Integrating Geometric Algebra for Early Cancer Detection and Personalized Treatment Strategies in Indonesia

Grace Evelyn Simon - 13523087[1,2]
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
[1]13523087@std.stei.itb.ac.id, [2]graceevelynsimon@gmail.com

***Abstract*— Cancer remains one of the leading causes of death in Indonesia, with late diagnosis and limited access to advanced treatment options contributing to high mortality rates. This paper introduces the integration of geometric algebra into early cancer detection and personalized treatment strategies as a transformative approach to improve healthcare outcomes. Combined with machine learning, this method enables precise tumor detection from medical imaging data, such as CT and MRI scans, at early stages. The study also explores the application of Singular Value Decomposition (SVD) for dimensionality reduction and Principal Component Analysis (PCA) for feature selection, enhancing computational efficiency while preserving critical tumor features. These techniques facilitate the accurate classification of tumors and the development of treatment strategies tailored to individual patients' genetic and clinical profiles. This integrated approach addresses Indonesia's pressing healthcare challenges, aspiring to reduce cancer-related mortality and improve overall health outcomes.**

***Keywords*— cancer detection, geometric algebra, machine learning, PCA, SVD.**

I. INTRODUCTION

Health is one of the most critical factors in ensuring the quality of life. It is a cornerstone of societal development and a fundamental element of the Sustainable Development Goals (SDGs), particularly SDG 3: ensuring healthy lives and promoting well-being for all at all ages. Achieving global health goals faces significant challenges, with progress slowing since 2015 in areas like maternal mortality, premature deaths from major noncommunicable diseases, and access to essential healthcare. To meet the SDG 3 targets by 2030, substantial investment and focus are needed to address these challenges, including tackling inequality and environmental factors. In Indonesia, these global health challenges are mirrored by the growing burden of diseases, including cardiovascular conditions, respiratory illnesses, and particularly cancer. Among these, cancer has emerged as one of the leading causes of death, with thousands of lives lost annually due to late diagnoses and limited access to advanced treatment options. This alarming trend underscores the urgent need for innovative technologies to improve early cancer detection and enhance treatment strategies.

Early detection plays a crucial role in increasing survival rates for cancer patients. Tumors identified at early stages are often more treatable, allowing for less invasive and more effective interventions. However, in Indonesia, resource constraints and limited access to cutting-edge medical technologies hinder timely diagnoses, creating an urgent demand for technological interventions that can bridge this gap.

One promising approach lies in the application of geometric algebra, a mathematic framework that excels at representing and analyzing complex multidimensional data. By utilizing geometric algebra, it is possible to process and interpret medical imaging data, such as CT and MRI scans, with high precision to detect early-stages tumors. This capability makes it possible to detect subtle abnormalities indicative of early-stage tumors. Combined with machine learning, geometric algebra can further enhance the accuracy of cancer detection, enabling the classification of tumor types and the development of personalized treatment strategies tailored to individual patients based on their genetic and clinical profiles.

This paper explores the integration of geometric algebra into early cancer detection and personalized treatment in Indonesia, aiming to address the critical gaps in Indonesia's healthcare infrastructure. By integrating advanced computational method, this approach aspires to reduce the mortality rate caused by cancer and improve overall health outcomes. Through early diagnosis and individualized care, this application offers a transformative pathway toward enhancing Indonesia's ability to combat cancer effectively.

II. FUNDAMENTAL THEOREM

A. *Singular Value Decomposition*

Singular Value Decomposition (SVD) is a mathematical method used to factorize matrix A into three components:

$$A = U\sum V^T$$

in which U, $\sum$, and V have sizes $m\ x\ m$, $m\ x\ n$, and $n\ x\ n$ respectively, and in which

(a) $V = [\ v_1\ v_2\ ...\ v_n\ ]$ orthogonally diagonalizes A^TA

(b) The nonzero diagonal entries of $\sum$ are $\sigma_1 = \sqrt{\lambda_1}$, $\sigma_2 = \sqrt{\lambda_2}, \ldots, \sigma_k = \sqrt{\lambda_k}$, where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the nonzero eigenvalues of $A^T A$ corresponding to the column vectors of V

(c) The column vectors of V are ordered so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k \geq 0$

(d) $u_i = \frac{Av_i}{||Av_i||} = \frac{1}{\sigma_i} Av_i \qquad (i = 1, 2, \ldots, k)$

(e) $\{u_1, u_2, \ldots, u_k\}$ is an orthonormal basis for col(A)

(f) $\{u_1, u_2, \ldots, u_k, u_{k+1}, \ldots, u_m\}$ is an extension of $\{u_1, u_2, \ldots, u_k\}$ to an orthonormal basis for R^m

SVD generalizes the concept of eigenvalues and eigenvectors, particularly for non-square matrices. It allows us to decompose A into orthogonal bases for its row space and column space, along with a scaling component represented by the singular values.

B. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical and linear algebraic method used to reduce the dimensionality of data while preserving as much variance as possible. PCA transforms a dataset into a new coordinate system, where the axes correspond to the directions of maximum variance. PCA use the concept of SVD, with key steps such as:

(a) Compressing dataset into a single row matrix with equation $X = [x_1, x_2, \ldots, x_n]$

(b) Mean-centering data with formula $\mu_j = \frac{1}{m} \sum_{i=1}^{m} X_{ij}$, and then subtracted with $X_{centered} = X - \mu$

(c) Compute covariance matrix C, $C = \frac{1}{m-1} X_{centered}^T X_{centered}$

(d) Eigenvalue decomposition on the covariance matrix C with eigenvalue equation $Cv_i = \lambda_i v_i$

(e) Eigenvalue sort and select, selection is based on the top k eigenvectors to retain the most important components

(f) Project data onto principal components with $X' = X_{centered} V_k$ with X' as the reduced m x k matrix

(g) Reconstruction if needed using $X_{reconstructed} = X' V_k^T + \mu$

C. Geometric Algebra (GA)

Geometric Algebra (GA) is used for analyzing and manipulating multidimensional geometric structures. It generalizes traditional linear algebra concepts (vectors, matrices) to higher dimensions and enables operations such as rotations, projections, and transformations. Geometric Algebra is defined over a vector space R^n with a set of basis vectors $\{e_1, e_2, \ldots, e_n\}$ that satisfy:

$$e_i \cdot e_j = \begin{cases} 1, if\ i = j \\ 0, if\ i \neq j \end{cases}$$

(a) Geometric product

The geometric product of two vectors a and b is $ab = a \cdot b + a \wedge b$ where $a \cdot b$ is inner product (scalar) and $a \wedge b$ is outer product (bivector). Some properties of geometric product:

- Non-commutativity: $ab \neq ba$
- Distributive: $a(b + c) = ab + ac$

(b) Outer product

The outer product $a \wedge b$ is antisymmetric and represents the oriented area formed by vectors a and b. Meanwhile for three vectors a, b, c, the outer product extends to volumes $(a \wedge b) \wedge c$. Some propertiesof outer product:

- Antisymmetry: $a \wedge b = -b \wedge a$
- Associative: $(a \wedge b) \wedge c = a \wedge (b \wedge c)$

(c) Rotations with rotors

A rotor is a multivector that encodes a rotation in R^n. Rotating a vector v about a plane is given by: $v' = RvR^{-1}$ where $R = e^{B\theta}$ is the rotor, B is bivector representing the plane of rotation, θ is rotation angle.

(d) Projections

Projection of vector b onto vector a is $proj_a(b) = \frac{a \cdot b}{a \cdot a} a$

(e) Reconstructing complex geometries

Higher-grade multivector can be reconstructed with $T = \lambda_1(e_1 \wedge e_2) + \lambda_2(e_2 \wedge e_3) + \lambda_3(e_3 \wedge e_1)$, where λ_i are scalars representing magnitudes along each oriented area

D. Image Segmentation

Image segmentation is the process of dividing an image into meaningful regions using relevant theorems and equations:

(a) Gaussian smoothing

Gaussian kernel $G(x, y)$ is used to reduce noise using equation $G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$

(b) Edge detection

Edge detection is used in the image to highlight boundaries between regions. Gradient of the image is computed using $\nabla I = [\frac{\partial I}{\partial x} \ \frac{\partial I}{\partial y}]$ and the equation for magnitude of gradient (edge detection) is $|\nabla I| = \sqrt{(\frac{\partial I}{\partial x})^2 + (\frac{\partial I}{\partial y})^2}$

(c) Thresholding

Thresholding segments the image by dividing it into regions based on pixel intensity based on equation $T(I_{ij}) = \begin{cases} 1, if\ I_{ij} > \tau, \\ 0, if\ I_{ij} \leq \tau, \end{cases}$ where τ is the threshold value

(d) Segmentation

Segment the image into connected regions using K-Means clustering methods. The image is divided into k clusters by minimizing intra-clusters variance

using $min \sum_{i=1}^{k} \sum_{x \epsilon C_i} ||x - \mu_i||^2$, where C_i is the i-th cluster, and μ_i is its centroid

(e) Normalized cut (graph-based segmentation)
To minimize a cost function, partition the graph with normalized cut $NCut(A,B) = \frac{Cut(A,B)}{Assoc(A,V)} + \frac{Cut(A,B)}{Assoc(B,V)}$ where A and B are partitions, and $Cut(A,B)$ measures edge weights between A and B

E. Machine Learning

Machine learning for classification is the process of assigning data points to predefined categories based on their features. The dataset is represented as X, where X is an $m\ x\ n$ matrix, with m as number of samples and n as number of features.

(a) Logistic regression
Logistic regression models the probability that a sample belongs to a class using the sigmoid function $P(y=1|x) = \frac{1}{1+e^{-(w.x+b)}}$ where w is weight vector and b is bias term.

(b) SVM decision boundary
SVM finds the hyperplane that maximizes the margin between classes. $\max \frac{1}{||w||}$ subject to $y_i(w.x_i + b) \geq 1$

(c) Decision trees
Split the dataset into subsets based on feature thresholds to maximize information gain: $information\ gain = H(y) - \sum_i \frac{|S_i|}{|S|} H(S_i)$, where $H(S)$ is the entropy of set S

(d) Neural networks
Neural networks learn complex patterns through layers of neurons. Each neuron computes $z = w.x + b, a = \sigma(z)$, where σ is the activation function

(e) Loss function
Train the model by minimizing a loss function using optimization techniques $L = -\frac{1}{m}\sum_{i=1}^{m}[y_i \log(\widehat{y_i}) + (1-\hat{y}_i)\log(1-\hat{y}_i)$

III. PROBLEM ANALYSIS

Cancer is one of the leading causes of death in Indonesia. The high mortality rate from cancer is largely due to late diagnoses and limited access to advanced early detection technologies. Early diagnosis plays a crucial role in improving patients' chances of recovery, as tumors detected at an early stage are often more treatable. However, there are several key challenges in detecting cancer in Indonesia:

(a) Limitations of Conventional Technology:
Imaging technologies such as CT and MRI often produce large and complex datasets, making manual analysis difficult and there is no standardized approach to detecting and classifying tumors with high accuracy

(b) Limited Adoption of AI and Mathematical Methods:
The use of machine learning-based technologies, such as Geometric Algebra (GA), SVD/PCA, and image segmentation, remains limited in Indonesia

(c) Processing Medical Imaging Data
CT/MRI scans generate high-dimensional data with significant noise and manual analysis is challenging due to the complexity of the data

(d) Tumor Classification and Geometric Representation of Tumors:
After detecting a tumor, classifying the type requires highly accurate algorithms. Precisely determining the shape and orientation of tumors is difficult with traditional methods.

Based on the problem analysis, the proposed approach includes:

1. Integration of Geometric Algebra (GA):
A mathematical representation to understand the shape, orientation, and projection of tumors.
2. Utilization of SVD/PCA:
Dimensionality reduction to handle large-scale imaging data.
3. Application of Machine Learning:
Improving accuracy in detecting and classifying tumors.
4. Mathematical Image Segmentation:
Precise extraction of tumor regions to ensure better diagnosis.

This approach aims not only to improve the accuracy of tumor detection but also to aid in developing personalized treatment strategies tailored to the unique conditions of each patient.

IV. IMPLEMENTATION

The implementation of the system for early cancer detection and classification is achieved using the Python programming language. The program consists of several functions to achieve the objectives defined in the study. Each function is designed to fulfill a specific task within the system, following the theoretical foundation provided in Chapter II. This implementation uses Python version 3.10 and employs the libraries NumPy, OpenCV, and scikit-learn to handle matrix operations, image processing, and machine learning, respectively. Each section below provides an explanation of the function, its purpose, and its implementation in Python.

(a) Preprocessing
This step normalizes the input CT/MRI image data and prepares it for analysis. The goal is to ensure that pixel intensity values are scaled to a consistent range for uniformity during processing.

```python
import cv2
import numpy as np

def normalize_image(image_path):
    image = cv2.imread(image_path, cv2.IMREAD_GRAYSCALE)
    normalized = (image - np.min(image)) / (np.max(image) - np.min(image))
    return normalized
```

Fig. 4.1 Implementation of Preprocessing CT/MRI Image Data
Source: Author

(b) Singular Value Decomposition (SVD)

SVD is used to reduce the dimensionality of the image while preserving critical features, as described in Chapter II.

```
def svd_reduction(image, k):
    U, S, Vt = np.linalg.svd(image, full_matrices=False)
    reconstructed = np.dot(U[:, :k], np.dot(np.diag(S[:k]), Vt[:k, :]))
    return reconstructed
```

Fig. 4.2 Implementation of SVD
Source: Author

(c) Thresholding for Tumor Segmentation

Thresholding is applied to separate the tumor region from the background. A binary threshold is used to isolate high-intensity regions representing potential tumors.

```
def threshold_image(image, threshold=0.5):
    _, thresholded = cv2.threshold(image, threshold, 1.0, cv2.THRESH_BINARY)
    return thresholded
```

Fig. 4.3 Implementation of Thresholding Image
Source: Author

(d) Geometric Feature Extraction

Geometric Algebra principles are applied to extract tumor features such as area and centroid. Area calculation is used to compute the total number of pixels in the tumor region, meanwhile the centroid calculation is used to determine the geometric center of the tumor region.

```
def calculate_area(segmented_image):
    return segmented_image.sum()

def calculate_centroid(segmented_image):
    coords = np.argwhere(segmented_image == 1)
    centroid = np.mean(coords, axis=0)
    return centroid
```

Fig. 4.4 Implementation of Area Calculation and Centroid Calculation
Source: Author

(e) Machine Learning-Based Classification

The classification step predicts the tumor type based on extracted features. Its purpose is to train a machine learning model using labeled data to classify tumors based on area, compactness, and other features.

```
def train_classifier(features, labels):
    clf = LogisticRegression()
    clf.fit(features, labels)
    return clf

def predict_tumor(clf, new_features):
    prediction = clf.predict(new_features)
    return "Malignant" if prediction[0] == 1 else "Benign"
```

Fig. 4.5 Implementation of Machine Learning-Based Classification Tumor
Source: Author

The code then integrated for use with a Kaggle dataset with dataset consists of CT/MRI scans images and labels for training. Diagnosis B indicates benign tumor type. A benign tumor is a non-cancerous growth in the body. Benign tumors grow slowly and remain localized (they do not spread to other parts of the body). Examples of benign tumors are lipoma (fat tissue growth), fibroid (non-cancerous growth in the uterus), and adenoma (non-cancerous growth in glandular tissue). Diagnosis M indicates malignant tumor type. A malignant tumor is a cancerous growth that can invade nearby tissues and spread to other parts of the body. Malignant tumors grow quickly and aggressively, they invade and destroy nearby healthy tissues. Malignant tumors are often life-threatening and require immediate treatment. Examples of malignant tumors are carcinoma (cancer of skin or lining organs), sarcoma (cancer of connective tissues), and leukemia (cancer of blood-forming tissues). Identifying whether a tumor is benign or malignant is crucial for determining the appropriate treatment and prognosis.

The dataset used for this paper is “Breast Cancer Data Set” contains the characteristics of patients diagnosed with cancer. The dataset contains a unique ID for each patient, the type of cancer (diagnosis), the visual characteristics of the cancer and the average values of these characteristics.

Fig. 4.6 Breast Cancer Data Set
Source: https://www.kaggle.com/datasets/erdemtaha/cancer-data

V. TESTING AND ANALYSIS

The program could be executed using Kaggle. To get an insight into the dataset, the code represents the head of the dataset, specifically a breast cancer dataset, which is used for binary classification to predict whether a tumor is Malignant (M) or Benign (B).

```
Dataset Head:
         id diagnosis  radius_mean  texture_mean  perimeter_mean  area_mean  \
0    842302         M        17.99         10.38          122.80     1001.0
1    842517         M        20.57         17.77          132.90     1326.0
2  84300903         M        19.69         21.25          130.00     1203.0
3  84348301         M        11.42         20.38           77.58      386.1
4  84358402         M        20.29         14.34          135.10     1297.0

   smoothness_mean  compactness_mean  concavity_mean  concave points_mean  \
0          0.11840           0.27760          0.3001              0.14710
1          0.08474           0.07864          0.0869              0.07017
2          0.10960           0.15990          0.1974              0.12790
3          0.14250           0.28390          0.2414              0.10520
4          0.10030           0.13280          0.1980              0.10430

   ...  texture_worst  perimeter_worst  area_worst  smoothness_worst  \
0  ...          17.33           184.60      2019.0            0.1622
1  ...          23.41           158.80      1956.0            0.1238
2  ...          25.53           152.50      1709.0            0.1444
3  ...          26.50            98.87       567.7            0.2098
4  ...          16.67           152.20      1575.0            0.1374

   compactness_worst  concavity_worst  concave points_worst  symmetry_worst  \
0             0.6656           0.7119                0.2654          0.4601
1             0.1866           0.2416                0.1860          0.2750
2             0.4245           0.4504                0.2430          0.3613
3             0.8663           0.6869                0.2575          0.6638
4             0.2050           0.4000                0.1625          0.2364

   fractal_dimension_worst  Unnamed: 32
0                  0.11890          NaN
1                  0.08902          NaN
2                  0.08758          NaN
3                  0.17300          NaN
4                  0.07678          NaN

[5 rows x 33 columns]
```

Fig. 5.1 Head of The Dataset
Source: Author

The id is a unique identifier for each patient or tumor sample, diagnosis indicating the type of tumor: M: Malignant (cancerous), B: Benign (non-cancerous). Then,

radius_mean, texture_mean, perimeter_mean, area_mean are statistical features calculated from the cell nuclei in the tumor.

- radius_mean: Average radius of the nuclei
- texture_mean: Measure of variation in the texture
- perimeter_mean: Average perimeter of the nuclei
- area_mean: Average area of the nuclei

Meanwhile, smoothness_mean, compactness_mean, concavity_mean, concave points_mean describe the shape and smoothness of the nuclei.

- smoothness_mean: Regularity of the nuclei's boundary
- compactness_mean: Compactness of the nuclei
- concavity_mean: Severity of concave parts of the nuclei boundary
- concave points_mean: Number of concave portions on the boundary

The dataset consists of features derived from the tumor's cell nuclei properties, such as size, shape, texture, and boundary regularity, which are used to classify the tumor as Malignant (M) or Benign (B). The table shows the first 5 rows of the dataset, highlighting the range of feature values for both malignant and benign tumors.

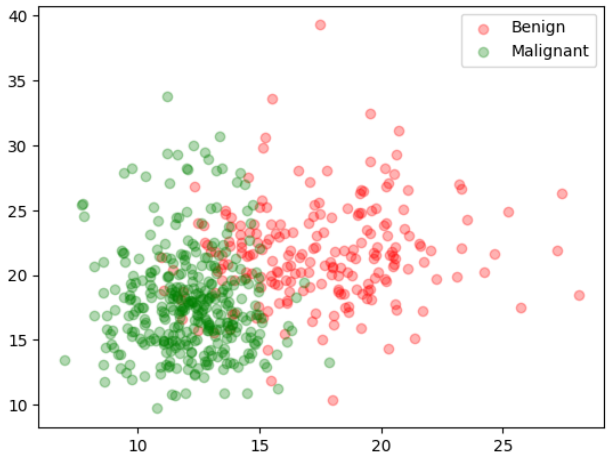

Fig. 5.2 Dataset Distribution
Source: Author

This scatter plot visualizes the distribution of Benign and Malignant tumor samples based on two features. Red points represent Benign tumors, green points represent Malignant tumors. There is a noticeable clustering of red points (benign) in one region and green points (malignant) in another. Malignant tumors typically have higher values along the axes (larger size or irregular features).

The code first tested using an image of malignant type of breast tumor. This section summarizes the model's performance on the test dataset (114 samples split into 71 benign and 43 malignant cases). The model is tested on a new image (test3.png).

1. Extracted Features:
   - Area: 2460
   - Mean: 0.15
   - Prediction Probabilities:
     a. Benign (0): 0.00 (0% confidence in benign).
     b. Malignant (1): 1.00 (100% confidence in malignant).
2. Final Prediction:
   The model classifies the tumor as Malignant.

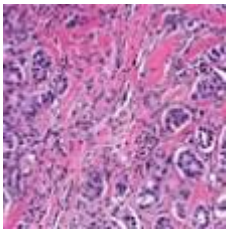

Fig. 5.3 Malignant Samples
Source: https://medium.com/analytics-vidhya/breast-cancer-diagnostic-dataset-eda-fa0de80f15bd

```
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.90      0.93        71
           1       0.85      0.93      0.89        43

    accuracy                           0.91       114
   macro avg       0.90      0.92      0.91       114
weighted avg       0.92      0.91      0.91       114

Prediction for new tumor in /kaggle/input/data-test/data-test1/test3.png: Malignant
```

Fig. 5.4 Malignant Testing
Source: Author

The code then tested using an image of benign type of breast tumor. This section summarizes the model's performance on the test dataset (114 samples split into 71 benign and 43 malignant cases). The model is tested on a new image (test1.png).

1. Extracted Features:
   - Area: 274.5
   - Mean: 0.08
   - Prediction Probabilities:
     a. Benign (0): 1.00 (100% confidence in benign).
     b. Malignant (1): 0.00 (0% confidence in malignant).
2. Final Prediction:
   The model classifies the tumor as Benign.

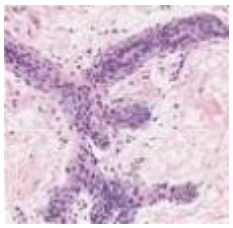

Fig. 5.5 Benign Samples
Source: https://medium.com/analytics-vidhya/breast-cancer-diagnostic-dataset-eda-fa0de80f15bd

```
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.90      0.93        71
           1       0.85      0.93      0.89        43

    accuracy                           0.91       114
   macro avg       0.90      0.92      0.91       114
weighted avg       0.92      0.91      0.91       114

Prediction for new tumor in /kaggle/input/data-test/data-test1/test3.png: Malignant
```

Fig. 5.6 Benign Testing
Source: Author

The application of Singular Value Decomposition (SVD), Principal Component Analysis (PCA), and concepts from Geometric Algebra significantly contribute to the process of tumor classification in this task. Singular Value Decomposition (SVD) is primarily utilized to reduce the dimensionality of the input image data. Medical images, such as the ones used in this study, are essentially matrices where each pixel's intensity represents a data point. SVD decomposes these matrices into three components. By retaining the top singular values, which capture the most critical aspects of the image, SVD helps reduce the dimensionality of the data. This process not only eliminates noise but also retains the structural integrity of the image, which is crucial for subsequent steps like segmentation and feature extraction. The reduced representation is then used to segment the tumor and calculate meaningful features like area and mean intensity.

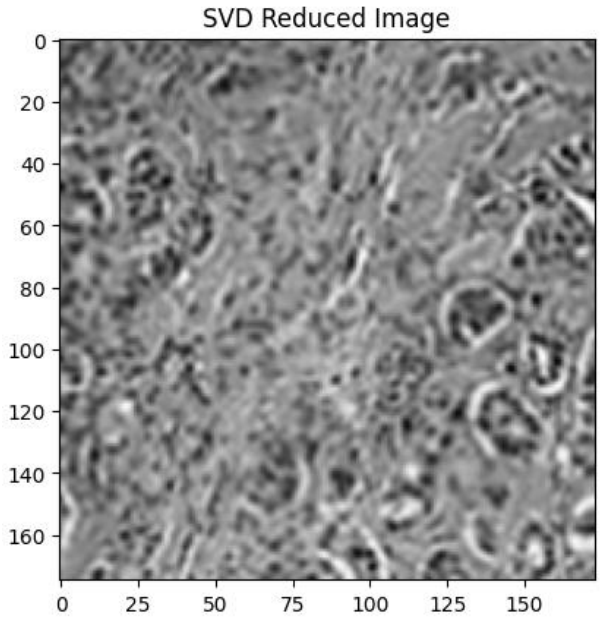

Fig. 5.7 SVD Reduced Image of test3.png
Source: Author

Similarly, Principal Component Analysis (PCA) plays a vital role in feature selection and extraction from the dataset. PCA identifies directions in the feature space, known as principal components, that capture the maximum variance in the data. In this context, it could be used to select the most informative features from the dataset, such as area_mean and smoothness_mean, while reducing the dimensionality of the feature space. By focusing on the components that contribute the most to variance, PCA ensures that the model works with the most relevant features, improving computational efficiency and reducing the risk of overfitting. This approach is particularly beneficial when dealing with high-dimensional datasets, as it prioritizes features that are critical for classification while discarding redundant or less informative ones. With the help of Geometric Algebra, it could provide additional advantages in image processing tasks, particularly in tumor segmentation and feature extraction. Geometric Algebra offers a powerful mathematical framework for representing shapes and spatial relationships. It could enhance segmentation accuracy by mathematically capturing the irregularities in tumor boundaries, thereby allowing for more precise delineation of tumor regions.

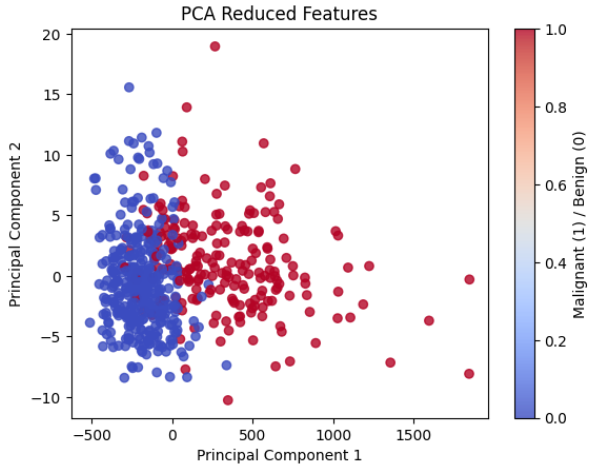

Fig. 5.8 PCA Reduced Features of The Dataset
Source: Author

The model's output demonstrates robust performance, achieving an overall accuracy of 91% and strong metrics across both benign and malignant classes. The classification report shows high precision and recall for benign tumors, with slightly lower precision for malignant tumors. This discrepancy indicates that the model might occasionally classify benign tumors as malignant, leading to some false positives. Despite this limitation, the model is effective at distinguishing between the two classes.

VI. Conclusion

The integration of geometric algebra, Singular Value Decomposition (SVD), and Principal Component Analysis (PCA) in early cancer detection and personalized treatment is an innovation for addressing critical healthcare challenges in Indonesia. By utilizing advanced computational methods, this approach enhances the precision of tumor detection and classification from medical imaging data, enabling timely diagnosis and effective treatment strategies. These methods not only reduce computational complexity but also ensure that critical features of the tumor are preserved for accurate analysis.

The study highlights the transformative potential of mathematical frameworks and machine learning in bridging technological gaps in healthcare infrastructure. In a country like Indonesia, where late-stage diagnoses and

limited access to advanced technologies are widespread, this approach provides a scalable and efficient solution to improving cancer outcomes.

Future research can expand on this work by exploring the integration of additional geometric or computational techniques to enhance accuracy further. Additionally, implementing this approach in real-world clinical settings, with larger and more diverse datasets, will be crucial for validating its effectiveness and adaptability. The application of such innovations not only addresses immediate healthcare needs but also sets the foundation for broader advancements in personalized medicine and data-driven healthcare systems in Indonesia.

VII. ACKNOWLEDGMENT

The completion of this paper would not have been possible without the guidance of Ir. Rila Mandala, M.Sc., Ph.D., whose dedication to teaching and insightful guidance throughout the Linear Algebra and Geometry course have been invaluable. The writer extends gratitude for providing the necessary knowledge during the development of this paper. A debt of gratitude is also owed to Dr. Ir. Rinaldi Munir, whose extensive learning resources and contributions to the field have greatly enriched the writer's understanding and facilitated the completion of this work.

REFERENCES

[1] G. Strang, “Introduction to Linear Algebra”, 5th ed. Wellesley, MA: Wellesley-Cambridge Press, 2016, pp. 145–200.
[2] I. Jolliffe, “Principal Component Analysis”, 2nd ed. New York: Springer, 2002, pp. 35–78.
[3] D. C. Lay, S. R. Lay, and J. J. McDonald, “Linear Algebra and Its Applications”, 5th ed. Boston, MA: Pearson, 2015, ch. 7.
[4] R. Munir, “Aljabar Geometri Bagian 1-4,” [Online]. Available: https://informatika.stei.itb.ac.id/~rinaldi.munir/AljabarGeometri/. Accessed: Dec. 28, 2024.
[5] S. Shankar, “Predicting Breast Cancer with Logistic Regression,” [Online]. Available: https://www.kaggle.com/code/shankarat/predicting-breast-cancer-logistic-regression. Accessed: Dec. 31, 2024.
[6] P. Anderson, “Geometric Algebra for Physicists”. Oxford, UK: Oxford University Press, 2007, ch. 3.

STATEMENT

Hereby, I declare that this paper I have written is my own work, not a reproduction or translation of someone else's paper, and not plagiarized.

Bandung, 2 Januari 2025

Grace Evelyn Simon, 13523087