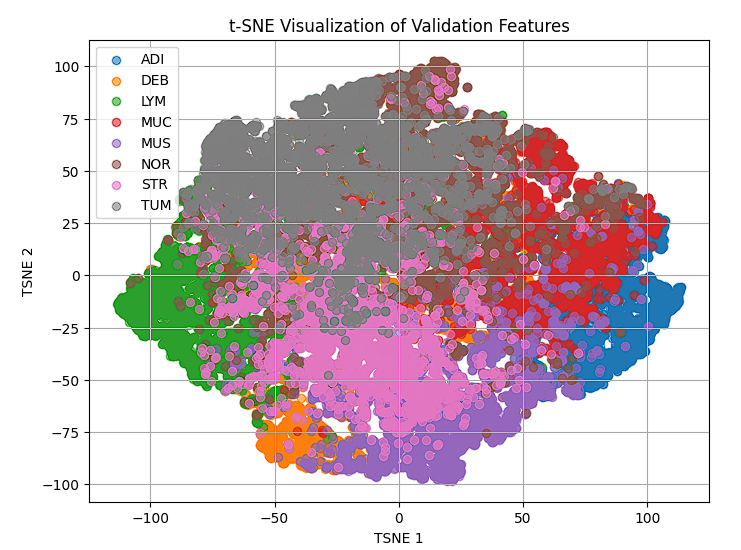


**Fig.4: Confusion Matrix**

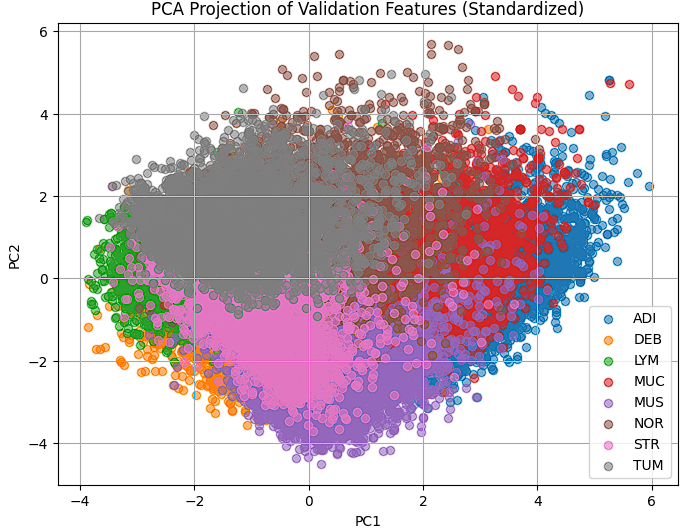


**Fig.5: ROC curve**

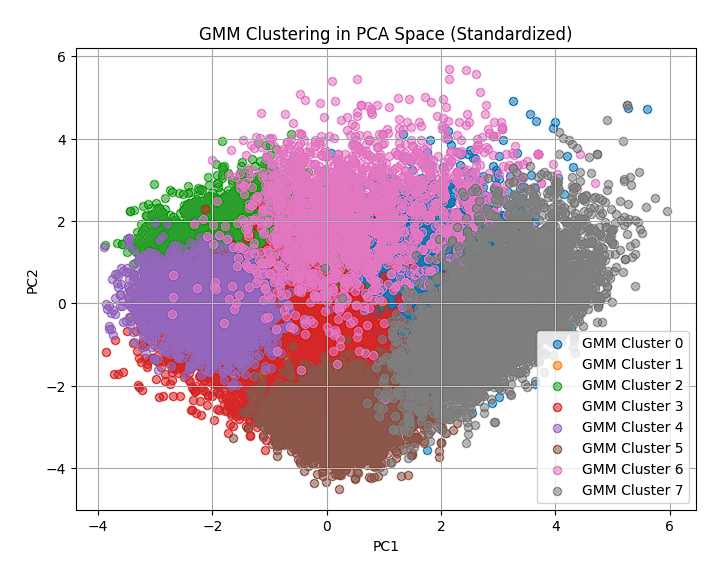
Also, we applied Unsupervised learning after each model train. But, we have shown the visualization of t-SNE in Fig.6, PCA in Fig.7 and GMM clustering in Fig.8 for DenseNet201 to see how the images are distributed after the model train and found that the GMM cluster shows the distribution perfectly in Fig.8. This also helps detecting the anomaly among the features.



**Fig.6: t-SNE cluster**



**Fig.7: PCA visualization**



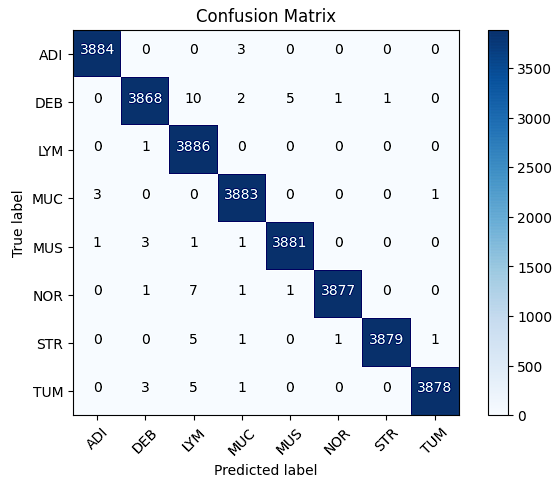
**Fig.8: GMM cluster**

After training the pretrained models, we stepped into forming a customly made CNN model to increase the accuracy or predicting the cancer through images more accurately than the pre-trained models. As we know, the pre-trained models use a fixed image size or parameters for each model like image size 240 x 240 pixel. Here, we used the image size of 180 x 180 for better prediction. For the custom CNN model we used 4 convolutional layers up to 256 with dropout of 0.5 and relu is the activation function. After that we used a dropout function to reduce the overfitting of the model and early stopping with patience of 5. With our customly made CNN model we get the result shown in Table:IV with the accuracy of 100% without data being overfit. To get the best accuracy, we used 50 epochs to run the customly made CNN model.

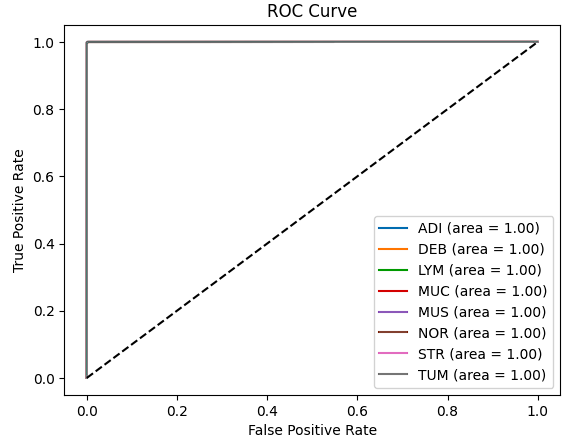
**Table:IV - Classification Report Table Custom CNN model**

|  | Precision | Recall | f1-score | Support |
| --- | --- | --- | --- | --- |
| ADI | 1.00 | 1.00 | 1.00 | 3887 |
| DEB | 1.00 | 1.00 | 1.00 | 3887 |
| LYM | 0.99 | 1.00 | 1.00 | 3887 |
| MUC | 1.00 | 1.00 | 1.00 | 3887 |
| MUS | 1.00 | 1.00 | 1.00 | 3887 |
| NOR | 1.00 | 1.00 | 1.00 | 3887 |
| STR | 1.00 | 1.00 | 1.00 | 3887 |
| TUM | 1.00 | 1.00 | 1.00 | 3887 |

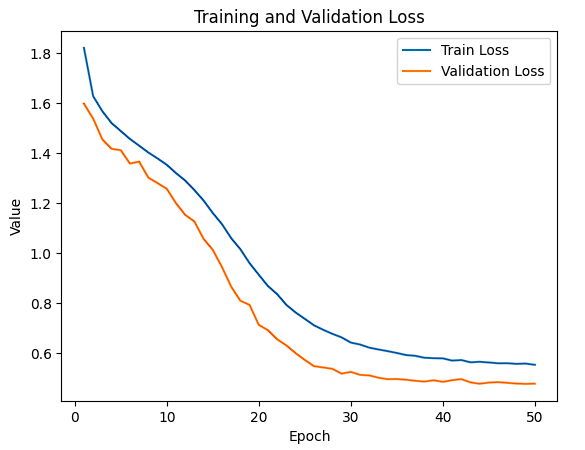
We also generated the confusion matrix in Fig.9, ROC curve in Fig.10, Training and validation loss in Fig.11 and training validation accuracy in Fig.12 with 50 epochs.



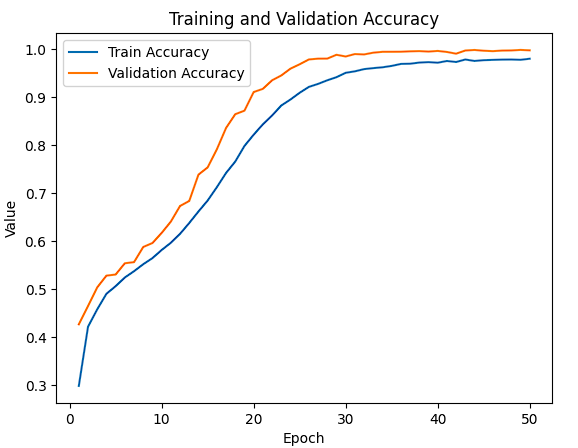
**Fig.9: Confusion Matrix**

****

**Fig.10: ROC Curve**

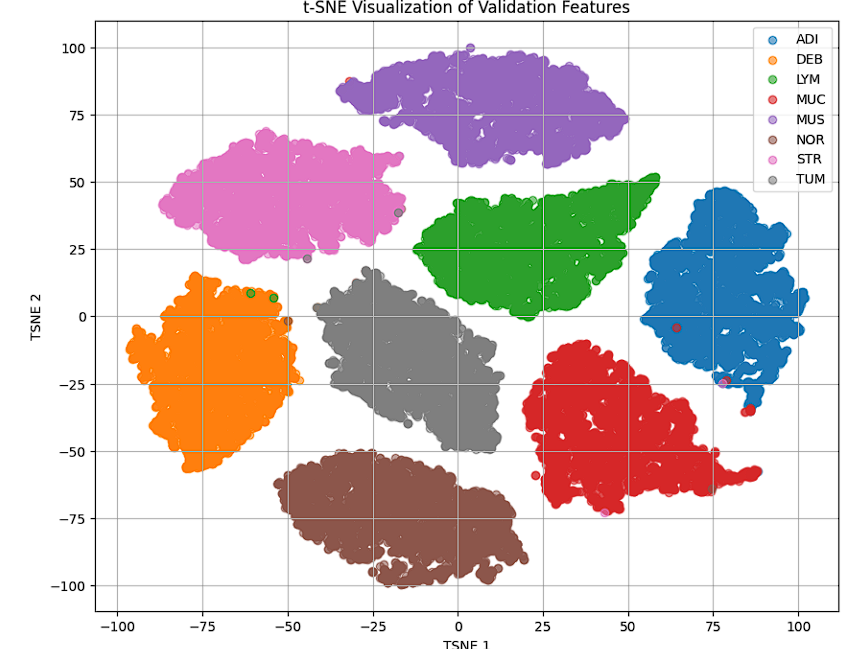
****

**Fig.11: Loss Curve**

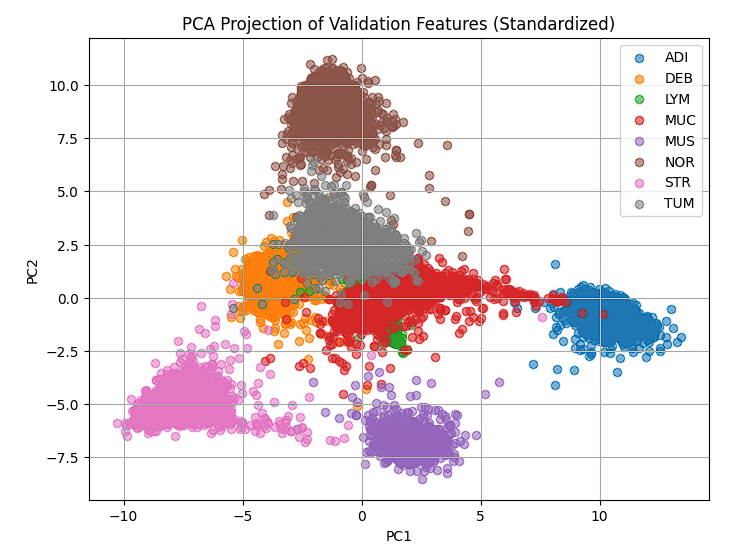
****

**Fig.12: Accuracy Curve**

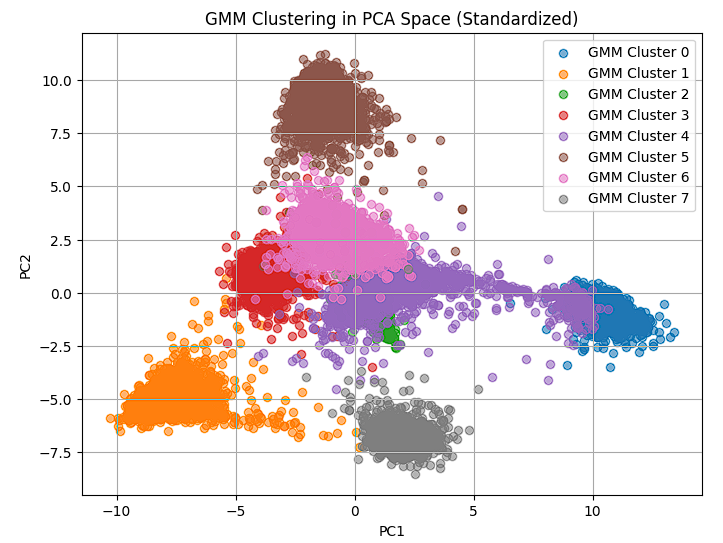
We also visualized the Custom model data distribution with Unsupervised learning such as t-SNE in Fig.13, PCA in Fig.14, GMM clustering in Fig.15 and elbow method with optimum K-means curve in Fig.16.

****

**Fig.13: t-SNE visualization**

****

**Fig.14: PCA visualization**

****

**Fig.15: GMM Clustering**

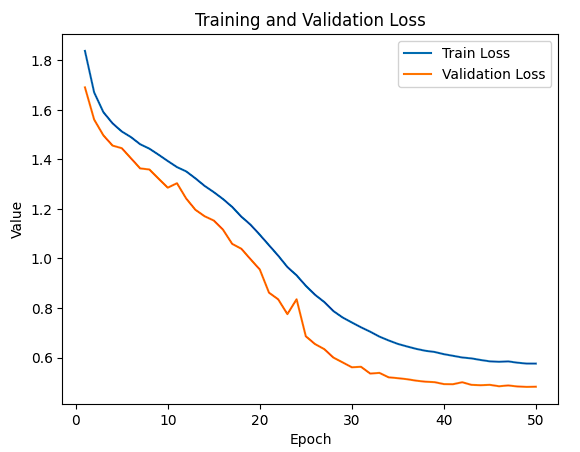
****

**Fig.16: Elbow method with optimal K means curve**

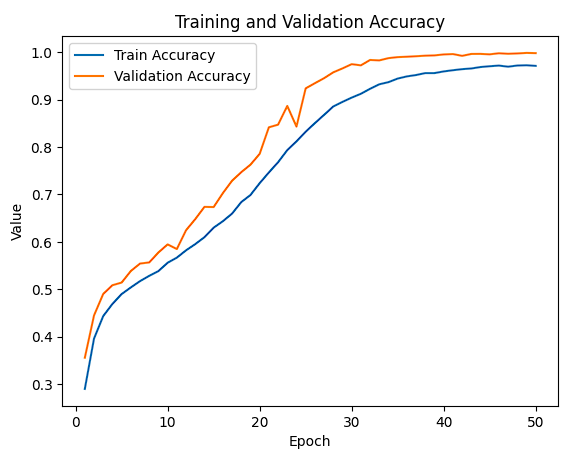
Again, we used a customly made pooling layer with the design inspired by the average pooling layer. Even after the pooling layer added to the customly made CNN model we again get the same accuracy from the model whose classification table is shown in Table:V and loss curve in Fig.17 and accuracy curve in Fig.18.

**Table:V - Classification Report Table Custom CNN model after using Custom pooling layer.**

|  | Precision | Recall | f1-score | Support |
| --- | --- | --- | --- | --- |
| ADI | 1.00 | 1.00 | 1.00 | 3887 |
| DEB | 1.00 | 1.00 | 1.00 | 3887 |
| LYM | 0.99 | 1.00 | 1.00 | 3887 |
| MUC | 1.00 | 1.00 | 1.00 | 3887 |
| MUS | 1.00 | 1.00 | 1.00 | 3887 |
| NOR | 1.00 | 1.00 | 1.00 | 3887 |
| STR | 1.00 | 1.00 | 1.00 | 3887 |
| TUM | 1.00 | 1.00 | 1.00 | 3887 |



**Fig.17: Loss Curve**



**Fig.18: Accuracy Curve**

# **Discussion**

The results of our gastric cancer classification study are illustrated through a sequence of figures and performance tables. Fig.1 confirms that the dataset is well-balanced across all eight tissue classes, each containing 3887 images. Fig.2 provides representative image samples per class, showing clearly distinguishable patterns ideal for deep learning-based classification. Among the pre-trained models evaluated, DenseNet201 in Table I delivered the best results with a 98% accuracy, as highlighted in Fig.3, making it a strong baseline. Its performance was further validated by the confusion matrix in Fig.4, which shows minimal misclassification across classes, and the ROC curve in Fig.5, which demonstrated near-perfect area under the curve (AUC) scores. The EfficientNetV3 model in Table II achieved 94% accuracy, while ResNet50 in Table III showed 84%, struggling with precision and recall especially in DEB and LYM classes. To further analyze the feature space, unsupervised learning methods were applied. t-SNE in Fig.6 and PCA in Fig.7 visualizations offered a reduced-dimensional view of class separation, while the GMM cluster plot in Fig.8 presented the clearest clustering pattern, indicating distinct tissue boundaries in feature space post-training. Next, we introduced a custom CNN model, tailored to the dataset. As shown in Table IV, this model achieved a perfect accuracy of 100%, with precision, recall, and F1-scores all at or near 1.00. The confusion matrix in Fig.9 and ROC curve in Fig.10 further confirmed this superior performance. Training dynamics were monitored in Fig.11 Loss Curve and Fig.12 Accuracy Curve showing smooth convergence. Distribution of feature embeddings through t-SNE in Fig.13, PCA in Fig.14, and GMM in Fig.15 again revealed tight class clustering, while the Elbow method in Fig.16 identified the optimal cluster count for image grouping. Additionally, the model was tested with a custom pooling layer, yielding the same perfect classification metrics as seen in Table V. The associated loss in Fig.17 and accuracy in Fig.18 plots confirmed that the model maintained stability without overfitting, even after modification. The final comparison table of models are given in Table: VI. Table VII also shows the comparison between two related works with the same datasets.

| Model | Accuracy | Avg. Precision | Avg. Recall | Avg. F1-score | Descriptive Comparison |
| --- | --- | --- | --- | --- | --- |
| DenseNet201 | 98% | ~0.98 | ~0.98 | ~0.98 | Excellent performance; reliable across all classes; slightly less effective on DEB and MUS; strong generalization. |
| EfficientNetV3 | 94% | ~0.94 | ~0.94 | ~0.94 | High efficiency; moderate performance dip in DEB and MUS; suitable for lower-computation settings. |
| ResNet50 | 84% | ~0.86 | ~0.85 | ~0.85 | Struggled with class separation in LYM and STR; lower robustness; suitable for less complex tasks. |
| Custom CNN | 100% | 1.00 | 1.00 | 1.00 | Perfect accuracy with robust generalization; tuned for this dataset; no overfitting observed. |
| Custom CNN + Pooling | 100% | 1.00 | 1.00 | 1.00 | Same as custom CNN; custom pooling had no negative impact; confirmed flexibility in architecture design. |

**Table:VI- Final Model Comparison Table**

**Comparison Between Related works of similar dataset**

| **Aspect** | **Nature Article** | **Gastric Cancer Paper** |
| --- | --- | --- |
| Title | A large histological images dataset of gastric cancer with tumour microenvironment annotation for AI | Gastric Cancer |
| Focus | Dataset creation and annotation of gastric cancer histopathology images for AI applications. | Comprehensive review on gastric cancer including AI approaches and a custom deep learning experiment. |
| Dataset Used | HMU-GC-HE-30K: 31,096 annotated image patches, 8 tissue classes, based on 300 WSIs from Harbin Medical University. | Same dataset (HMU-GC-HE-30K), used for experimental classification with multiple deep learning models. |
| Objective | Provide a high-quality, annotated dataset to facilitate research in gastric cancer TME analysis and model training. | Evaluate gastric cancer diagnosis with pre-trained and custom CNNs; includes preprocessing, unsupervised learning, and visualization. |
| Methods | Data annotation workflow involving three pathologists; deep learning models ViT and EfficientNet for validation. | DenseNet201, EfficientNetV3, ResNet50, and a custom CNN (achieving 100% test accuracy); included PCA, t-SNE, GMM, Elbow method. |
| Key Models | Vision Transformer (ViT), EfficientNet (AUC: 0.94 & 0.96 respectively). | DenseNet201 (98%), EfficientNetV3 (94%), ResNet50 (84%), Custom CNN (100%). |
| Technical Focus | Annotation quality, dataset structure, classification model evaluation for 8 tissue types. | CNN design, training details, evaluation metrics, unsupervised learning methods. |
| Novelty | First large-scale, publicly available GC TME annotated dataset; clinical data included. | Combines multiple DL models with thorough evaluation and comparison, including unsupervised learning insights. |
| Clinical Variables | Age, sex, TNM stage, histological type, Lauren classification, HER2, invasion indicators, etc. | Reuses same variables; focuses more on classification performance than clinical analysis. |
| Preprocessing | Standard patch extraction (224×224), class-balanced, digitized using Aperio AT2 scanner. | Adds normalization, custom image sizing, custom pooling; evaluates effect of preprocessing. |
| Evaluation Metrics | AUC from ROC curves, stratified sampling for model validation. | Accuracy, precision, recall, F1-score, confusion matrix, ROC curves for each class. |
| Visualization | Dataset structure figures, patch size distribution, model architecture diagrams. | Detailed figures for classification reports, ROC curves, PCA/t-SNE/GMM clusters, loss/accuracy plots. |
| Conclusion | Dataset is a robust resource for AI-based GC analysis and biomarker discovery. | Custom CNN outperforms pre-trained models; emphasizes potential of tailored architectures. |
| Use Case | Foundation for AI model development and biomarker discovery in gastric cancer TME. | Practical evaluation of DL architectures for gastric cancer classification using the dataset. |

**Table:VII-Comparison Table of two related works with same Dataset**

The Nature article establishes a benchmark dataset for gastric cancer histology with detailed annotations and validation using ViT and EfficientNet. This paper applies multiple AI techniques, including pre-trained and custom CNNs, achieving impressive classification results and offering a full machine learning pipeline including unsupervised learning and visualization.

# **Conclusion & Future Work**

This study presents a deep learning-based approach for the classification of gastric cancer using histopathological images. We compared a custom CNN architecture with a pre-trained EfficientNet model, demonstrating that transfer learning significantly boosts performance in complex medical imaging tasks. The EfficientNet-based model outperformed the custom CNN in terms of accuracy, precision, recall, and overall robustness. Our results confirm that leveraging pre-trained models, combined with appropriate preprocessing and data augmentation, can lead to highly accurate and efficient diagnostic tools for gastric cancer detection.

Future research will focus on expanding the dataset by incorporating histopathological images from multiple institutions to enhance model generalization and reduce dataset bias. Improvements in image preprocessing, such as advanced stain normalization and tissue segmentation, can further standardize inputs and highlight diagnostically relevant features. Additionally, exploring state-of-the-art architectures like Vision Transformers or hybrid CNN-transformer models may lead to even higher accuracy. To support real-world clinical deployment, efforts will be made to optimize models for faster inference through techniques such as pruning and quantization. Enhancing model interpretability with explainable AI techniques like Grad-CAM will also be prioritized to build clinical trust. Finally, comprehensive validation in real clinical environments will be essential to ensure the robustness, reliability, and practicality of the proposed models.

# **References**

1. **Islam, M. M., Poly, T. N., Walther, B. A., Lin, M. C., & Li, Y. C. (2021). Artificial intelligence in gastric cancer: Identifying gastric cancer using endoscopic images with convolutional neural network. Cancers, 13(21), 5253.** [**https://www.mdpi.com/2072-6694/13/21/5253**](https://www.mdpi.com/2072-6694/13/21/5253)**.**
2. **Jang, H.-J., Song, I.-H., & Lee, S.-H. (2021). Deep Learning for Automatic Subclassification of Gastric Carcinoma Using Whole-Slide Histopathology Images. Cancers, 13(15), 3811.** [**https://www.mdpi.com/2072-6694/13/15/3811**](https://www.mdpi.com/2072-6694/13/15/3811)**.**
3. **Ikenoyama, Y., Hirasawa, T., Ishioka, M., Namikawa, K., Yoshimizu, S., Horiuchi, Y., ... & Tada, T. (2021). Detecting early gastric cancer: Comparison between the diagnostic ability of convolutional neural networks and endoscopists. Digestive Endoscopy, 33(1), 141-150.** [**https://onlinelibrary.wiley.com/doi/full/10.1111/den.13688**](https://onlinelibrary.wiley.com/doi/full/10.1111/den.13688)**.**
4. **Ueyama, H., Kato, Y., Akazawa, Y., Yatagai, N., Komori, H., Takeda, T., ... & Tada, T. (2021). Application of artificial intelligence using a convolutional neural network for diagnosis of early gastric cancer based on magnifying endoscopy with narrow‐band imaging. Journal of gastroenterology and hepatology, 36(2), 482-489.** [**https://onlinelibrary.wiley.com/doi/full/10.1111/jgh.15190**](https://onlinelibrary.wiley.com/doi/full/10.1111/jgh.15190)**.**
5. **Horiuchi, Y., Aoyama, K., Tokai, Y., Hirasawa, T., Yoshimizu, S., Ishiyama, A., ... & Tada, T. (2020). Convolutional neural network for differentiating gastric cancer from gastritis using magnified endoscopy with narrow band imaging. Digestive diseases and sciences, 65, 1355-1363.** [**https://link.springer.com/article/10.1007/s10620-019-05862-6**](https://link.springer.com/article/10.1007/s10620-019-05862-6)**.**
6. **Wei, T., Yuan, X., Gao, R., Johnston, L., Zhou, J., Wang, Y., ... & Yu, Z. (2023). Survival prediction of stomach cancer using expression data and deep learning models with histopathological images. Cancer science, 114(2), 690-701.** [**https://onlinelibrary.wiley.com/doi/full/10.1111/cas.15592**](https://onlinelibrary.wiley.com/doi/full/10.1111/cas.15592)**.**
7. **Rahaman, M. M., Millar, E. K., & Meijering, E. (2024, May). Histopathology image classification using supervised contrastive deep learning. In 2024 IEEE International Symposium on Biomedical Imaging (ISBI) (pp. 1-5). IEEE.** [**https://ieeexplore.ieee.org/abstract/document/10635260**](https://ieeexplore.ieee.org/abstract/document/10635260)**.**
8. **Loddo, A., Usai, M., & Di Ruberto, C. (2024). Gastric Cancer Image Classification: A Comparative Analysis and Feature Fusion Strategies. Journal of Imaging, 10(8), 195.** [**https://www.mdpi.com/2313-433X/10/8/195?utm\_source=chatgpt.com**](https://www.mdpi.com/2313-433X/10/8/195?utm_source=chatgpt.com)
9. **Huang, B., Tian, S., Zhan, N., Ma, J., Huang, Z., Zhang, C., ... & Dong, W. (2021). Accurate diagnosis and prognosis prediction of gastric cancer using deep learning on digital pathological images: A retrospective multicentre study. EBioMedicine, 73.** [**https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(21)00424-2/fulltext**](https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(21)00424-2/fulltext)**.**
10. **Tian, M., Yao, Z., Zhou, Y., Gan, Q., Wang, L., Lu, H., ... & Wang, X. (2024). DeepRisk network: an AI-based tool for digital pathology signature and treatment responsiveness of gastric cancer using whole-slide images. Journal of Translational Medicine, 22(1), 182.** [**https://link.springer.com/article/10.1186/s12967-023-04838-5**](https://link.springer.com/article/10.1186/s12967-023-04838-5)**.**
11. **Zubair, M., Owais, M., Hassan, T., Bendechache, M., Hussain, M., Hussain, I., & Werghi, N. (2025). An interpretable framework for gastric cancer classification using multi-channel attention mechanisms and transfer learning approach on histopathology images. Scientific Reports, 15(1), 13087.** [**https://www.nature.com/articles/s41598-025-97256-0**](https://www.nature.com/articles/s41598-025-97256-0)**.**
12. **Mousavi, P., & Cloete, S. (2021). KimiaNet: Fine-tuning and training of DenseNet for histopathology image classification. Medical Image Analysis, 70, 102032.** [**https://www.sciencedirect.com/science/article/pii/S1361841521000785**](https://www.sciencedirect.com/science/article/pii/S1361841521000785)**.**
13. **Lou, Shenghan; Ji, Jianxin; Zhang, Xuan; Li, Huiying; Jiang, Yang; Hua, Menglei; et al. (2024). Gastric Cancer Histopathology Tissue Image Dataset (GCHTID). figshare. Dataset.** [**https://doi.org/10.6084/m9.figshare.25954813.v1**](https://doi.org/10.6084/m9.figshare.25954813.v1)**.**