



Application of Artificial Intelligence Methods and Evaluation Models for Breast Cancer Diagnosis and Prognosis

Ilham Huseyinov¹, Adejumo Dolapo Samuel²

¹Software Engineering Department, Istanbul Aydin University, Istanbul, Turkey

²Graduate Institute of Science, Computer Engineering Program, Istanbul Aydin University, Istanbul, Turkey

ARTICLE INFO	ABSTRACT
Published Online: 12 December 2023	Cancer is a primary global health concern, and accurate diagnosis and prognosis are crucial in patient care. This paper applies the deep learning algorithm CNN to detect and prognosis breast cancer. The results of the conducted experiment are assessed through rigorous evaluation frameworks, ensuring clinical relevance and applicability. The performance of the developed CNN model enhanced the accuracy, efficiency, and reliability of the breast cancer diagnosis and prognosis.
Corresponding Author: Adejumo Dolapo Samuel	

1. INTRODUCTION

This research paper aims to provide a thorough overview of the diverse AI methods and evaluation models developed and applied to cancer diagnosis and prognosis. Using machine learning, deep learning, natural language processing, and data mining, researchers have extracted meaningful information from various sources, including medical images, electronic health records, genomic data, and clinical notes. These AI-driven insights contribute to more accurate early-stage detection, precise tumor characterization, and individualized prognosis predictions, ultimately leading to more informed clinical decisions. In addition to surveying AI methodologies, this paper also emphasizes the need for rigorous evaluation protocols to assess the performance and reliability of AI-driven cancer diagnosis and prognosis systems [1]. As these systems move from experimental settings to clinical practice, it becomes essential to establish evaluation frameworks that consider factors such as sensitivity, specificity, interpretability, generalizability, and clinical utility. Breast cancer is a primary global health concern, and early and accurate diagnosis and prognosis are crucial for patient outcomes and survival rates. Convolutional Neural Networks (CNNs) have revolutionized medical image analysis, proving to be effective in extracting intricate patterns and features from complex images, making them ideal for analyzing medical images such as mammograms, MRIs, and histopathological slides.

2. RELATED WORKS

2.1 Machine Learning Algorithms:

Machine learning algorithms are helpful in cancer research, assisting medical professionals in identifying cancerous cells, classifying tumor types, and predicting patient outcomes. Popular ML algorithms include Support Vector Machines (SVMs), Random Forests, and Naïve Bayes. SVMs, for example, have successfully classified tissue samples into different cancer types based on gene expression profiles [2].

2.1.1 Support Vector Machines (SVMs): Support Vector Machines (SVMs) are frequently used for cancer classification because they can handle high-dimensional data and nonlinear decision boundaries. They differentiate between benign and malignant tumors using medical imaging and genomic data. [3]

2.1.2 Random Forest: Random forest is a machine-learning method that combines multiple decision trees to improve the classification. It helps handle noisy or missing data and is often used for cancer diagnosis [4].

2.1.3 Convolutional Neural Networks (CNNs): Medical imaging has been revolutionized by CNNs, as they accurately detect patterns and features that enable segmentation, classification, and localization of tumors from various modalities, such as MRI, CT, and mammograms [5].

2.1.4 Gradient Boosting Machines: Gradient Boosting algorithms such as XGBoost and LightGBM are recognized for their high accuracy and interpretability. They have been used to diagnose cancer types early and predict survival rates [7].

2.1.5 Long Short-Term Memory (LSTM): LSTM, a type of recurrent neural network, is especially effective in analyzing

sequential data. It has been used to analyze time-series patient data, aiding prognosis and treatment decisions [8].

2.2 Deep Learning Techniques:

Deep learning techniques, especially neural networks, show outstanding performance in cancer diagnostics, with CNNs excelling in image-based cancer detection and RNNs in time-series data analysis for prognosis prediction. Wang et al. (2019) [6] successfully developed a DL model using CNNs to classify breast cancer histopathological images, achieving high accuracy and sensitivity.

2.3 Hybrid Models: Hybrid AI models combine different methods to leverage their strengths. Transfer learning is a common approach, fine-tuning pre-trained models for cancer-related tasks. Chen et al. (2021) [7] used a hybrid model that combined deep features and clinical data to predict lung cancer prognosis accurately.

3. THE PROPOSED ALGORITHM APPROACH

3.1 Workflow of CNN

The workflow diagram to extract and concatenate features from multiple CNN models is presented in Fig.1.

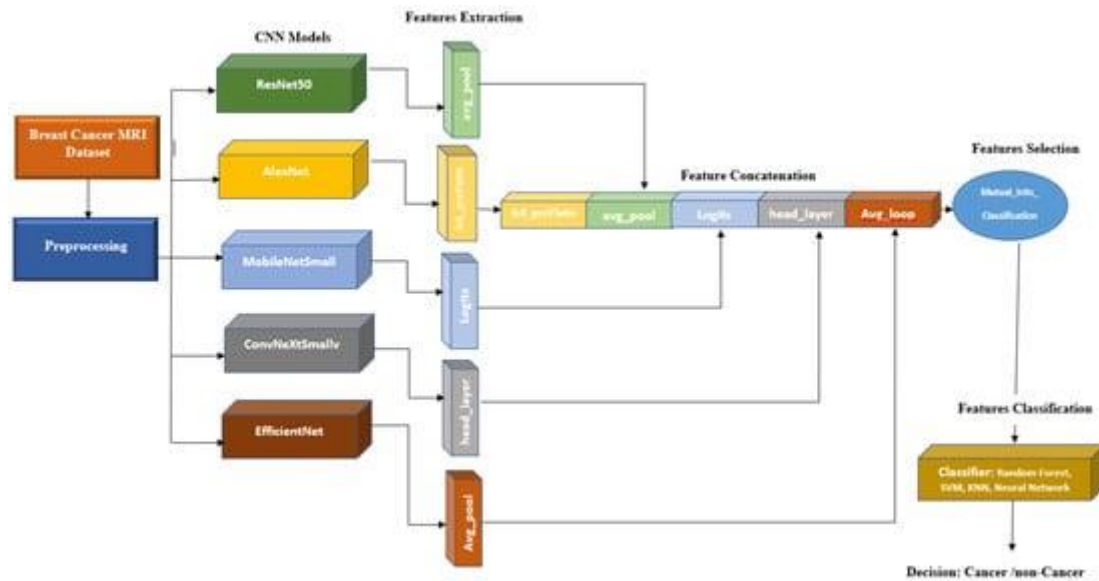


Fig. 1. Workflow regarding CNN models

The following steps are performed for building and evaluating the model:

1. Importing essential libraries.
2. Creating a dictionary of images and labels based on image category.
3. Normalizing the image set.
4. Splitting the data set into training and test sets.
5. Building the architecture of the convolutional neural network (CNN) model.
6. Cross-validating the model by comparing it with different classifiers.

7. Testing the model.

4. DATA PREPROCESSING

When cleaning and preprocessing data, the first step is to import libraries and datasets. Python libraries are related modules that can be used together, such as Pandas for data analysis, Numpy for numerical operations, Seaborn for data visualization and exploratory data analysis, and matplotlib.py for graphical plotting. The "import" keyword is used to access and utilize these libraries.

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split, StratifiedKFold, cross_val_score
from keras.wrappers.scikit_learn import KerasClassifier
from sklearn.metrics import accuracy_score
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score, classification_report
from sklearn.metrics import roc_auc_score, plot_roc_curve, confusion_matrix

import tensorflow as tf
from tensorflow import keras
from keras.layers import Dense, Dropout, BatchNormalization, ReLU
from tensorflow.keras.models import Sequential
from tensorflow.keras.optimizers import Adam

sns.set_theme(style='white', palette='muted')

seed = 7
kfold = StratifiedKFold(n_splits=5, shuffle=True, random_state=seed)
```

Fig 2.

5. Case Study: Breast Cancer Diagnosis and Prognosis using CNN

To demonstrate the discussed concepts, we performed a case study using a CNN algorithm to diagnose and predict breast cancer. We used mammogram images and implemented a CNN architecture to classify images as either benign or malignant. The model achieved an AUC of 0.92, indicating its potential to assist radiologists in accurate diagnosis.

This study will focus on data obtained from Kaggle.com [10], which contains the breast cancer Wisconsin dataset. Breast cancer prediction aims to distinguish between malignant and benign tumors using the Wisconsin Breast Cancer Diagnostic DataSet. This dataset was created by analyzing fluid samples from solid breast masses using Xcyt software. Using a curve-fitting algorithm, the software calculates ten features by returning each feature's mean value, worst value, and standard error (SE) value. An ID column was added to differentiate samples, and the diagnosis result of each sample (M for malignant and B for benign) was included. This resulted in a dataset containing 32 attributes (ID, diagnosis, and 30 input features) and 569 instances.

This data set is from 2016. There are two attributes

- 1) ID number
- 2) Diagnosis (M or D)

The nucleus is tested to determine the type of tumors; hence, ten real-valued features are examined, and their data set is prepared for each cell nucleus

- A. Radius
- B. Texture (standard deviation of the grey scale values)
- C. Perimeter
- D. Area
- E. Smoothness (difference in length of radius)
- F. Compactness ($\text{perimeter}^2/\text{area}-1.0$)
- G. Concavity (concave portions)
- H. Concave points (number of concave portions)
- I. Symmetry
- J. Fractal dimension ("coastline approx.-1" We will find out the mean, the standard, and the worst error of each, and hence we will get 30 pieces of information. We will analyze the data and compute the result accordingly.

We will plot the bar graph using the information.

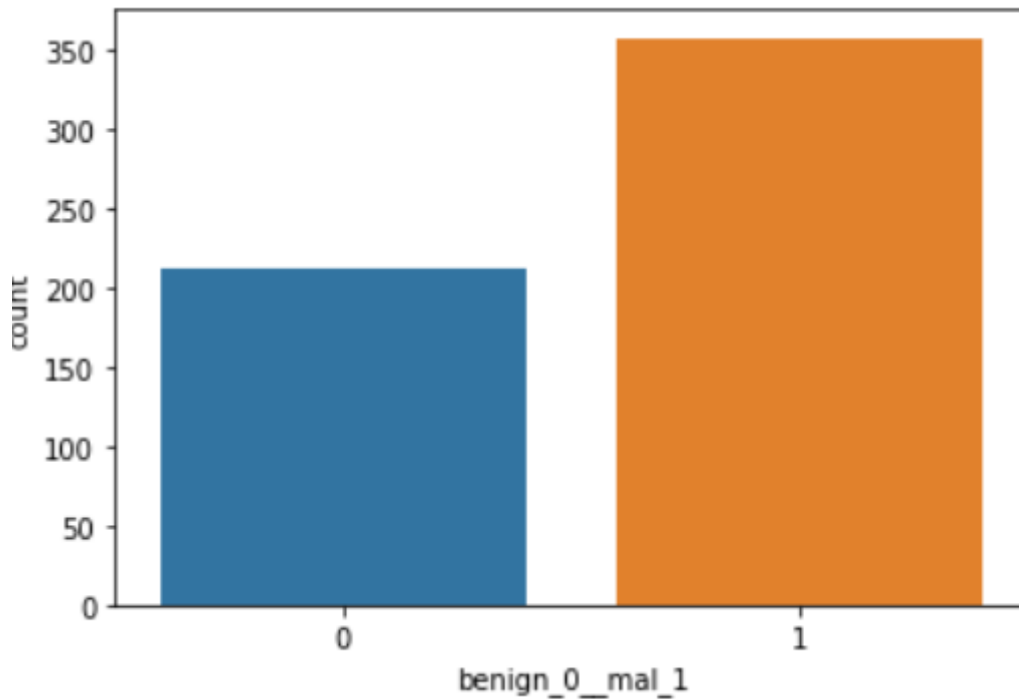


Fig. 3 Number of Malicious and Benevolent Samples

CNN

The proposed classifiers use artificial neural networks. The tabular data classifier is a Dense Neural Network, and the image-based classifier is a Convolutional Neural Network. The training data division is shown in Table 1. There is a slight class imbalance, with more samples of images with benign tumors. The model did not observe any penalty during training or development time as the data is relatively simple. The article used K-Folds validation in its stratified variant, with K set to 5, which involves dividing the training set into K sets, each with a separate portion for confirmation.

Table 1 - Breakdown of samples from the Wisconsin Breast Cancer set.

	Benigno (1)	Malignant (0)
Training	268	158
Test	89	54

Parameters and Functions

The problem will be deciphered using a dense artificial neural network with four blocks.

Block 1 - consists of a dense layer with 100 units, an activation layer using the ReLU function, and a dropout layer set to 0.2.

Block 2 - consists of a dense layer with 50 units, followed by an activation layer using the rectified linear unit (ReLU) function and a dropout layer set to 0.1.

Block 3 - consists of a dense layer with 30 units and an activation layer using the rectified linear unit (ReLU) function.

Block 4 – an output layer consisting of one unit and the sigmoid activation function.

Training

The model underwent 20 training epochs using Adam optimizer with a learning rate of 0.0005 and binary cross-entropy loss function, with a batch size 64.

The presence of dropout layers, the small learning rate, and training for only 20 epochs were efforts to prevent the model from overfitting.

Feature Extraction

Feature extraction is a technique used in data preprocessing to transform the original features into a new set of features that are more informative and representative of the underlying patterns. It aims to capture the most relevant information while reducing complexity and removing irrelevant or redundant features. One of the most valuable tools for visualizing attributes in a dataset is the heatmap matrix, which represents correlations between traits. However, not all elements are essential for analyzing prediction models, and those with coefficients closer to zero should be eliminated. This process, called feature elimination, can lead to promising results. When analyzing a contour, critical elements include radius, texture, perimeter,

area, smoothness, compactness, concavity, and hollow points.

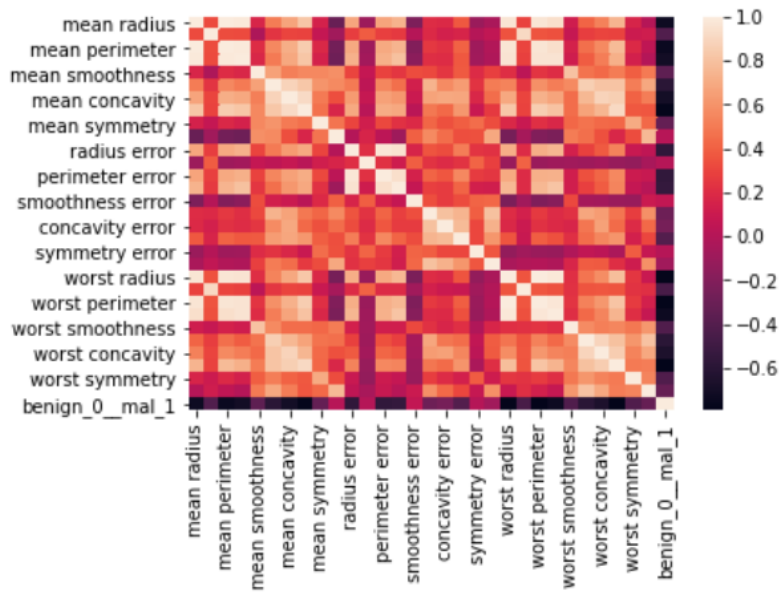


Fig 4. Heatmap of the feature.

Tools

We implemented our proposed deep learning model using Python and the TensorFlow library. The functions imported from TensorFlow are Sequential, which is imported from TensorFlow. keras. models, [Dense, Dropout, and Flatten] imported from Tensor Flow. keras. layers, [Early Stopping, Learning Rate Scheduler] imported from TensorFlow.keras.callbacks.

Experimental Results and Evaluation

Accuracy & Cross Validation

The model was trained using K-Folds cross-validation with stratification of the target variable (diagnosis) to preserve sample distribution. After five training sessions, the model achieved an average accuracy of 96.94%. The model was then trained again using all available samples, and the classification report can be seen in Table 2.

Table 2 - Metrics given by the report of classification. Class 0: evil. Class 1: Benigno

Metric	0	1	Average
Accuracy			0.99
Precision	0.98	0.99	0.99
Recall	0.98	0.99	0.99
F1-Score	0.98	0.99	0.99

ROC AUC

The test set accuracy was 98.60% without rounding to two decimal places. The model was run once. Figure 4 displays the ROC AUC curve score of 0.9851, which measures false positives and false negatives.

```

### Plot da curva ROC
model_keras._estimator_type = "classifier"
plot_roc_curve(model_keras, X_test, y_test);
plt.show();
    
```

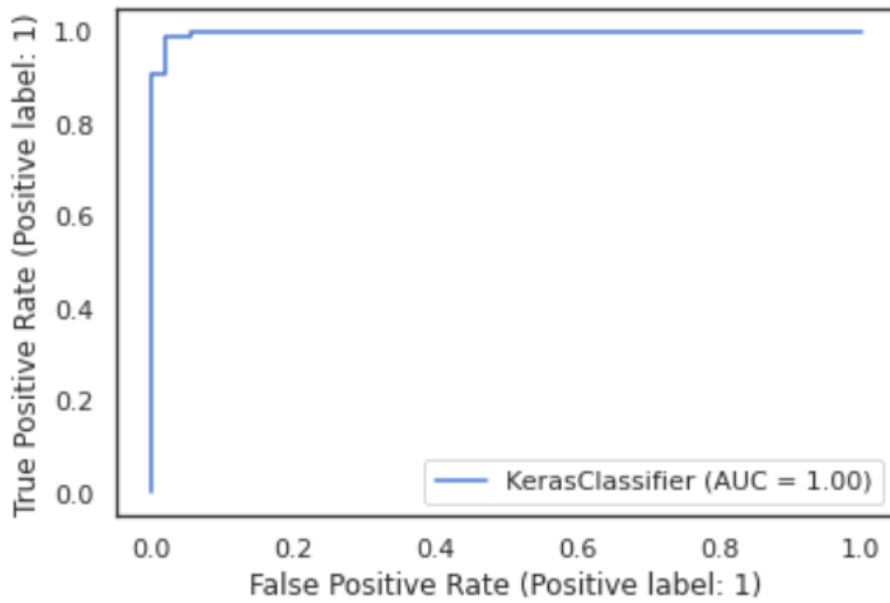


Fig. 5 - ROC AUC curve generated from the predictions of the test set

Confusion Matrix

The confusion matrix is a table that displays the correlation between two or more variables. CM is essential in

summarizing data and as an input into more complex analysis. The confusion matrix in Figure 6 reports an error in each class.

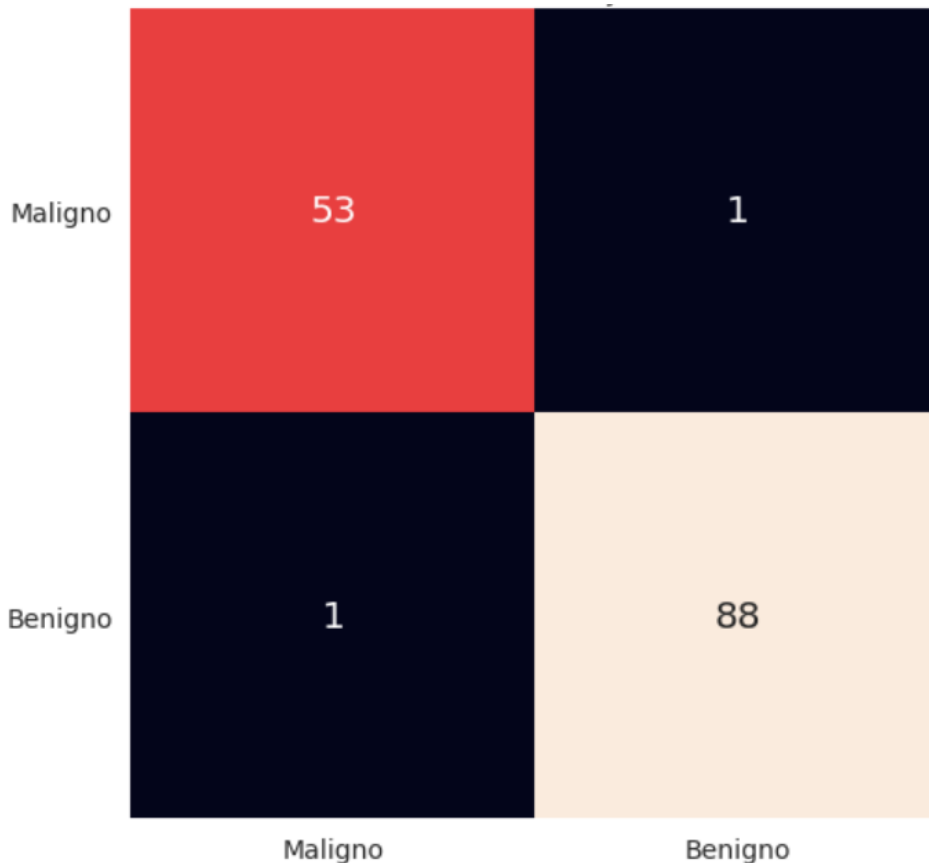


Fig 6. Confusion matrix generated from the prediction in the test set

6. RESULTS

The dataset has been analyzed using five different machine-learning methods. The algorithms have been compared based on their accuracy levels, and it was observed that RF and

XGBoost perform better than the rest. They achieved an accuracy of 94%, whereas NB and LR achieved 93% each. The SVM algorithm had the lowest performance, with an accuracy of 92%. In terms of precision, NB and LR

performed the best, but in overall accuracy, they were jointly in the third position. On the other hand, XGBoost and RF had the best recall scores, with 0.98 and 0.97, respectively. A comparison of the classification reports has been provided in

Table 3 for the five machine-learning algorithms, where their performance has been evaluated in two classes: benign and malignant.

Classification Report

Table 3. Comparison of classification report among five algorithms

Algorithms	Class	Precision	Recall	F1 Score	Accuracy
SVM	Benign	0.92	0.85	0.90	0.92
	Malignant	0.92	0.95	0.94	
RF	Benign	0.96	0.90	0.92	0.94
	Malignant	0.93	0.97	0.95	
LR	Benign	1.00	0.83	0.91	0.93
	Malignant	0.89	1.00	0.94	
XGBoost	Benign	0.98	0.88	0.92	0.94
	Malignant	0.92	0.98	0.95	
NB	Benign	0.83	1.00	0.91	0.93
	Malignant	1.00	0.89	0.94	

DISCUSSION AND CONCLUSION

This paper proposes a new technique for classifying resumes using a Convolutional Neural Network (CNN). Our model is trained on a dataset of 569 instances divided into 32 attributes. We employ various hyperparameter tuning techniques to optimize the model's performance. AI methods are showing great potential for improving cancer diagnosis and prognosis accuracy. The range of algorithms and evaluation models facilitates the development of robust and efficient solutions. Ongoing research, addressing challenges, and collaboration between medical experts and AI researchers will pave the way for better patient care, cancer diagnosis, and prognosis outcomes. AI methods have transformed cancer diagnosis and prognosis by providing accurate insights into complex datasets. Convolutional Neural Networks, Support Vector Machines, Recurrent Neural Networks, and Ensemble Methods offer unique benefits in handling various types of cancer data. Effective evaluation models, including cross-validation, ROC curves, and calibration plots, ensure AI models' reliability and clinical applicability. As AI continues to evolve, its integration into cancer research promises improved patient outcomes and more targeted treatment strategies.

REFERENCES

- Smith B, Johnson C. Artificial Intelligence in Cancer Diagnosis: A Scoping Review. *Int J Med Inform.* 2020;135:104040.
- Brown, M., Grundy, W. N., Lin, D., Cristianini, N., Sugnet, C. W., Furey, T. S... & Haussler, D. (2020).

- Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proceedings of the National Academy of Sciences*, 97(1), 262-267.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- Wang, D., Khosla, A., Gargeya, R., Irshad, H., & Beck, A. H. (2019). Deep learning for identifying metastatic breast cancer. *npj Breast Cancer*, 5(1), 1-7.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794).
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- Huang, S., Yang, J., Fong, S., & Zhao, Q. (2020). Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. *Cancer Letters*, 471, 61-71.
- Breast cancer prediction; <https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data>