**Laporan Proyek Akhir**

**Computational Biology**

Topik/judul : Breast Cancer Classification Using CNN

Nomor Kelompok : 1

Kelas : LV01

NIM – Nama Lengkap Ketua Kelompok : 2602101514 – Valentcia Angelica

NIM – Nama Lengkap Anggota 1 : 2602145030 – Aulia Yasmin

**Kontribusi dalam project:**

|  |  |  |  |
| --- | --- | --- | --- |
| NIM | Nama Lengkap | Kontribusi | Keterangan |
| 2602101514 | Valentcia Angelica | 1. Melakukan Review Artikel Sejumlah 12 | Artikel yang saya review dapat dilihat pada link :   1. <https://doi.org/10.3390/ASEC2022-13791> 2. <https://doi.org/10.1109/TENCON58879.2023.10322406> 3. <https://doi.org/10.1016/j.engappai.2023.106902> 4. <https://doi.org/10.53730/ijhs.v6nS1.8378> 5. <https://doi.org/10.1109/ACCESS.2023.3265327> 6. <https://doi.org/10.32604/cmes.2022.017030> 7. <https://doi.org/10.11591/ijece.v13i5.pp5885-5897> 8. <https://doi.org/10.1016/j.procs.2023.03.009> 9. <https://doi.org/10.3390/app13095260> 10. <https://doi.org/10.1016/B978-0-12-818366-3.00005-8> 11. <https://doi.org/10.1016/j.neucom.2020.07.061> 12. https://doi.org/10.1016/j.compbiomed.2021.104931 |
| 1. Membuat Desain Workflow | Desain workflow dibuat sesuai dengan alur kerja pada eksperimen ini. |
| 1. Merancang Code Klasifikasi menggunakan Arsitektur EfficientNetB0, VGG16, DenseNet169, ResNet50 | Code ini terinspirasi dari Kaggle link: <https://www.kaggle.com/code/thesnak/breast-cancer-classification-96-89>. Code tersebut diperbaharui sesuai dengan model yang di apply. |
| 1. Melakukan training pada code EfficientNetB0, VGG16, DenseNet169, ResNet50 | Memberharuan code sesuai dengan model yang diusulkan. |
| 1. Menulis bagian Methodology, Result, and Disscussion pada laporan | Menjelaskan bagan kerja sesuai dengan workflow, menganalisis arsitektur model yang dikembangkan, serta menjelaskan hasil dari model. |
| 1. Memperbaiki bagian literature review yang telah dibuat oleh Aulia | Menambahkan beberapa detail terkait literature review yang dibaca. |
| 2602145030 | Aulia Yasmin | 1. Melakukan Review Artikel Sejumlah 11 | Artikel yang saya review dapat dilihat pada link :   * + - 1. <https://doi.org/10.1186/s13058-023-01687-4>       2. <https://doi.org/10.3390/pharmaceutics13050723>.       3. <https://doi.org/10.1177/10732748211038735>       4. <https://doi.org/10.1002/cac2.12207>       5. <https://doi.org/10.3322/caac.21754>       6. <https://doi.org/10.3390/jpm11080808>       7. <https://doi.org/10.1186/s13104-022-05936-1>       8. <https://api.semanticscholar.org/CorpusID:218597317>       9. <https://doi.org/10.3390/info14070410>.       10. https://doi.org/10.1109/ICIT52682.2021.9491631       11. https://doi.org/10.3390/diagnostics13010103 |
| 1. Berkontribusi dalam memperbaiki dan mengaplikasikan beberapa code dari Valentcia Angelica | Membantu ketika ada codingan yang harus diengkapi.  Memperbaiki code EfficientNetB0, ResNet50, VGG16, DenseNet169.  Melakukan training terhadap code.  Membantu dalam pengumpulan data. |
| 1. Membantu dalam proses pengumpulan data untuk dapat digunakan dalam menganalisis model-model untuk mendeteksi kanker payudara. | Membaca artikel dan jurnal ilmiah untuk memperoleh informasi sebanyak mungkin, sehingga dapat dijadikan bahan referensi. |
| 1. Membantu memperbaiki laporan dan merapikan laporan | Sebelum diperbaiki, laporan masih terdapat beberapa bagian yang salah, sehingga memerlukan perbaikan. |
| 1. Membuat bab introduction dan conclusion. | Melakukan review jurnal lalu diimplementasikan ke dalam introduction dan membuat conclusion |

Breast Cancer Classification Using CNN

**Introduction**

Breast cancer is the cause of cancer death for women. Breast cancer has well-established risk factors, and lowering risk is essential to lowering the disease's occurrence. Breast cancer normally develops silently, and is diagnosed during routine screening in the Western world. Without screening, breast cancer is commonly identified as a palpable breast lump. Depending on the stage and kind of tumor, a combination of surgery, radiation, chemotherapy, and immunotherapy may be used to treat breast cancer. Overall survival and patient-reported outcomes have significantly improved as a result of advancements in various therapy approaches (M. Alkabba et al., 2024).

Most American women are diagnosed and die from breast cancer. The American Cancer Society recommends yearly screening mammography for early detection for this cancer. Early detection and treatment give a better result for patient outcomes. MRI is generally more sensitive and offers more detailed pathophysiological information and is also less cost effective compared to mammography for population-based screening (Adam et al., 2023).

Patients with small tumor sizes at the time of diagnosis had a much greater survival rate and a lower risk of death from cancer. As a result, several novel technologies are being developed for the early diagnosis of primary tumors, distant metastases, and recurrent disease in order to manage breast cancer effectively. Theranostics has evolved as a new paradigm for simultaneously diagnosing, imaging, and treating malignancies. It offers the ability to provide more prompt and effective patient care through tailored therapy. Nanotheranostics allows cell-specific targeting moieties, imaging agents, and therapeutic compounds to be integrated into a single formulation for successful therapy (Bhushan et al., 2021).

Such progress has been accelerated in recent decades by technological and conceptual advances in a variety of fields, including massive next-generation sequencing, the incorporation of "omic" sciences, high-resolution microscopy, molecular immunology, flow cytometry, individual cell analysis and sequencing, new cell culture techniques, and the development of animal models, among others. Nonetheless, there are numerous questions and difficulties that need to be addressed in relation to this disease. As a result, oncological research must be regarded as urgent (Piña-Sánchez et al., 2021).

(Adam et al., 2023) shows that one in eight American women (13%) will be diagnosed with breast cancer at some point in their lives, and one in 39 (3%) will die from it. The American Cancer Society recommends that women have yearly screening mammography for early diagnosis of breast cancer, which can begin at age 40. Approximately 2%-5% of women in the general population in the United States have a lifetime risk of breast cancer of 20% or greater, though this figure varies depending on the population studied and the risk assessment method utilized. The American Cancer Society advises yearly breast magnetic resonance imaging (MRI) in addition to mammography for women who have a lifetime risk of 20-25% or higher. Early detection and treatment are likely to produce better patient outcomes.

In 2020, there were an estimated 2.3 million new breast cancer cases and 685,000 breast cancer deaths worldwide. The age-standardized incidence and mortality rates varied by country, Belgium having the highest incidence at 112.3 per 100,000 population and Iran having the lowest at 35.8 per 100,000 population. Fiji had the highest mortality rate at 41.0 per 100,000 population and South Korea had the lowest at 6.4 per 100,000. Some Asian and African countries had a peak age for breast cancer more than ten years sooner than European or American countries. From 2000 to 2012, age-standardized incidence rates of breast cancer grew in China and South Korea but fell in the United States. Between 2000 and 2015, age-standardized death rates climbed dramatically in China and South Korea while decreasing in the UK, USA, and Australia (Lei et al., 2021).

Breast cancer incidence rates have increased by 0.5% per year over the last four decades, primarily due to localized-stage and hormone receptor-positive illnesses. However, mortality rates have dropped since 1989, with a 43% reduction between 1989 and 2020. Black women have the lowest 5-year relative survival of any racial/ethnic group, with the largest absolute gaps for hormone receptor-positive/human epidermal growth factor receptor 2-negative disease, hormone receptor-negative/human epidermal growth factor receptor 2-positive disease, and stage III disease (Giaquinto et al., 2022).

For the latest news, breast cancer accounts for 30% of female malignancies, with 276,480 new cases and more than 42,000 expected deaths in 2020. Based on the expression of the estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), invasive breast cancer can be classified into four major molecular subgroups using immunohistochemistry. Luminal A BC (ER+ and/or PR+, and HER2-) accounts for around 60% of BC and is associated with a good prognosis. Luminal B BC (ER+ and/or PR+ and HER2+) accounts for 30% of all BC and is associated with elevated ki67 (>14%), a proliferation marker, and a poor prognosis. HER2 BC (ER-, PR-, and HER2+) accounts for 10% of all BC cases and is linked with a poor prognosis (Burguin et al., 2021).

The growing demand for breast MRI, along with a radiology scarcity, has resulted in an increase in radiologists' workload. Machine learning methods help radiologists improve the accuracy with which they interpret breast MRI images, as well as support clinical decision-making and patient outcomes. By examining massive datasets of MRIs, machine learning algorithms can learn to identify and classify questionable areas, thereby lowering the incidence of false positives and false negatives and boosting diagnostic accuracy. A few studies have found that machine learning can outperform radiologists in diagnosing breast cancer using MRIs. Machine learning could also assist radiology departments prioritize their worklists (Adam et al., 2023).

Invasive Ductal Carcinoma (IDC) is one of the most subtype of breast cancer that is need to detect early among women. (Barsha et al., 2021) proposed DenseNet-121, DenseNet-201, ResNet-101v2, and ResNet-50 to build an ensemble model for grading the IDC. This study used the private dataset that produced by the Hospital of the University of Pennsylvania and The Cancer Institute of New Jersey, including the IDC positive and non-IDC. This experiment got an overall accuracy of 69.31%, 75.07%, 61.85%, and 60.50% on one validation cohort and 62.44%, 79.14%, 76.62%, and 71.05% on the second validation.

In 2023 (Uysal & Köse, 2022) proposed three models to clasify three classes of breast cancer, using shared BreakHis dataset from Kaggle, including 780 ultrasound images of benign, malignant, and normal. The proposed models are ResNet50, ResNetXt50, and VGG16 with randomly distributred datasets, diveded into 70%, 30% of training and testing, respectively. the highest accuracy achieves by ResNetXt50 with 85.83%, while ResNet50 gets 85.4%, and VGG16 achieves 81.11%.

(Gautam & Singh, 2023) conducted several models, such as DenseNet121, DenseNet169, DenseNet201, EfficientNetB0, EfficientNetB5, EfficientNetV2B0, and EfficientNetV2S. These models used BreakHis dataset, including 2,480 benign class images, and 5,429 of images in malignant class. The experiment was split into train, test, and validation of 70%:20%:10%, respectively. the result accuracy of DenseNet121, DenseNet169, DenseNet201, EfficientNetB0, EfficientNetB5, EfficientNetV2B0, and EfficientNetV2S models are 66.96%, 60.45%, 42.32%, 57.17%, 66.83%, 56.98%, and 58.43%, respectively. overall, the accuracy shows that the models could not performed well. In this context, the experiment needs more training to evaluate a better performance for each model.

In (El Agouri et al., 2022) Resnet50 and Xception models produced equivalent results, with Xception extracted features outperforming the former somewhat. They showed high levels of overall correct classification accuracy (88%), as well as sensitivity (95%), for detecting carcinoma cases, which is critical for diagnostic pathology workflow in order to assist pathologists in accurately identifying breast cancer. Despite the small amount of the data, the results of the current investigation revealed that the built classification model performs well in predicting breast cancer diagnosis.

(Abdel Rahman et al., 2020) employed pre-trained CNN models, InceptionV3 and ResNet50, on the DDSM dataset to discriminate benign and malignant mammography lesions. Due to the low amount of data, transfer learning, pre-processing, and data augmentation approaches were employed. ResNet50 achieved 85.7% accuracy, whilst InceptionV3 obtained 79.6% (Jafari & Karami, 2023). In the field of BC detection, reducing false negatives is critical to ensuring accurate diagnosis and avoiding the possible harm caused by missing positive cases. This study presents a unique CNN-based strategy to improve the accuracy of breast cancer identification, with a specific focus on X-ray picture datasets. The creation of an improved and dependable system for cancer detection has enormous promise for improving patient outcomes and furthering the field of medical imaging diagnostics (Jafari & Karami, 2023).

In (Albashish et al., 2021), the authors explore how to distinguish between benign and malignant cases in the BreaKHis dataset [**16]** using different magnifications. They discuss how the authors before adapted AlexNet by changing the last layer to include two classes. Then they utilized it as both a feature extractor and a classifier. They improved F Measure to 94.6% for binary classification at 40 magnification. However, they disregard the multiclass task, which is a difficult task in this domain. VGG16 was employed for BreakHis categorization at 40X magnification. They scored 89.6% in multiclass classification. One reason for the poor results is that they both used CNN as a classifier model, which may necessitate fine-tuning of its parameters. Thus, providing the features to the classifiers may provide satisfactory results.In consideration of prior investigations, some points are observed: First, utilizing pre-trained CNN as a feature extraction method is preferable to combining extraction and classification in a single model. Second, the multiclass classification problem in BreaKHis is difficult and deserves greater attention from scholars. Third, to lessen the imbalance in the dataset, a data augmentation approach is required.

The aims of this project are to evaluate and compare how CNN architecture works, especially using the ResNet50 model, DenseNet169 model, VGG16 model, and EfficientNetb0 model to classify breast cancer. By using advanced deep learning models we knew how to improve accuracy in detecting and classifying breast cancer from histopathological images. We will determine which model is the most effective for this specific application based on evaluation metrics, such as accuracy, precision, recall, and F1-score. We also examine machine learning algorithms used to improve the accuracy and efficiency of classification models. The benefits that will be gained from this project are acquiring practical knowledge of how to apply and optimize various CNN architectures for image classification, generating models that can provide objective and consistent assessments, and reducing diagnostic variability among pathologists

**Methodology**

A collage of images of cells

Description automatically generated

Fig. 1. Labelled images from IDC + and IDC – class.

The dataset was collected from Kaggle, Breast Histopathology Images, using the Invasive Ductal Carcinoma (IDC) images, which are one of the most common subtypes of all breast cancers. This is a public dataset and can be accessed on Kaggle. Paul Mooney published this dataset.

Table 1. Number of images from each class in the Kaggle dataset.

|  |  |
| --- | --- |
| Classes | Number of Images |
| IDC positive | 78,786 |
| IDC negative | 198,738 |
| **Total** | **277,524** |

The dataset only contains two categories of breast cancer, which are IDC positive, and IDC negative. This dataset includes of 277,524 patches of size 50 x 50 pixels (198,738 IDC negative and 78,786 IDC positive). The example of microscopic images of breast tissue is shown on Fig.1.

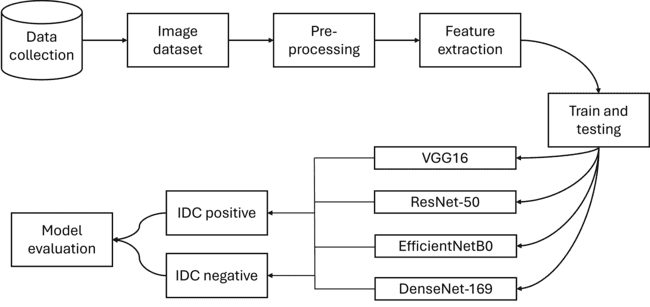


Fig. 2. Proposed workflow.

There are four steps that are used in this study. First step is collecting dataset, which is from Kaggle (BreakHis) dataset. After the dataset was loaded, the dataset went through preprocessing, including filter images, image normalization, combine and shuffling the images. After the preprocessing, all labelled datasets will be processed, including splitting the dataset, training and testing using four models, and lasty, performance evaluations.

The first step is data collection, with all data downloaded from Kaggle. This process took more time because the dataset contains a large number of images. After the dataset successfully loaded, we visualised it to verify that it was correct.

There are some steps in the preprocessing step, the first one is image filtering, that filters images based on their files name. the dataset was explained that images with file names ending with ‘0’ are considered to IDC negative, meanwhile those ending with ‘1’ are IDC positive. Due to a significant difference between the positive and negative IDC case, we limit the dataset for the training and testing with 10,000 of positive images, and 10,000 of negative images.

Then, these images will be resized into 50 x 50 pixels for each using linear interpolation. This step helps each model extract features better.

Next, the combining and shuffling process refer to process that combine all IDC negative and IDC positive and shuffles them to randomness in the dataset. So, the total number of used datasets is 20,000 of images. After that, the dataset will be processed into train and testing. In this study, the percentage of the training data to the testing and validation is 80%:10%:10%, respectively. Next, all the splitting data will be trained using VGG16 model, ResNet50 model, EfficientNetB0 model, and DenNet model. All of these models were using 20 epochs, because it is the number that the result shown more accurately.

A black and white rectangular object with text

Description automatically generated with medium confidence

Fig. 3. VGG16 model architecture

Visual Geometry Group 16 (VGG16) is one of the most common deep convolutional neural network (CNN) which designed for image classification. This architecture was introduced by Simonyan and Zisserman at the Oxford University (Arjun & Kumar, 2022). As figured in Fig. 3. we used all the 13 convolutional layers without top layer, and three top custom layers.

The base model includes 13 convolutional layers, using the weights pre-trained on the ImageNet dataset. It excludes the top layer of VGG16 because we want to add custom layers for a better result.

For the custom layers, the Flatten layer converts the 3D output from the last convolutional layer of the VGG16 model into a 1D feature vector. The first Dense layer used a connected layer with 128 neurons, ReLU activation, and He uniform initialization. The first BatchNormalization layer normalizes the previous layer, helping to stabilize the training and performance. The first Dropout layer applies the dropout regulation, setting 50% of the neurons to zero to prevent overfitting.

Then, the second dense layer adds another fully connected layer with 64 neurons, also using the ReLu activation and He uniform initialization. As above, the second BatchNormalization and Dropout layer normalize and apply the dropout regulation. The last dense layer adds 2 neurons to the fully connected layer, using the sofmax activation to apply the output.

The Rectified Linear Unit (ReLu) is a activation function to eliminate the negative-valued input in the preserving the positive-valued inputs in neral network algorithm. This following equation (1) is the formula to the ReLu activation.

(1)

The Softmax activation is a normalized exponential function that is used in the output layers of classification algoritm (Mehra et al., 2023). This activation can be alternative to the sigmoid activation. The SoftMax activation formula can be seen in the equation (2).

(2)

A diagram of a computer program

Description automatically generated with medium confidence

Fig. 4. EfficientNetB0 architecture

EfficientNetb0 has a total of 237 layers, excluding the top layers. The base model includes 237 layers, contains all the convolutional, batch normalization, rescaling, and all other layers. Because this large number of the base layers, the shown Fig. 4. of efficientNetB0 architecture does not show all the layers. In this study we used a total of 10 custom top layers on the top of EfficientNetB0 model.

Table 2. The number of layer(s) in EfficientNetB0 architecture

|  |  |
| --- | --- |
| **Model** | **Number of Layer(s)** |
| Flatten Layer | 1 Layer |
| First Dense Layer | 1 Layer (128 units) |
| First Batch Normalization Layer | 1 Layer |
| First Dropout Layer | 1 Layer (50%) |
| Second Dense Layer | 1 Layer (64 units) |
| Second Batch Normalization Layer | 1 Layer |
| Third Dense Layer | 1 layer (64 units) |
| Second Dropout Layer | 1 Layer (30%) |
| Third Dense Layer | 1 Layer (24 units) |
| Output Dense layer | 1 Layer (2 units) |

Table 2. shows the custom layers of the EfficientNetB0 model. The Flatten layer converts the 3D tensor output from EfficientNetB0 into a 1D vector. The first Dense layer adds a fully connected layer with 128 neurons, using ReLu activation function and He uniform initializer. Same as above, all the Batch Normalization layer aims to normalize the output of the previous Dense layer. The first Dropout layer was dropping 50% of neurons to zero, prevented overfitting. The second Dense layer using 64 fully connected neurons, while the third Dense layer using 24 unit of neurons. All the Dense layer using ReLu activation function and He uniform initializer. Lastly, the Output Dense layer applies the fully connected layer with 2 neurons, which is each class in binary classiffication uses one neuron.

In the efficientNEtB0 architecture, the Rescaling layer normalizes the input images and processes them to scale the pixel values. The scaling step balances the depth d, width w, and resolution r (Raza et al., 2023). This process helps stabilize and improve the training process. The formula for calculating the scaling process is shown in equation (3).

(3)

A diagram of a computer

Description automatically generated

Fig. 5. ResNet-50 architecture

ResNet-50 is an architecture model that is improved neural network. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun produced this model in 2015 (Bichri et al., 2023)**.** The ResNet-50 model has a total of 50 layer, including 48 conventional layer, an pool layer, and one MaxPool layer. The convolutional layers produced a feature map, a BatchNormalization layer, and a ReLu and SoftMax activation function. The Fig. 5. shows the ResNet-50 architecture model but only shows the beginning and the end of the architecture due to a large number of layers.

Table 3. The number of Layer(s) in ResNet50 architecture

|  |  |
| --- | --- |
| **Model** | **Number of Layer(s)** |
| Flatten Layer | 1 Layer |
| First Dense Layer | 1 Layer (128 units) |
| First Batch Normalization Layer | 1 Layer |
| First Dropout Layer | 1 Layer |
| Second Dense Layer | 1 Layer (64 units) |
| Second Batch Normalization Layer | 1 Layer |
| Second Dropout Layer | 1 Layer |
| Third Dense Layer | 1 Layer (24 units) |
| Output Dense layer | 1 Layer (2 units) |

Table 3. shows the number of custom layers used in this experiment. The Flatten layer converts the 3D output from ResNet50 into a 1D vector. This model used four Dense layers with 128, 64, 24, and 2 neurons for the first, second, third, and output Dense layers. All the Dense layer applies ReLu activation function, and He uniform initializer. The output layer used the SoftMax activation function to output class probabilities. In building the model, we used the Adam optimizer to compile it and combine it with a 0.0001 learning rate. This combination of the Adam optimizer and the learning rate increases the model's convergence accuracy of the model (Zhang et al., 2023). the Adam algorithm function is shown below.

(4)

In the equation above, the recommended value of is 0.999 (Bichri et al., 2023), gi is the number of random batches, and the vt is the time of momentum.

A diagram of a computer

Description automatically generated

Fig. 6. DenseNet169 architecture

DenseNet169 has a total of 169 layers, which includes all types of layers (convolutional, batch normalization, activation, etc.). When loaded without the top layers (include\_top=False). while the number of customized top layers is 10. Table () shows the number of layers in the processed ResNEt-50 model.

Table 4. The number of layer(s) in DenseNet169

|  |  |
| --- | --- |
| **Model** | **Number of Layer(s)** |
| Flatten Layer | 1 Layer |
| First Dense Layer | 1 Layer (128 units) |
| First Batch Normalization Layer | 1 Layer |
| First Dropout Layer | 1 Layer (50%) |
| Second Dense Layer | 1 Layer (64 units) |
| Second Batch Normalization Layer | 1 Layer |
| Third Dense Layer | 64 units |
| Second Dropout Layer | 1 Layer (30%) |
| Fourth Dense Layer | 1 Layer (24 units) |
| Output Dense layer | 1 Layer (2 units) |

As shown is Table 4., the Flatten layer used to convert the multi-dimensional data into a one-dimensional array, which is required before feeding it into the dense layers. In this case, it's flattening the output of the DenseNet169 model. Four fully connected layers for the Dense layers (128, 64, 64, 24) are responsible for learning features from the previous flattened data (CC-BY 4.0, 2024). The BatchNormalization layers help speed up the training process and make the model more stable by reducing internal covariate shifts(Leo, 2024). There are also two Dropout layers, using a specific dropout rate of 50 % and 30% to prevent overfitting. And the Output Dense layer uses 2 neurons, which represent the two classes, and uses the softmax activation function to output probabilities of each class.

In evaluating each model, we used the confusion matrix to display the model’s performance. A confusion matrix is a measurement used to solve classification tasks. Four different parameters used in the confusion matrix, such as True Negative, explain the value of negative examples that are classified accurately. As with TN, the True Positive shows the number of positive classes that are classified correctly. The False Positive represents the number of actual negative units classified as positive. And the False Negative is the number of the actual positive classes that are classified as negative(Kulkarni et al., 2020)

In this study, the model’s performance were measured using four parmeters, including accuracy, precision, recall, and F1-score. Accuracy measures reflection of all the classes made as a result of the test implemented against the actual class. A common performance measurement tool that can be applied in the assessment of this model is shown in the equation (5) below.

Accuracy = (TP+TN) / (TP+TN+FP+FN) (5)

The ratio of accurately predicted positive classes to all predicted positive remarks is known as precision. The equation (6) below can be used to calculate the precision.

Precision = TP / (TP+FP) (6)

Recall defined as the total of actual class positive classes divided by the correctly predicted positive classes. The equation (7) below can be used to calculate the recall.

Recall = TP / (TP+FN) (7)

The measure of F1-score is the average of the precision and the recall weighted by the harmonic mean of both. The equation (8) below can be used to calculate the F1-score.

F1 = (2\*Precision\*Recall) / (Precision + Recall) (8)

**Results and Discussion**

These models used the SkLearn and TensorFlow framework, and coded in Python using Jupyter notebook on Kaggle Kernel. These models were performed on the following configuration: Intel 16 CPU cores, 32 GB RAM, and GPU T4 x2 accelerator. We used some libraries in order to apply the method and function that is built. For instance, Keras library, matplotlib, and sklearn.

Table 5. Hyperparameters in each model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Batch size | Learning rate | Epochs | Optimizer |
| VGG16 | 35 | 0.0001 | 20 | Adam |
| EfficientNetB0 | 35 | 0.0001 | 20 | Adam |
| ResNet-50 | 35 | 0.0001 | 20 | Adam |
| DenseNet-169 | 35 | 0.0001 | 20 | Adam |

As shown in the Table 5. all models using the same hyperparameters for the training process, because we want to achieve consistency across folds. But in some cases, we have to consider different factors to find the best performing parameter for each model (Yang & Shami, 2020).

Table 6. Model performance

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | VGG16 | | | EfficientNetB0 | | | ResNet50 | | | DenseNet169 | | |
|  | Precision | Recall | F1-S | Precision | Recall | F1-Score | Precision | Recall | F1-Score | Precision | Recall | F1-Score |
| IDC - | 81.871 | 84.475 | 83.154 | 80.632 | 84.124 | 82.341 | 79.867 | 86.701 | 83.144 | 79.776 | 90.439 | 84.774 |
| IDC + | 83.801 | 84.106 | 82.431 | 84.413 | 80.971 | 82.656 | 86.378 | 79.417 | 82.752 | 88.305 | 75.897 | 81.633 |
| Accuracy |  |  | 82.800 |  |  | 82.500 |  |  | 82.950 |  |  | 83.350 |

Table 7. Accuracy of each model

|  |  |
| --- | --- |
| Models | Accuracy |
| VGG16 | 82.8% |
| EfficientNetB0 | 82.5% |
| ResNet-50 | 82.95% |
| **DenseNet-169** | **83.35%** |

Table 6. shows that the highest accuracy performed by DenseNet169 model with 83.35% of accuracy. In this model, the IDC – class achieved about 79.776% correcrly predicted as IDC -, while for the class IDC+, the precision is 88.305%, indicated that the model got 88.305% of instantes predicted as class IDC+ were actually class IDC+. In Class 0, recall is 0.90439, meaning that the model correctly identified about 90.439% of all actual Class 0 instances. In Class 1, recall is 0.75897, indicating that the model identified about 75.897% of all actual Class 1 instances.

The other three models did not show a significant difference of accuracy. The EfficientNetB0 acchieves the lowest accurqacy of 82.5%, meanwhile the VGG16 model, and the ResNet50 achieve the accuracy of 82.8%, and 82.95%, respectively. it can be seen in the table, the number of parameters in the model architecture tends to affect the accuracy of the model, models with larger parameters generalize to large, more diverse datasets, thus allowing for richer and more abstract feature representation..

The number of parameters is not always the main influence on accuracy results, there are also several factors that become the background for accuracy results, such as the number of datasets, the quality of the dataset, and whether the model's ability can carry out functions with high complexity or not.

Table 8. Shows the comparison of the number of parameters in each model. It can be seen that the VGG16 model has the fewest total parameters compared to the other three models. This can cause a decrease in model performance, because the more parameters in the model, the more complex patterns the model can train in the data. In this context, the dataset used is quite large, so large parameters are needed to be able to train the model better.

Table 8. Number of parameters in each model

|  |  |  |  |
| --- | --- | --- | --- |
| Models | Trainable params | Non-trainable params | Total |
| VGG16 | 74,434 | 14,715,072 | 14,789,506 |
| EfficientNetB0 | 669,898 | 4,049,955 | 4,719,853 |
| ResNet-50 | 1,063,114 | 23,588,096 | 24,651,210 |
| DenseNet-169 | 227,530 | 12,643,264 | 12,870,794 |

A screenshot of a computer screen

Description automatically generatedA collage of graphs

Description automatically generated

Fig. 7. Model confusion matrix and accuracy plot

From the confusion matrix in Fig. 7, it can be concluded that the four models have high scores in the False Negative (FN) section. This shows that the model makes many prediction errors for the actual positive class, causing a lack of correct classification for the positive class. On the other hand, these models also succeeded in predicting positive and negative classes correctly. This can happen because the amount of training data is quite large so that the model can carry out more in-depth exploration of each class.

Table 9. comparison performance of current study and previous studies.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Author** | **Model(s)** | **Class classification** | **Dataset** | **Accuracy (%)** |
| (Uysal & Köse, 2022) | ResNet50 | Benign/malignant/normal | BreakHis | 85.4 |
| ResNetXt50 | 85.83 |
| VGG16 | 81.11 |
| (Gautam & Singh, 2023) | DenseNet121 | Benign/malignant | BreakHis | 66.96 |
| DenseNet169 | 60.45 |
| DenseNet201 | 42.32 |
| EfficientNetB0 | 57.17 |
| EfficientNetB5 | 66.83 |
| EfficientNetV2B0 | 56.43 |
| EfficientNetV2S | 58.43 |
| (Barsha et al., 2021) | DenseNet-121 | IDC+/IDC- | Private | 62.44 |
| DenseNet-201 | 79.14 |
| ResNet-101v2 | 76.62 |
| ResNet-50 | 71.05 |
| (El Agouri et al., 2022) | Resnet50 | Carcima/non-carcima | Private | 84.5 |
| Xception | 88.0 |
| (Abdel Rahman et al., 2020) | ResNet50 | Benign/malignant | DDSM | 85.7 |
| InceptionV3 | 79.6 |
| Ours | VGG16 | IDC+/IDC- | BreakHis | 82.8% |
| EfficientNetB0 | 82.5% |
| ResNet-50 | 82.95% |
| DenseNet-169 | 83.35% |

Table 9. above shows a comparison between the accuracy results of the model we created and previous research. The efficientNetB0 model in this experiment showed better results than in previous research. This happens because in research (Gautam & Singh, 2023), the EfficientNetB0 model has underfitting, so the accuracy results are not optimal. On the other hand, in this experiment, the EfficientNetB0 model can be said to work quite well, due to the custom layers in this research help the model to explore the data mode accurately.

For the ResNet50 model, the best results were obtained by research (Uysal & Köse, 2022), with accuracy reaching 85.4%. The ResNet50 model run by (Uysal & Köse, 2022) shows good performance in stability and class-based results. As for this previous research, the image size used was 400 x 400 pixels, so that the model could work better in understanding details and information..

The higher accuracy of the VGG16 and DenseNet169 models was obtained in this experiment. This may be due to the inadequate number of datasets in the previous works. Moreover, the complex and divergent of datasets may help the model to explore the detail of the images, obtaining the better accuracy.

Based on the above explanation, the performance of the four models in this experiment can be said good enough to classify cancer types. However, it should be noted, there are still several things that can be improved in the model architecture to produce better accuracy.

**Conclusion**

The results of this study shows what is the difference of Convolutional Neural Network (CNN) architectures and the various effectiveness in classifying breast cancer histopathological images. We found that DenseNet169 achieved the highest accuracy of 83.35%. The other three models also got a good performance, there is no significant different of accuracy between these models.From this study, we know that the potential of deep learning models in aiding early and accurate diagnosis of breast cancer, which really important to improve the result of the patient and survival rates.

Future work can explore some ways to optimize our current experimental setup. Enhancing the dataset through augmentation technique could improve model robustness and generalization. Also using more advanced architecture might give the better performance. Combining advanced preprocessing steps such as stain normalization and more comprehensive feature extraction metods could refine the model. Utilizing ensemble learning techniques and cross-validation could provide more reliable and generalized results, then giving more efficieny for breast cancer detection.

**Reference**

Abdel Rahman, A. S., Belhaouari, S. B., Bouzerdoum, A., Baali, H., Alam, T., & Eldaraa, A. M. (2020). Breast Mass Tumor Classification using Deep Learning. *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, 271–276. https://doi.org/10.1109/ICIoT48696.2020.9089535

Adam, R., Dell’Aquila, K., Hodges, L., Maldjian, T., & Duong, T. Q. (2023). Deep learning applications to breast cancer detection by magnetic resonance imaging: a literature review. In *Breast Cancer Research* (Vol. 25, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13058-023-01687-4

Albashish, D., Al-Sayyed, R., Abdullah, A., Ryalat, M. H., & Ahmad Almansour, N. (2021). Deep CNN Model based on VGG16 for Breast Cancer Classification. *2021 International Conference on Information Technology, ICIT 2021 - Proceedings*, 805–810. https://doi.org/10.1109/ICIT52682.2021.9491631

Arjun, K. P., & Kumar, K. S. (2022). A combined approach of VGG 16 and LSTM transfer learning technique for skin melanoma classification. *International Journal of Health Sciences*, 13504–13516. https://doi.org/10.53730/ijhs.v6ns1.8378

Barsha, N. A., Rahman, A., & Mahdy, M. R. C. (2021). Automated detection and grading of Invasive Ductal Carcinoma breast cancer using ensemble of deep learning models. *Computers in Biology and Medicine*, *139*, 104931. https://doi.org/10.1016/j.compbiomed.2021.104931

Bhushan, A., Gonsalves, A., & Menon, J. U. (2021). Current state of breast cancer diagnosis, treatment, and theranostics. In *Pharmaceutics* (Vol. 13, Issue 5). MDPI. https://doi.org/10.3390/pharmaceutics13050723

Bichri, H., Chergui, A., & Hain, M. (2023). Image Classification with Transfer Learning Using a Custom Dataset: Comparative Study. *Procedia Computer Science*, *220*, 48–54. https://doi.org/10.1016/j.procs.2023.03.009

Burguin, A., Diorio, C., & Durocher, F. (2021). Breast cancer treatments: Updates and new challenges. *Journal of Personalized Medicine*, *11*(8). https://doi.org/10.3390/jpm11080808

CC-BY 4.0. (2024, May). *Build a Convolutional Neural Network*.

El Agouri, H., Azizi, M., El Attar, H., El Khannoussi, M., Ibrahimi, A., Kabbaj, R., Kadiri, H., BekarSabein, S., EchCharif, S., Mounjid, C., & El Khannoussi, B. (2022). Assessment of deep learning algorithms to predict histopathological diagnosis of breast cancer: first Moroccan prospective study on a private dataset. *BMC Research Notes*, *15*(1). https://doi.org/10.1186/s13104-022-05936-1

Gautam, A., & Singh, S. K. (2023). Analysis of Deep Learning Models to Detect Breast Cancer from Histopathology Images. *TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)*, 438–443. https://doi.org/10.1109/TENCON58879.2023.10322406

Giaquinto, A. N., Sung, H., Miller, K. D., Kramer, J. L., Newman, L. A., Minihan, A., Jemal, A., & Siegel, R. L. (2022). Breast Cancer Statistics, 2022. *CA: A Cancer Journal for Clinicians*, *72*(6), 524–541. https://doi.org/10.3322/caac.21754

Jafari, Z., & Karami, E. (2023). Breast Cancer Detection in Mammography Images: A CNN-Based Approach with Feature Selection. *Information (Switzerland)*, *14*(7). https://doi.org/10.3390/info14070410

Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). Foundations of data imbalance and solutions for a data democracy. In *Data Democracy* (pp. 83–106). Elsevier. https://doi.org/10.1016/B978-0-12-818366-3.00005-8

Lei, S., Zheng, R., Zhang, S., Wang, S., Chen, R., Sun, K., Zeng, H., Zhou, J., & Wei, W. (2021). Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020. *Cancer Communications*, *41*(11), 1183–1194. https://doi.org/10.1002/cac2.12207

Leo, C. (2024, May 8). *The Math Behind Batch Normalization Explore Batch Normalization, a cornerstone of neural networks, understand its mathematics, and implement it from scratch.*

M. Alkabba, F., Menon, G., & Ferguson, T. (2024). *StatPearls [Internet]*.

Mehra, S., Raut, G., Purkayastha, R. Das, Vishvakarma, S. K., & Biasizzo, A. (2023). An Empirical Evaluation of Enhanced Performance Softmax Function in Deep Learning. *IEEE Access*, *11*, 34912–34924. https://doi.org/10.1109/ACCESS.2023.3265327

Piña-Sánchez, P., Chávez-González, A., Ruiz-Tachiquín, M., Vadillo, E., Monroy-García, A., Montesinos, J. J., Grajales, R., Gutiérrez de la Barrera, M., & Mayani, H. (2021). Cancer Biology, Epidemiology, and Treatment in the 21st Century: Current Status and Future Challenges From a Biomedical Perspective. In *Cancer Control* (Vol. 28). SAGE Publications Ltd. https://doi.org/10.1177/10732748211038735

Raza, R., Zulfiqar, F., Khan, M. O., Arif, M., Alvi, A., Iftikhar, M. A., & Alam, T. (2023). Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images. *Engineering Applications of Artificial Intelligence*, *126*. https://doi.org/10.1016/j.engappai.2023.106902

Uysal, F., & Köse, M. M. (2022). Classification of Breast Cancer Ultrasound Images with Deep Learning-Based Models †. *Engineering Proceedings*, *31*(1). https://doi.org/10.3390/ASEC2022-13791

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, *415*, 295–316. https://doi.org/10.1016/j.neucom.2020.07.061

Zhang, L., Bian, Y., Jiang, P., & Zhang, F. (2023). A Transfer Residual Neural Network Based on ResNet-50 for Detection of Steel Surface Defects. *Applied Sciences (Switzerland)*, *13*(9). https://doi.org/10.3390/app13095260