Project Report

Information Retrieval (CS60092)

Project Report

Information Retrieval (CS60092)

Breast Cancer Image Retrieval

Group – 18

D G Pranathi - 19CS10075

Sonu Kumar Yadav - 20CS10061

Bhagoti - 20CS30010

Abstract:

In this project, our aim is to develop an image retrieval system for the early detection and diagnosis of breast cancer. Breast cancer is a common and deadly form of cancer that affects women worldwide, and early detection is crucial in ensuring successful treatment outcomes. The proposed system employs deep learning techniques, specifically convolutional neural networks (CNNs), for the accurate classification of breast cancer images. A content-based image retrieval (CBIR) technique is used to retrieve similar images from the database. The system was evaluated on a dataset of breast cancer images, achieving high accuracy in both classification and retrieval.

Introduction:

Breast cancer image retrieval is a technique used in medical imaging to retrieve and analyze images of breast tissue. It involves the use of computer algorithms and machine learning techniques to analyze and retrieve images of breast cancer. The main purpose of breast cancer image retrieval is to assist physicians in making accurate diagnoses and treatment decisions by providing them with access to large amounts of medical imaging data.

Breast cancer image retrieval typically involves several steps, including image acquisition, pre-processing, feature extraction, and image retrieval. During image acquisition, medical imaging technologies such as mammography, ultrasound, and magnetic resonance imaging (MRI) are used to capture images of the breast tissue. These images are then pre-processed to remove any artifacts or noise that may interfere with the analysis

Next, feature extraction techniques are used to identify and extract relevant features from the images. These features can include the shape, texture, and intensity of the breast tissue. Machine learning algorithms are then used to analyze these features and generate a set of relevant features that can be used to identify and retrieve similar images.

Finally, image retrieval techniques are used to search through a database of medical images and retrieve images that are similar to the ones being analysed. This can be done using a variety of techniques, including similarity-based retrieval and content-based retrieval.

Breast cancer image retrieval has the potential to significantly improve the accuracy and speed of breast cancer diagnosis and treatment. By providing physicians with access to large amounts of medical imaging data, they can make more informed decisions about the best course of treatment for each individual patient.

In earlier research, Elaheh Mahraban Nejad et al. performed Content Based Medical Image Retrieval on breast cancer images using the

BreakHis dataset, utilising the VGG-19 pre-trained deep CNN model for feature extraction and Euclidean Distance to determine the similarity distance of the captured images. The highest trial results from the five studies carried out produced a MAP value of 0.80 [5]. Similar research on CBMIR on breast cancer images was carried out by Yun Gu et al. using the BreakHis dataset. The approach used in this study, CNN-based Densely-Connected Multi-Magnification Hashing (DCMMH), applies different magnification levels to the applied data, including 40X, 100X, 200X, and 400X. The two techniques, such as low-magnification and high-magnification pictures, are further applied in this work.

With an average MAP for each magnification level of 0.95, the test results show that the overall employed methodologies were capable of producing an enhanced accuracy in the picture retrieval task compared to the prior model. On the main-class dataset, this study hardly manages to retrieve any images. Moreover, Yun Gu, et al. conducted a new study in 2019 using the same dataset as the previous one. This study uses this research as its primary source of information. In this study, Yun Gu and colleagues carried out a retrieval technique for the main-class and sub-class categories of class datasets. Multi-Magnification Correlation Hashing was the technique employed (MMCH) Overall, the main-class dataset performance of the suggested technique was satisfactory, with the highest MAP value of 0.9416.

Datasets

The utilized dataset in this study is BreakHis dataset retrieved from https://web.inf.ufpr.br/vri/databases/breast-cancerhistopathological-database-breakhis/. The Breast Cancer Histopathological Database (BreakHis) is a publicly available dataset of histopathological images of breast tumor tissues. The dataset contains 7,909 digitized histological images of breast tumor tissues, which are categorized into two main classes: benign and malignant. The images were obtained from 82 patients, with each patient having 50 images of both benign and malignant tumors. The images were captured at different magnification sizes (40X, 100X, 200X, and 400X) and the tissue samples were stained with Hematoxylin and Eosin (H&E).

Training:

In this study, the training process will first start with an image separation method where 80% of the images are used for training, 10% are used for validation, and the remaining 10% are used to measure the retrieval value (test). The CNN-based Autoencoder method will be used in this study to transform the original image into a reconstructed image through the 2 phases of encoder and decoder. The encoder network will take in the input image and progressively reduce its dimensions by applying convolutional layers followed by pooling layers. The number of convolutional and pooling layers, as well as their parameters, will depend on the specific dataset and problem at hand. The encoder network will end with a bottleneck layer that will represent the compressed representation of the input image. The decoder network will take the compressed representation from the encoder network and reconstruct the original image by applying deconvolutional layers followed by upsampling layers. The number of deconvolutional and upsampling layers, as well as their parameters, will depend on the specific dataset and problem at hand. The decoder network will end with a reconstruction layer that will output the reconstructed image. The two networks will be connected such that the output of the encoder network will serve as the input to the decoder network. The loss function for the autoencoder will be the difference between the original and reconstructed images, and the model will be trained using backpropagation to minimize this loss. The hyperparameters of the model, such as the learning rate, batch size, and number of epochs, will need to be optimized through experimentation to achieve the best performance.

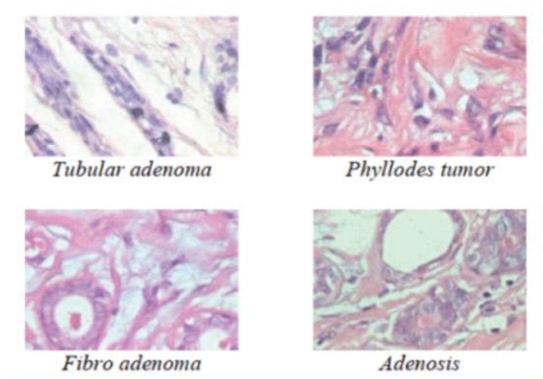

Figure 1: Example of Benign Tumor Image

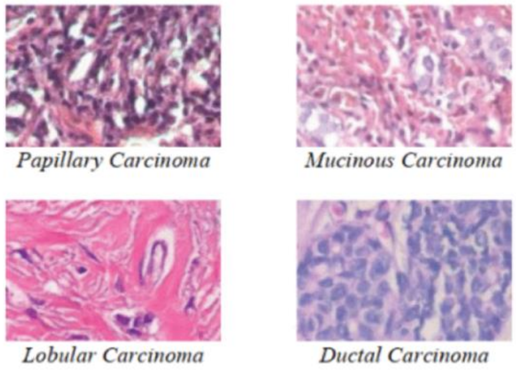

Figure 2: Example of Malignant Tumor Image

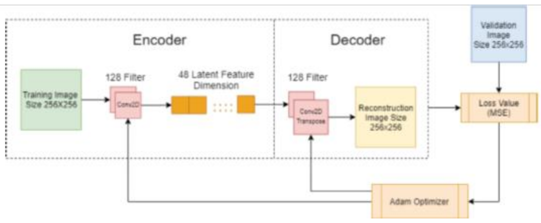

Figure 3: Flow of Training Stages

At the encoder stage, the input data will be converted into smaller dimensions or also entitled as compression, which in this study will be converted into a latent dimension with a size of 48 nodes by using Conv2D(2D Convolution Layer). Conv2D layer generates a tensor of outputs by convolving a convolution kernel with the layer input, in this study two modified default parameter were used. The first one is strides, stride is set to 2 for striding of the convolution along the height and width, another one is using “same” padding value which mean padding convolution with zero values is evenly distributed to the left/right or up/down of the input, resulting in output with the same height or width dimension as the input. The number of filters used is 128 with a size of 3 X 3 pixels. LeakyReLU is a popular activation function used in many deep learning models, including CNN-based autoencoder applications for breast cancer image retrieval. The LeakyReLU activation function, which adds a slight slope to the negative part of the function, is a variation on the ReLU (Rectified Linear Unit) activation function. By doing this, it is possible to prevent the "dying ReLU" problem, which can happen when the ReLU function becomes frozen at zero for negative inputs, resulting in the disappearance of the gradient and slowing down the training process. The LeakyReLU function is defined as: f(x) = x for x > 0 f(x) = ax for x <= 0 where a is a small positive constant, typically set to 0.01 remove back ground color.

In a CNN-based autoencoder application for breast cancer image retrieval, the LeakyReLU activation function is commonly used in the convolutional layers of the encoder and decoder. This helps to introduce non-linearity into the model, allowing it to capture more complex patterns in the data. The LeakyReLU function is also computationally efficient, making it suitable for use in large-scale models with many layers.

The decoder uses the same convolution method configuration as the encoder, which uses 128 filters with a 3 X 3-pixel filter size and the same activation function processed differently. Data from the latent dimension is transposed by the decoder into the reconstructed image. The stage moves on to the error calculation and optimization phase when the encoder and decoder stages are finished. At this point, each iteration of the process involves calculating the loss for each function to see which has the lowest loss value. Mean Square Error serves as the study's loss function (MSE). The MSE measures the average squared difference between the predicted output and the actual target output, and is defined as:

$$MSE = (1/N) * \sum (y - \hat{y})^2$$

Where N = number of samples of dataset
y = actual target output
ŷ = predicated output

The goal of the model is to minimize the MSE loss by adjusting the model parameters through backpropagation and gradient descent. By minimizing the MSE loss, the model is able to generate high-quality reconstructed images that closely resemble the original input images. This is important in breast cancer image retrieval, where accurate and detailed images are necessary for accurate diagnosis and treatment.

Feature Extraction

The encoder part of the autoencoder handles feature extraction. Convolutional layers in the encoder component apply filters to the input image to extract features at different levels of abstraction. Each layer of convolution generates a set of feature maps that indicate whether or not specific visual patterns are present in the image. Following this, these feature maps are subjected to a pooling layer that decreases their spatial dimensions and a non-linear activation function, such as ReLU, that incorporates non-linearity into the feature representation. The final step is to flatten the resulting feature vectors and feed them into a fully connected layer, which converts them into a representation of a lower-dimensional latent space.

Replacing AE with VAE:

VAEs model the latent space as a probabilistic distribution, typically a Gaussian distribution, as opposed to a deterministic latent space in AEs providing a more robust and flexible representation of the data, allowing for better handling of uncertainty and generating more diverse samples. VAEs are capable of generating new samples by sampling from the latent space. This allows for creative exploration of the data distribution and generation of novel data points that may not be present in the original dataset. VAEs employ a regularization term in their loss function, known as the Kullback-Leibler (KL) divergence, which encourages the learned latent space to follow a specific distribution (e.g., Gaussian). This helps in preventing overfitting and encourages the model to learn meaningful representations of the data.

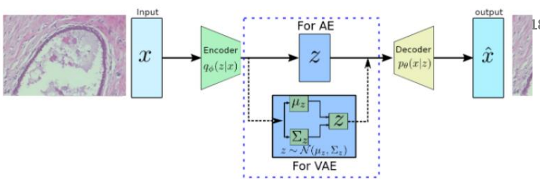

Figure 4: Overview of AE and VAE

Using ImageNet model head as Encoder:

Current SOTA model uses only one layered CNN model as an encoder. This limits the learning capability of the encoder considering the images size used is 256x256. To increase the complexity of encoder we use SOTA ImageNet models such as Swin Transformer, DenseNet, EfficientNet, ResNet as encoders by removing the classification layer. To match the complexity of encoder and decoder we increase the number of layers/stages of Conv + Upsample Layers in decoder till we reach the comparable model complexity. For our task we have replaced 1 stage Conv + Upsample network with 4 stage Conv + Upsample layers. Considering our dataset is entirely different from ImageNet we didn't use any pretrained weights before training.

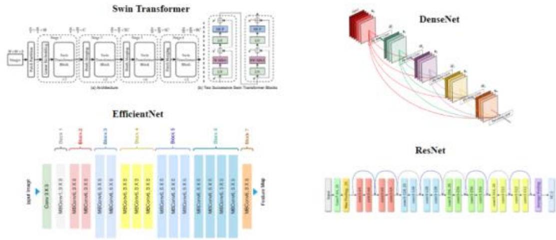

Figure 5: Various ImageNet models

Conditioning Encoder and Decoder with Label:

Conditional Variational Autoencoders (CVAEs) are a variant of Variational Autoencoders (VAEs) that incorporate additional conditional information during training. CVAEs learn a probabilistic latent space that is conditioned on the input conditions, which allows for a more meaningful and structured representation of the data. This can result in better disentanglement of the factors of variation in the data, making CVAEs suitable for tasks such as disentangled representation learning, style transfer, and data manipulation. We have used embedding layer followed by flattening it's output to convert a label to 48D embedding. Later this embedding is concatenated with encoder output. Finally a FC Layer is added on top of match the original AE's encoder's feature vector dimension.

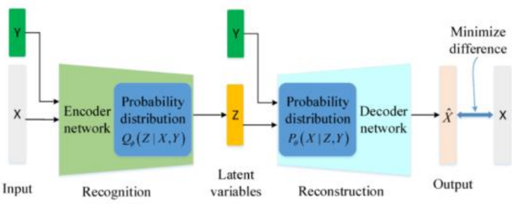

Figure 6: Conditional VAE

Results:

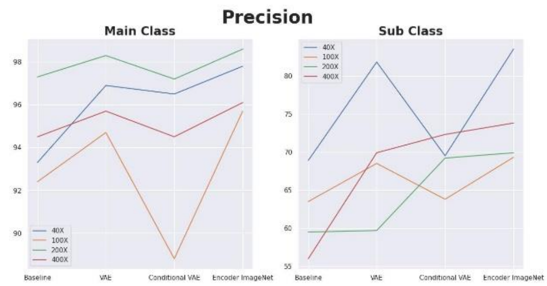

Figure 7: Precision Comparison for all experiments

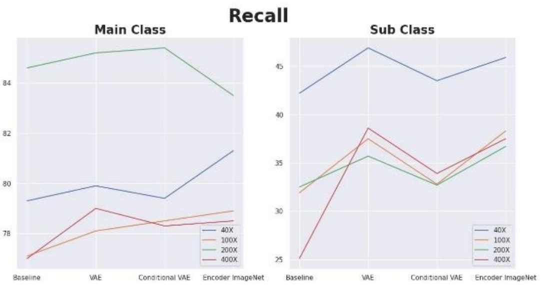

Figure 8: Recall Comparison for all experiments

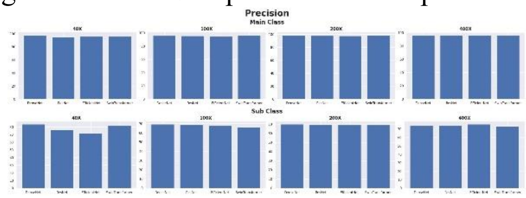

Figure 9: Precision comparison for different ImageNet encoders

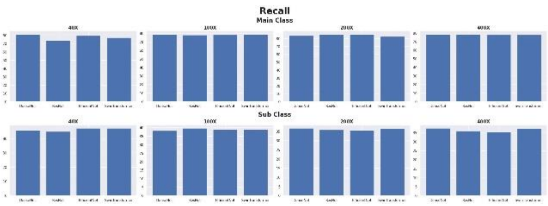

Figure 10: Recall comparison for different ImageNet encoders

Observations:

From Figure 7 and 8 we can clearly see that for both main class and sub class level classifications irrespective of magnifications, ImageNet encoder approach gave best results in-terms of both Precision and Recall. Figure 9, 10 shows precision and recall scores of different ImageNet encoders (Swin Transformer, ResNet, DenseNet, EfficientNet). All the encoders performed almost the same with slight difference in performance. By considering the average Precision and Recall for all the experiments varying class level and magnifications, DenseNet outperformed other ImageNet encoders.

References:

[1] Z. Liu et al., ‘Swin Transformer: Hierarchical Vision Transformer using Shifted Windows’, CoRR, vol. abs/2103.14030, 2021.

[2] G. Huang, Z. Liu, and K. Q. Weinberger, ‘Densely Connected Convolutional Networks’, CoRR, vol. abs/1608.06993, 2016.

[3] K. He, X. Zhang, S. Ren, and J. Sun, ‘Deep Residual Learning for Image Recognition’, CoRR, vol. abs/1512.03385, 2015.

[4] D. P. Kingma and M. Welling, ‘Auto-Encoding Variational Bayes’, arXiv [stat.ML]. 2022.

[5] A. E. Minarno, K. M. Ghufron, T. S. Sabrila, L. Husniah, and F. D. S. Sumadi, ‘CNN Based Autoencoder Application in Breast Cancer Image Retrieval’, in 2021 International Seminar on Intelligent Technology and Its Applications (ISITIA), 2021, pp. 29–34.

[6] L. Wei, Y. Yang, and R. M. Nishikawa, ‘Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis’, Pattern Recognition, vol. 42, no. 6, pp. 1126–1132, 2009.

[7] Y.-L. Huang, D.-R. Chen, and Y.-K. Liu, ‘Breast cancer diagnosis using image retrieval for different ultrasonic systems’, in 2004 International Conference on Image Processing, 2004. ICIP ’04., 2004, vol. 5, pp. 2957-2960 Vol. 5.

[8] R. S. Bressan, P. H. Bugatti, and P. T. M. Saito, ‘Breast cancer diagnosis through active learning in content-based image retrieval’, Neurocomputing, vol. 357, pp. 1–10, 2019.

[9] J. Zhou, C. Feng, X. Liu, and J. Tang, ‘A texture features based medical image retrieval system for breast cancer’, in 2012 7th International Conference on Computing and Convergence Technology (ICCCT), 2012, pp. 1010–1015.

Work Distribution :

Name	Work
D G Pranathi (19CS10025)	• Wrote the necessary code and scripts to carry out the experiments. • Tested and validated the code to ensure accuracy and reliability of results. • Analyzed the data obtained from the experiments to draw meaningful conclusions.
Sonu Kumar Yadav (20CS10061)	• Conducted thorough research on the assigned topic. • Helped Bhagoti in Report making. • Ensured that the report adhered to the guidelines of the ARR format.
Bhagoti (20CS30010)	• Conducted thorough research on the assigned topic. • Gathered information from reliable sources. • Compiled the findings into a well-structured report.