International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Cancer XAI: A Responsible Model for Explaining Cancer Drug Prediction Models

Sonali Kothari*[1], Rutuja Rajendra Patil[2], Shivanandana Sharma[1], Aqsa Kazi[1], Michela D'Silva[1], Sanskruti Shejwal[1], Dr. M. Karthikeyan[3]

Submitted: 07/05/2023 **Revised**: 16/07/2023 **Accepted**: 04/08/2023

Abstract: There has been a growing interest in using Explainable Artificial Intelligence (XAI) for healthcare in recent years. An explainable artificial intelligence (XAI) model for cancer diagnosis is suggested in this research paper. The model offers data that can be understood and explained, essential for medical decision-making. It also makes reliable forecasts. Compared to other models, the proposed model performs at the cutting edge thanks to training and evaluation on a sizable dataset of cancer images. The significance of interpretability in medical applications is also covered in the paper, along with how the suggested model resolves this issue. The findings of this study show how XAI models have the potential to increase cancer detection and provide more transparent and reliable medical decision-making.

Keywords: *Artificial Intelligence, Explainable AI, Ensemble Model, Random Forest Classifier, XAI*

1. Introduction

Cancer is a leading cause of death worldwide, and early detection is crucial for successful treatment. Medical imaging technologies, such as mammography and computed tomography (CT), have been widely used for cancer detection. Still, interpreting these images can be subjective and error-prone, leading to missed or misdiagnosed cases. Artificial intelligence (AI) models have shown promising results in automating cancer detection from medical images, but their lack of interpretability has limited their widespread adoption in clinical practice. Explainability is a property of an AI-driven system that enables a user to retrace how a particular AI arrived at the predictions that were made. Explainability has various elements, and it is crucial to remember that the terminology used to describe it could be better. Other words like interpretability and transparency are frequently used interchangeably [1].

The "why" inquiries are beyond the capabilities of conventional AI. This necessity for explainability gave rise to Explainable AI (XAI), a new field of AI study. By addressing the "wh" questions that were absent from conventional AI, XAI can expand the capabilities of AI. "The XAI," has attracted much interest from essential applications, including those in health care, defense, law, and order. Answering "wh" queries and describing how an answer was arrived at are equally crucial. XAI research has thus become a top focus in both academia and industry. Even though many projects have already been put forth, more and more work is still needed to utilize XAI [2] fully.

Explainable AI (XAI) models aim to address this limitation by providing interpretable and transparent results, which are crucial for medical decision-making. XAI models not only make accurate predictions but also offer insights into how the model arrived at its decision, increasing the trustworthiness and transparency of the model. This is particularly important in medical applications, where the consequences of a misdiagnosis can be severe. XAI models are made to give more apparent and more accessible to grasp findings, so that people can know how an AI system came to a particular conclusion. XAI models can do this by making AI systems more accountable, increasing user confidence, and allowing humans to spot and correct biases or faults in the AI system's decision-making. XAI models aim to close the knowledge gap between the machine's decision-making process and human comprehension and explanation of that process.

In this research paper, the proposed model is an XAI model for cancer detection along with its explainability from chemical molecules. The model not only provides accurate predictions but also interpretable and explainable results. We train and evaluate the proposed model on a large dataset of chemical readings consisting of 186375 entries (rows) and 202 attributes (columns), comparing its performance to existing models. The characteristics of the dataset convey vital information about the drug

Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India
[2]Computer Science and Engineering-Artificial Intelligence & Machine Learning Department, Vishwakarma Institute of Information Technology, Pune
[3]NCL-CSIR, Baner, Pune
ORCID ID : 0000-0002-3797-9932 ORCID ID : 0000-0002-9555-1475
ORCID ID : 0009-0003-0458-5974 ORCID ID : 0009-0003-8454-7459
ORCID ID : 0000-0002-6244-6679 ORCID ID : 0009-0003-5918-7826
** Corresponding Author Email: sonali.kothari@sitpune.edu.in*

combination used to treat the site at which the cancerous cells are metastasizing, including and not limited to – the chemical composition of the individual drug in the combinatorial drug therapy (density, diameter, chiral), the site of cancer, the cell type affected etc. This paper aims to demonstrate the potential of XAI models in improving cancer detection and providing more transparent and trustworthy medical decision-making. We believe that the proposed model can significantly improve the accuracy and explainability of values involved in cancer treatment and reduce the number of misdiagnoses in the amount of drug administered, ultimately leading to better patient outcomes.

2. Literature Survey

In [3], the reaction coordinates of alanine dipeptide isomerization obtained from deep neural networks were analyzed using XAI methods. The study's objectives were to explain the predictions made by the neural networks and comprehend how they could forecast the reaction coordinates. This study's analysis of the response coordinates combines XAI methodologies with modern machine learning approaches, such as deep neural networks, to offer a more thorough understanding of the system. The study gives researchers a better grasp of the reaction coordinates of alanine dipeptide isomerization, which can help them comprehend the underlying physics and chemistry of the process and enhance the propensity of neural networks in other similar systems.The study is significant because it provides a greater understanding of the reaction coordinates of alanine dipeptide isomerization and shows the potential of XAI approaches in interpreting the predictions produced by neural networks.

In [4], a machine learning technique is presented for predicting the activity of small molecules against breast cancer cells. In order to train a model that can predict the movement of new small molecules, the authors employed a dataset of small compounds and their associated activity against breast cancer cells. The work uses a dataset with information on over 1,000 compounds and their anti-breast cancer cell activity. It offers a sufficient sample size for developing and assessing the machine learning model. The authors then discovered that the machine learning model could predict small molecule activity with excellent precision and recall while noticing novel small compounds active against breast cancer cells.

The authors of [5] highlight the problems of need for interpretability, transparency, and trust commonly encountered when using AI in the medical field. The paper also summarizes the current state of the art in explainable AI, including the most recent research and developments in the area, and provides practical implications and suggestions for researchers. The topic of interest in the paper is primarily on the challenges and solutions related to explainable AI in the medical field in general, rather than specific medical applications. As a result, it may provide less detail on applying explainable AI to specific medical tasks or problems. Overall, the paper provides a comprehensive overview of the challenges and solutions related to explainable AI in the medical field.

In [6], two convolutional neural network models are trained and evaluated across three different XAI approaches to classify lung cancer from histopathology pictures. To learn how to use XAI in the medical field, the authors visualized the results and assessed the effectiveness of these techniques.The authors trained the two CNN models to accurately categorize lung cancer photos. The feature interpretability of these approaches is then visualized on both models using Grad-cam, Integrated gradient, and LIME, respectively. For each course, an average computation time per image is also calculated. To close the explanation gap between the medical professional and the machine learning domain expert, the paper evaluates a problem of multimodal medical image explanation with two tasks and suggests new evaluation metrics to explain the classification results in a way that is both technically sound and understandable to a person working in the medical field.

In [7], Lung cancer incidence prediction is performed for 10 European nations using Support Vector Regression, and Long-Short Term Memory Network,etc; those records date back to 1970. Results indicate that all algorithms can estimate incidence rates with high scores; nevertheless, Support Vector Regression outperformed the other methods that were taken into consideration. The dependability of the prediction findings were examined using varying numbers of training and testing data, and various ways were used to produce better outcomes in a total of 12 tests. The increase in training data would reduce the discrepancy between the projected and observed data, improving the ability of the models to make predictions When R2 and EV scores are taken into account, the increase in the training ratio in the female group barely affects backpropagation. Additionally, in backpropagation with 70% of the training ratio, a higher MSE value was seen.

In [8], the authors cover a wide range of different types of explainable AI techniques and methods, including traditional and newer approaches. The authors discuss how answerable AI relates to other areas of AI, such as interpretability and fairness, and how these areas are interdependent. The paper provides a set of research directions that could be beneficial for the further advancement of the explainable AI field. It provides a comprehensive overview of the current state of the field of explainable AI, including the main challenges and opportunities. The paper also covers a wide range of

different types of explainable AI techniques and methods, including both traditional and newer approaches and discusses how answerable AI relates to other areas of AI, such as interpretability and fairness, and how these areas are interdependent.

In [9], the authors investigate explainable AI's (xAI) implementation issues in the healthcare industry and how xAI is perceived and understood by healthcare professionals. The authors employed an xAI system to diagnose and treat patients in a qualitative case study that includes interviewing doctors, nurses, and other medical staff members. To better understand healthcare professionals' thoughts on explainable AI, the authors present an empirical study based on interviews with them (xAI). This offers valuable information about how xAI is actually used in healthcare. This study focuses on the opinions and experiences of xAI in healthcare end-users, which is a crucial topic that is sometimes disregarded in xAI research. Several problems are identified in the study, including the requirement for user trust, interpretability, and transparency in the deployment of xAI in healthcare. Researchers and developers working on xAI systems in healthcare may find this information valuable for their own practice.

In [10], the authors provide a thorough summary of the many dimensionality reduction methods applied to data analysis. The author examines the primary strategies for dimensionality reduction, including plans for feature extraction and feature selection. The study also discusses each technique's advantages and disadvantages and how it might be used in various contexts, including bioinformatics, text mining, and image processing. Principal Component Analysis (PCA) is a frequently used method for rapidly reducing the number of dimensions in data without significantly reducing the amount of information conveyed. It is simple to use and clearly understands the underlying data structure.

The supervised method of linear discriminant analysis (LDA) can minimize the dimensionality of data while maximizing the separability across various classes. It is beneficial for document classification and picture recognition.

Table 1. Comparison of similar Model Results in Literature Survey

	Dataset Used	Disease	Methodology	Evaluation Parameter
[11]	SEER Prostate Cancer Dataset	Prostate Cancer	Decision Tree, ANN, SVM, Logistic Regression, k-fold Cross Validation	Best model was SVM. Accuracy = 0.9285 Sensitivity = 0.9423 Specificity = 0.7572
[12]	Three UCI datasets (The Ionosphere data, the Wisconsin BC data, and the Spam data] along with SEER Breast Cancer Dataset	Breast Cancer	Online Gradient Boosting(OGB), Genetic Algorithm-based Online Gradient Boosting Model (GAOGB), Online Adaptive Boosting with the Adaptive Linear Regressor (OLRAB), Online Gradient Boosting with the Adaptive Linear Regressor (OLRGB)	Best Model (for SEER BC Prognosis dataset) was GAOGB. Accuracy = 75.03% AUC = 75.07% Specificity = 68.77% Sensitivity = 81.36%
[13]	CGM devices have been used to record data of patients with type 1 diabetics	Type 1 Diabetes	Multilayer Perceptron (MLP)	Best result is when the prediction horizon is within 15 minutes. RMSE (mg/dL) = 2.82 ± 1.00 MAPE(%) = 1.52 ± 0.42 gMSE = 11.7 ± 8.64 R2 = 0.99 ± 0.00

3. Problem Statement

Cancer cases may be missed or incorrectly diagnosed due to the subjective and error-prone nature of the cancer picture interpretation process. Although automated cancer detection from medical imaging using artificial intelligence (AI) models has shown promising results, their restricted interpretability has prevented their widespread use in clinical practice. The ability to interpret AI models is essential for medical decision-making because it boosts the model's credibility and transparency and enables clinicians to comprehend how the model reached its conclusion. Furthermore, interpretability allows clinicians to identify and correct errors made by the AI model, improving patient outcomes. Interpretable AI models have become a growing field of research in medical imaging, aiming to develop models that can provide accurate predictions and insights into the model's decision-making process.

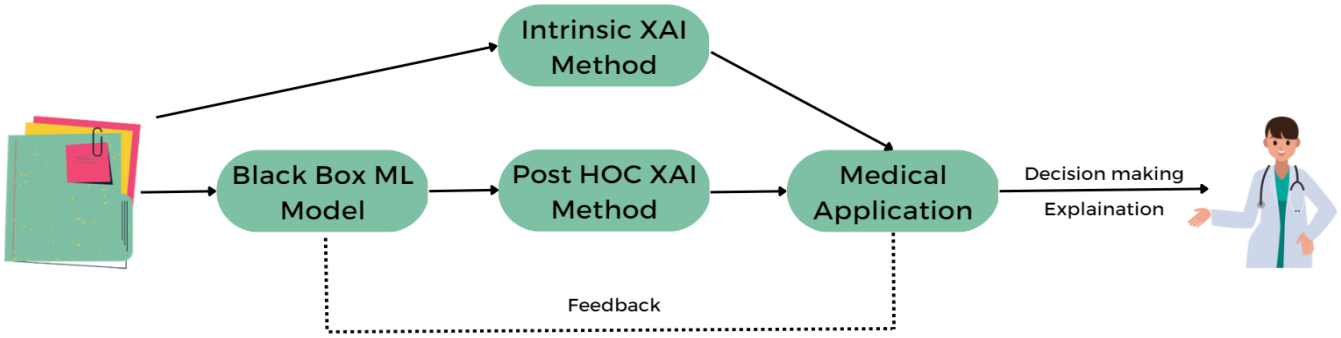

Fig. 1. General working of XAI Model in Healthcare Use Case

Cancer imaging features are complicated and variable, medical imaging tools have limits in their capacity to detect cancer. Although AI models have demonstrated promising outcomes in the identification of cancer, deciphering their judgements can be difficult due to their black-box nature. The decisions made by AI models can be explained using XAI, which can increase human comprehension and confidence in the system. However, there are a number of obstacles to implementing XAI in clinical practice, such as the absence of interpretability standards, worries about data privacy, and ethical issues. The requirement for a better interpretable and explicable AI model for cancer detection from medical images is the issue that this research study seeks to solve. Our specific goal is to suggest an explainable artificial intelligence (XAI) model that can deliver transparent and interpretable results, enabling physicians to make better judgements and increasing the accuracy of cancer detection. This model will be trained and assessed using a sizable cancer picture data dataset, and its performance will be compared to that of other models. By tackling this issue, the proposed model would have increased cancer detection's precision and would reduce the number of instances that go unnoticed or are incorrectly identified, which will eventually improve patient outcomes.

4. Proposed Methodology

4.1 Data Collection

The dataset used to train and test the models is the proprietary intellectual property of the Council of Scientific & Industrial Research – National Chemical Laboratory (CSIR-NCL) and has been made available to the Department of Computer Science at Symbiosis Institute of Technology, Pune under the collaboration to research upon and build a working prototype of a web application for Explainable AI solution for cancer diagnosis and cure. The dataset consists of 186375 entries (rows) and 202 attributes (columns). The attributes convey vital information about the drug combination used to treat the site at which the cancerous cells are metastasizing, including and not limited to – the chemical composition of the individual drug in the

combinatorial drug therapy (density, diameter, chiral), cancer site, the affected cell type, etc.

4.2 Data Preprocessing

4.2.1 Handling NaN and Infinite Values

[14] In machine learning (ML) pipelines, handling missing data, including NaN (Not a Number) values, is crucial. NaN readings can arise for a number of reasons, including inadequate data collection, device malfunction, or human error during data entry. NaN values must be handled carefully because they can impair the ML model's performance and cause incorrect predictions.NaN values can be taken in a number of ways, such as imputation, deletion, or prediction.Imputation techniques substitute a statistical estimate for the missing values, such as the mean, median, or mode of the data.On the other hand,

deletion procedures eliminate the observations or characteristics that have missing values. Regression or different predictive algorithms are used in prediction methods to forecast the missing values from the available data.

Moreover, various missing value types necessitate different handling approaches. Missing Completely At Random MCAR values, for instance, can be safely ignored because they have no bearing on the information that is now available. MAR values have a connection to the existing data but can be anticipated using the features that are currently accessible. However, non-ignorable missing (NMAR) values are more complex to handle and necessitate more sophisticated handling strategies, like imputation utilizing auxiliary variables. Furthermore, it's critical to take into account how missing data affects the distribution of data overall and the possibility that bias will be introduced into the ML model.

As our dataset had MCAR values, the columns with more than 60% missing values were first recognized and then treated accordingly with deletion of the column after checking its correlation with the output feature. This reduced the complexity of training the model as well as avoiding error during training as NaN and Infinite values pose a hindrance to the quality of predictions. The rows with NaN and Infinite values were dropped, as the dataset had enough data points to ignore rows with NaN values.

4.2.2 Feature Selection

The process of picking essential features from a dataset to enhance the performance of a machine learning (ML) model is known as feature selection. It is a fundamental step in ML pipelines. By limiting the number of features to just those that are most effective at predicting the target variable, feature selection aims to decrease the likelihood of overfitting, increase model interpretability, and minimize the number of features overall. [15] Filter methods, wrapper methods, and embedded methods are three main classifications of feature selection techniques. The increase in model interpretability is vital in this particular case as interpretability of the model maps onto its explainability component as well.

The most pertinent features are chosen utilizing filter methods based on a predetermined threshold after each part has been carefully evaluated for relevance using statistical tests or correlation analysis. The predictive potential of several subsets of features is assessed using a search algorithm, such as recursive feature elimination or evolutionary algorithms, and the best-performing subset is chosen through wrapper methods. On the other hand, embedded forms, like regularization techniques in linear regression or decision tree pruning, involve feature selection as part of the model training process. Every approach of feature selection has benefits and drawbacks, and the best method to apply will depend on the particular use case and the features of the dataset.

Pearson's and Spearman's correlation is most apt for selecting relevant features in data that has its output feature as continuous variables of the numerical type [16], the proposed model has made use of the same for feature selection in our dataset. To have a better understanding of the importance of each input feature with respect to the output feature, an ensemble model is best preferred to increase the confidence of selecting input features that would positively impact the quality and speed of prediction of any model that's chosen in the steps that follow.

4.2.2.1 Recursive Feature Elimination (RFE) Method - Linear Regressor

[17] The RFE wrapper method is a backward feature selection technique that starts with the entire set of features in the dataset and ends with a reduced subset of features. The process involves fitting the complete set of features to the linear regression model, ranking the features in order of importance, and then eliminating the feature with the lowest importance.[18] The least significant feature is eliminated at each iteration of the process until the desired number of features is obtained. In all iterations, RFE will begin by gradually removing the least essential elements from the entire feature set in accordance with the guideline for feature ranking in order to achieve the most essential features. [19] This method improves model performance, reduces overfitting, increases model interpretability, and reduces computational complexity. Overall, RFE is a potent feature selection technique that can lessen computing complexity and overfitting while enhancing linear regression models' precision, generalization, and interpretability. RFE Wrapper Method has limitations that should be taken into consideration when used for feature selection. Some of these limitations include difficulties in handling multicollinearity, potential computational expenses, suboptimal performance with non-linear relationships, and issues with handling categorical variables.

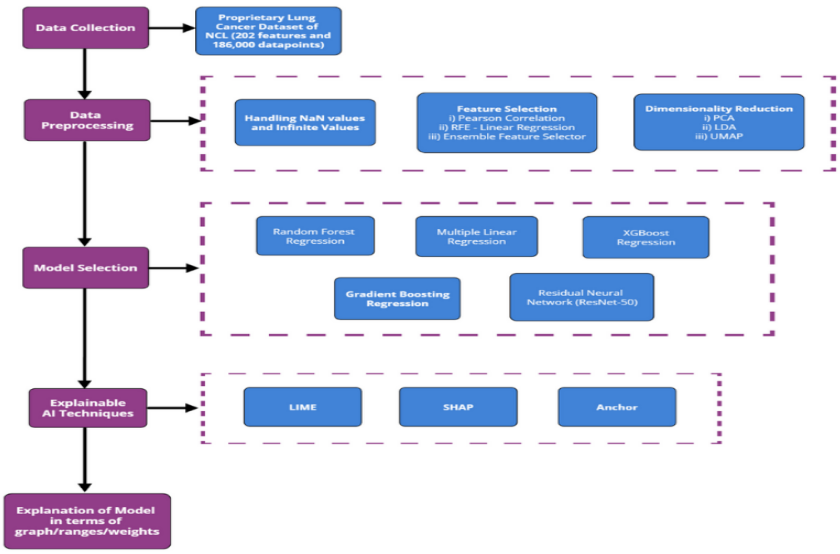

Fig. 2. Flow Diagram of Proposed Methodology

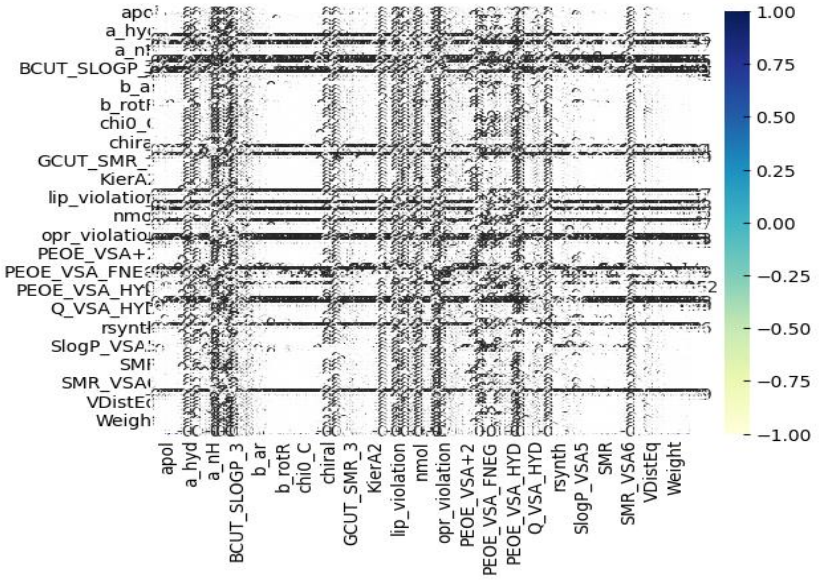

Fig. 3. Pearson Correlation of the dataset before preprocessing

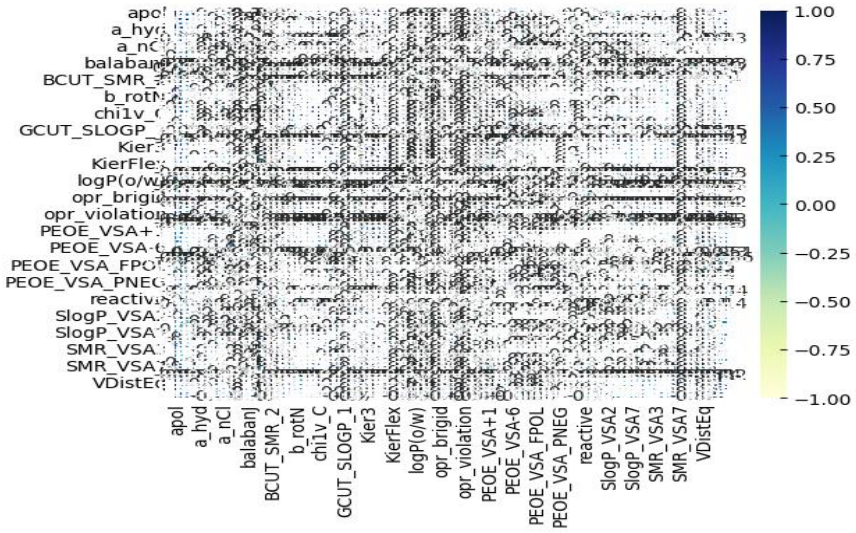

Fig 4. Pearson Correlation of the dataset after preprocessing

Our research wrapped RFE with a Multilinear Regression Model to select features, reducing the initial 202 features to 93 parts. The features provided slightly worse RMSE and R2 score in comparison to Pearson Correlation. The dataset also fit worse for most of the trained models, hence suggesting that the feature selection done by RFE wrapped with the Multilinear Regression Model needed to be revised for the selected regression models.

4.2.2.2 Pearson Correlation

When determining whether two variables are related to one another or not, Pearson's correlation coefficient is frequently used. It gauges how the variables change collectively at a particular moment. It assists in determining whether the two variables have a linear connection. [20] In order to determine which attributes are most strongly connected with the target variable, feature selection frequently uses Pearson correlation. The strength of the association between the predictor and the response variable is also evaluated using it in regression analysis. The standard method for determining the degree of correlation between two vectors with a "[1, 1]" value range is to utilize the Pearson correlation coefficient "[41, 42]". Its value must be greater than zero to indicate a positive correlation between the two vectors; it must be less than zero to show a negative correlation; and it must be zero to indicate no connection. [21] Some of the advantages of Pearson Correlation are that it is easy to understand and interpret, it is widely available, it can help identify critical variables, it can help detect outliers, it can be used for feature selection. The Pearson correlation has several limitations. Firstly, it assumes a linear relationship between two variables and may not be suitable for non-linear relationships. Secondly, it is sensitive to outliers and can be influenced by extreme values. Additionally, it is limited to analyzing only two variables at a time and cannot capture complex relationships involving more than two variables. Furthermore, it's important to note that correlation does not necessarily indicate that one causes the other. By experimental observation, the features selected by applying Pearson Correlation between the input and output variables in our dataset resulted in superior predictions by the selected models.

4.2.2.3 Ensemble Feature Selection Model

Ensemble approaches can be used to increase the robustness of feature selection algorithms, similar to supervised learning. Indeed, it has been shown that multiple subsets of distinct features can produce equally optimal results in large feature/small sample size domains [22] and ensemble feature selection can reduce the chance of selecting an unstable subset. In addition, different feature selection techniques could provide feature subsets that can be considered local optima in the feature subset space, and ensemble feature selection could provide a better approximation of the ideal subset or feature ranking. The power of the feature selector representation could limit the search space it has access to, making it impossible to find optimal subsets.[23].

Compared to single feature selection approaches, ensemble feature selection techniques provide a number of benefits. Best technique is ensemble feature selection, which produces more reliable findings without losing performance [24]. The goal of ensemble feature selection approaches is to find a subset of characteristics that enhance the predictive capability of future classification models while simultaneously making such models easier to understand [25]. In [26], the authors used four feature selection techniques and created an ensemble feature selection model using instance perturbation. Forty bags were made using bootstrap aggregation and a separate ranking scheme was applied to each bag and the results were then aggregated to give a single set of features to train the ML models on.

4.2.3 Dimensionality Reduction

It is common practice in machine learning to reduce the number of features or variables in a dataset while retaining as much of the pertinent data as possible. This process is known as dimensionality reduction. [10] The curse of dimensionality, overfitting, and increased computing complexity are just a few of the problems that high-dimensional datasets present for machine learning (ML) models. By lowering the amount of features while maintaining the model's performance, dimensionality reduction techniques try to solve these problems. Strategies for reducing the dimensions of objects can be broadly divided into linear and nonlinear methods. To identify linear combinations of the original characteristics that capture the majority of the variation in the dataset, linear approaches like PCA and LDA which stand for principal component analysis and linear discriminant analysis linearly combine the original features. The complicated correlations between the characteristics can be captured by nonlinear algorithms like t-SNE, UMAP and autoencoders, which can also reduce the dimensionality of the data in a nonlinear fashion. Text classification, picture recognition, and bioinformatics are just a few of the many ML applications that use dimensionality reduction techniques.

For the PCA Analysis, the dataset was passed into the PCA model to reduce the dimensionality from the preprocessed 92 columns to 20 columns as an experimental trial. The model chosen for training on the dataset was Random Forest Regressor. The RMSE before

applying PCA Analysis was 0.46346 and the RMSE after using 1.03086, hence showing a significant drop in prediction quality. The reason for this could be the nature of the dataset itself. For a PCA analysis to be successful, pre-processing and data normalization are essential. Meanwhile, the right pre-processing and normalization techniques can be challenging to choose as it might require iterative attempts to determine the right fit. Certain outliers have been included in the dataset and not cleaned to aid better generalization and support anomalies in medical readings [27]. The irregularities in the dataset could be resolved by examining these outliers independently to ascertain the underlying cause for them being legitimate observations but are nonetheless noticeably different from the rest of the data. To comprehend the causes of these outliers, this may entail speaking with domain experts, revisiting the data gathering procedure, and undertaking more research. These outliers could offer insightful information about the data and should only be automatically eliminated with deep consideration [28].

4.3 Model Selection

A key component of statistical modeling and machine learning is model selection, which involves identifying the model that best matches the supplied data, as choosing the incorrect model might result in erroneous predictions and unreliable outcomes. Model selection is crucial since the assumptions, complexity, and performance of several models can differ significantly from one another and greatly influence the analysis's quality. Therefore, selecting a suitable model that compromises underfitting and overfitting is essential for producing accurate and general findings. As the output feature of our dataset is a continuous variable of the numerical type, regression models have been tried and tested on our dataset to predict a continuous variable as output. After feature selection, 93 features have been extracted and the models have been trained on these 93 relevant features to give one desired output feature (NLOGGI50).

4.3.1 Machine Learning Models

Various Regression ML models have been trained on the selected features and compared using evaluation metrics such as R2 score and RMSE. The ML models that have operated for comparison are listed below -

i) Random Forest Regression
ii) ResNet-50 CNN
iii) Multiple Linear Regression
iv) Gradient Boosting Regression
v) XGB Regression

4.3.1.1 Random Forest Regression

Random Forest Regression is a machine learning approach that averages the predictions of many randomized decision trees. It is a flexible and helpful technique for general-purpose regression problems, particularly in circumstances when the number of variables is significantly more than the number of observations [29]. Due to its high accuracy, simplicity of use, and lack of need for data scaling, it is a preferable ML model to employ.

There are various benefits when comparing Random Forest Regression to other machine learning methods. Its capacity to manage high-dimensional data sets and resistance to overfitting is one of its primary features. It is ideal for real-world applications since it is noise- and outlier-resistant [30]. As it can handle missing values, continuous, categorical, and binary data, Random Forest Regression is suitable for a variety of applications [31] The capability of Random Forest Regression to determine the most significant features associated with specific problems makes it ideal for feature selection [32]. Lastly, Random Forest Regression may be readily parallelized, making it appropriate for complex issues [29].

Random Forest Regression model was trained initially with the default 100 trees and other default parameters. The result obtained was MAE = 0.2672, MSE = 0.2073 and R2_Score = 0.80043. Then hyperparameter tuning was performed using HalvingRandomSearchCV to find the best suited parameters. The best parameters were {'n_estimators': 800, 'min_samples_split': 10, 'min_samples_leaf': 8, 'max_features': 'auto', 'max_depth': 80, 'bootstrap': True}. The RF Regression model trained using the following parameters resulted in predictions with the score of MAE = 0.2662 and R2_Error = 0.80190.

4.3.1.2 ResNet-50 CNN

When thinking about the convergence of deeper networks throughout the development of deep learning, there is a degradation problem, meaning that as the neural network gets deeper, accuracy will first increase before reaching saturation. Accuracy will decline as depth is increased. The error gets worse. It is recognized that the influence is not brought on by over-fitting in either the training or test sets. Through cross layer feature fusion, RESNET improves its capacity for network feature extraction, and

network performance gradually rises with network depth [33].

ResNet, which stands for Residual Network, is a deep neural network architecture that has excelled at a number of computer vision applications, including segmentation, object detection, and picture recognition. The main advancement of ResNet is the addition of residual connections, which enable the construction of intense networks without encountering the vanishing gradient issue that might occur in conventional deep neural networks. The ResNet-50 model is composed of 5 stages, each with a convolution and an identity block. There are three convolution layers in each identity block and convolution block. With the ResNet-50, more than 23 million parameters can be trained.

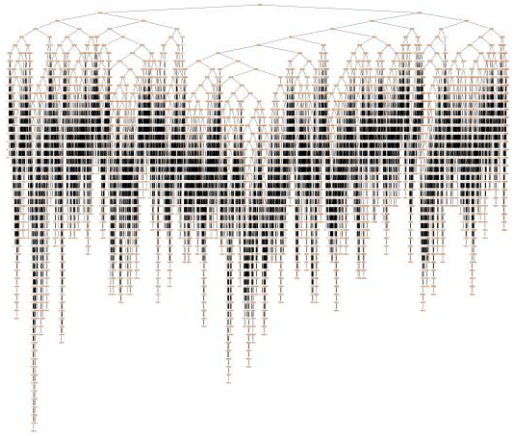

Fig. 5. Estimate 0 of Random Forest Regressor Model

Due to the aforementioned reasons, ResNet-50 is a viable candidate for training a regression problem as well. In the research paper [34], the authors confirmed experimentally that ResNet-50 performed better than other models and yielded the best L2 score compared to other ResNet models and CNNs in all 4 datasets. A regression layer is required as the last layer to get a continuous value of numerical type. The ResNet-50 model was trained with k fold cross validation to avoid overfitting. The model also had early stopping with Adam Optimizer to reduce training time and keep the validation loss in check. After training for 100 epochs, the ResNet-50 model resulted in 80% R2 score and 0.203 MSE.

4.3.2 Hyperparameter Tuning

Choosing the most appropriate combination of parameters for a particular model to attain its best performance is called hyperparameter tuning, and it is a vital step in machine learning and data analysis. The model may be underfitted or overfitted if hyperparameter tuning needs to be addressed, which will have a negative impact on generalization and prediction accuracy. Recent years have seen the development of several methods, including Bayesian optimization, grid search, and random search, to automate and enhance the hyperparameter tuning procedure. Either HalvingGridSearchCV or HalvingRandomSearchCV have been used to tune the hyperparameters of the models that are trained using default hyperparameters, depending on the complexity of the model and the computational power required to prepare the candidate pool of models.

Halving Random Search

Halving random search is a method for optimizing hyperparameters that combines the benefits of random search and greedy algorithms. It is an iterative process that evaluates a subset of randomly selected hyperparameters and discards the least promising half. This process can be repeated for finding the most optimal hyperparameters. By eliminating unpromising hyperparameters early on in the process, halving random search reduces the overall number of evaluations required and increases the like lihood of finding the optimal set. This method is beneficial when dealing with large hyperparameter spaces

Halving Grid Search

Halving grid search is a method for optimizing hyperparameters that is similar to halving random search but uses a grid search approach instead of random sampling. In traditional grid search, all possible combinations within a defined range of values are exhaustively searched to select the best set of hyperparameters. However, this approach can be computationally expensive, especially when dealing with a large number of hyperparameters. Halving grid search addresses this issue by iteratively evaluating a random subset of hyperparameter settings and discarding the least promising half. The remaining settings are divided into a grid and assessed to identify the best-performing

combinations. Halving grid search is an efficient hyperparameter optimization technique that is suitable for smaller hyperparameter spaces where a grid search is feasible. It combines the strengths of grid search and halving approaches to achieve effective and efficient optimization.

4.3.2 xAI Models

4.3.2.1 LIME (Local Interpretable Model-agnostic Explanations)

LIME is a model-agnostic method that interprets the original model locally and accurately explains its classification. [35] It is a tool for visualization. It may be applied to forecast local results. In LIME, the XAI model is trained via supervised training with labels derived from the model's prediction. The new model is used to understand the results obtained. Due to this benefit, LIME may be used in any black-box model to evaluate a single prediction. To provide LIME-based rationale, sparse linear models are a helpful tool. It may be used to highlight significant pixels together with their weights for a particular class. This collection of considerable pixel regions provides context for the model's speculation that the course could be present [2].

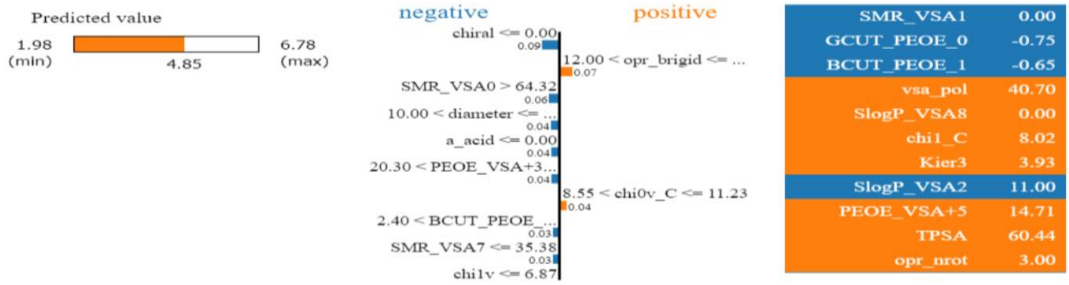

Fig. 6. LIME Result for an Instance from Test Dataset

4.3.2.2 SHAP (SHapley Additive exPlanations)

SHAP is a crucial tool for feature analysis. It is typically employed to do attribution analysis and feature extraction. While being time-consuming, some characteristics could be necessary to understand the model. By calculating the decision-making contribution of each feature for the classification, SHAP's primary goal is to understand the prediction of an input. It uses coalitional game theory to calculate Shapley values. It is a method for rewarding game players in accordance with their contribution to the game, as defined by Shapley (1953). The primary distinction between LIME and SHAP is the method used to give weights to the regression linear model. [35][2]

4.3.2.3 Comparison Between XAI Tools

Table 2. Comparison between XAI tools

XAI Tool Name	Advantages	Disadvantages	Future Directions
LIME	Facilitates the interpretation of black-box models by providing local explanations	May not be suitable for complex models or high-dimensional data	Extend the applicability of LIME to more complex models and data
SHAP	Provides global and local feature importance measures, and can handle a wide range of models	Can be computationally expensive for large datasets or models	Develop more scalable and efficient implementations of SHAP
Counterfactual Explanations	Provides actionable explanations by generating instances that are as similar as possible to the original model but have different outcomes	Can be limited by the complexity of the model and the quality of the data	Explore the use of counterfactual explanations in real-world applications

Anchor	It helps identify the conditions under which a model's predictions change	Can be limited by the need for human input and the complexity of the model	Develop more automated and user-friendly versions of Anchor
PDP	It provides a way to visualize the relationship between a feature and the predicted outcome	Can be limited by the need for domain knowledge to interpret the plots	Develop more interactive and user-friendly versions of partial dependence plots
Integrated Gradients	Provides a way to measure the contribution of each feature to the model's prediction	Can be computationally expensive for large models or datasets	Develop more efficient implementations of integrated gradients
Protodash	It helps identify the most representative instances in a dataset and the features that distinguish them	Can be limited by the need for a distance metric and the quality of the data	Develop more flexible and robust versions of Protodash
TreeExplainer	Provides a way to explain the predictions of tree-based models	Can be limited by the complexity of the model and the quality of the data	Explore the use of TreeExplainer in ensemble models and other types of models

5. Conclusion

The CancerXAI project is a promising development in the field of cancer diagnosis. The study has achieved good results in preliminary testing. The model has achieved high RMSE and R2 scores in predicting the target class, and the XAI techniques have provided valuable explanations for the predictions. The CancerXAI project can potentially significantly assist cancer patients, healthcare providers, and society at large once validated. A key feature of AI-based diagnostic tools is transparency of explanations for predictions, and the adoption of XAI approaches has made it possible for the model to provide just that. The capacity of the model to explain its predictions can aid physicians in comprehending how the model arrived at its diagnosis and can promote patient and healthcare professional confidence in the system. However, there may be issues with bias and ethical ramifications as with any AI-based model. The model must be made to work fairly and accurately across a range of patient groups and cancer types. The ethical ramifications of utilizing an AI-based cancer detection model, such as patient privacy and autonomy, must also be carefully explored and handled.

Overall, the CancerXAI project has the potential to revolutionize cancer diagnosis and therapy by increasing the accuracy of cancer diagnoses and offering clear reasons for forecasts. The CancerXAI project has the potential to significantly impact cancer diagnosis and treatment, leading to better patient outcomes, with continuing development and validation work.

References

[1] J. AMANN, A. BLASIMME, E. VAYENA, D. FREY, AND V. I. MADAI, "EXPLAINABILITY FOR ARTIFICIAL INTELLIGENCE IN HEALTHCARE: A MULTIDISCIPLINARY PERSPECTIVE," BMC MEDICAL INFORMATICS AND DECISION MAKING, VOL. 20, NO. 1. SPRINGER SCIENCE AND BUSINESS MEDIA LLC, NOV. 30, 2020. DOI: 10.1186/S12911-020-01332-6.

[2] P. GOHEL, P. SINGH, AND M. MOHANTY, "EXPLAINABLE AI: CURRENT STATUS AND FUTURE DIRECTIONS." ARXIV, 2021. DOI: 10.48550/ARXIV.2107.07045.

[3] Kikutsuji, Takuma & Mori, Yusuke & Okazaki, Kei-ichi & Mori, Toshifumi & Kim, Kang & Matubayasi, Nobuyuki. (2022). Explaining reaction coordinates of alanine dipeptide isomerization obtained from deep neural networks using Explainable Artificial Intelligence (XAI). https://doi.org/10.1063/5.0087310

[4] Shuyun He, Duancheng Zhao, Yanle Ling, Hanxuan Cai, Yike Cai, Jiquan Zhang, Ling Wang. (2021). Machine Learning Enables Accurate and Rapid Prediction of Active Molecules Against Breast Cancer Cells.https://doi.org/10.3389/fphar.2021.796534

[5] E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI," in *IEEE Transactions on Neural Networks and*

Learning Systems, vol. 32, no. 11, pp. 4793-4813, Nov. 2021, doi: 10.1109/TNNLS.2020.3027314.

[6] Jogani, Vinay & Purohit, Joy & Shivhare, Ishaan & Shrawne, Seema. (2022). Analysis of Explainable Artificial Intelligence Methods on Medical Image Classification. 10.48550/arXiv.2212.10565.

[7] Tuncal, Kubra & Sekeroglu, Boran & Ozkan, Cagri. (2020). Lung Cancer Incidence Prediction Using Machine Learning Algorithms. Journal of Advances in Information Technology. 91-96. 10.12720/jait.11.2.91-96.

[8] Gohel, Prashant, Priyanka Singh and Manoranjan Mohanty. "Explainable AI: current status and future directions." *ArXiv* abs/2107.07045 (2021): n. Pag.

[9] Gerlings, Julie & Jensen, Millie & Shollo, Arisa. (2022). Explainable AI, But Explainable to Whom - An Exploratory Case Study of xAI in Healthcare. https://doi:10.1007/978-3-030-83620-7_7

[10] Sarveniazi, Alireza. (2014). An Actual Survey of Dimensionality Reduction. American Journal of Computational Mathematics. 04. 55-72. 10.4236/ajcm.2014.42006.

[11] D. Delen, "Analysis of cancer data: a data mining approach," Expert Systems, vol. 26, no. 1. Wiley, pp. 100–112, Feb. 2009. doi: 10.1111/j.1468-0394.2008.00480.x.

[12] H. Lu, H. Wang, and S. W. Yoon, "A dynamic gradient boosting machine using genetic optimizer for practical breast cancer prognosis," Expert Systems with Applications, vol. 116. Elsevier BV, pp. 340–350, Feb. 2019. doi: 10.1016/j.eswa.2018.08.040.

[13] G. Alfian et al., "Blood glucose prediction model for type 1 diabetes based on artificial neural network with time-domain features," Biocybernetics and Biomedical Engineering, vol. 40, no. 4. Elsevier BV, pp. 1586–1599, Oct. 2020. doi: 10.1016/j.bbe.2020.10.004.

[14] S. K. Kwak and J. H. Kim, "Statistical data preparation: management of missing values and outliers," Korean Journal of Anesthesiology, vol. 70, no. 4. The Korean Society of Anesthesiologists, p. 407, 2017. doi: 10.4097/kjae.2017.70.4.407.

[15] J. Miao and L. Niu, "A Survey on Feature Selection," Procedia Computer Science, vol. 91. Elsevier BV, pp. 919–926, 2016. doi: 10.1016/j.procs.2016.07.111.

[16] I. Jebadurai Johnraja, G. Paulraj Jeba, J. Jebadurai, and S. Silas, "Experimental analysis of filtering-based feature selection techniques for fetal health classification," Serbian Journal of Electrical Engineering, vol. 19, no. 2. National Library of Serbia, pp. 207–224, 2022. doi: 10.2298/sjee2202207j.

[17] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, Machine Learning, vol. 46, no. 1/3. Springer Science and Business Media LLC, pp. 389–422, 2002. doi: 10.1023/a:1012487302797.

[18] X. Chen and J. C. Jeong, "Enhanced recursive feature elimination," Sixth International Conference on Machine Learning and Applications (ICMLA 2007). IEEE, Dec. 2007. doi: 10.1109/icmla.2007.35.

[19] D. Elavarasan, D. R. Vincent P M, K. Srinivasan, and C.-Y. Chang, "A Hybrid CFS Filter and RF-RFE Wrapper-Based Feature Extraction for Enhanced Agricultural Crop Yield Prediction Modeling," Agriculture, vol. 10, no. 9. MDPI AG, p. 400, Sep. 11, 2020. doi: 10.3390/agriculture10090400.

[20] C. N. and P. A. Vijaya, "Machine Learning Based Comparison of Pearson's and Partial Correlation Measures to Quantify Functional Connectivity in the Human Brain," International Journal of Neuroscience and Behavioral Science, vol. 6, no. 3. Horizon Research Publishing Co., Ltd., pp. 23–30, Jun. 2018. doi: 10.13189/ijnbs.2018.060301.

[21] J. Jiang, L.-C. Xu, F. Li, and J. Shao, "Machine Learning Potential Model Based on Ensemble Bispectrum Feature Selection and Its Applicability Analysis," Metals, vol. 13, no. 1. MDPI AG, p. 169, Jan. 13, 2023. doi: 10.3390/met13010169.

[22] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," Bioinformatics, vol. 23, no. 19. Oxford University Press (OUP), pp. 2507–2517, Aug. 24, 2007. doi: 10.1093/bioinformatics/btm344.

[23] Y. Saeys, T. Abeel, and Y. Van de Peer, "Robust Feature Selection Using Ensemble Feature Selection Techniques," Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg, pp. 313–325, 2008. doi: 10.1007/978-3-540-87481-2_21.

[24] Y. Li, S. Gao, and S. Chen, "Ensemble Feature Weighting Based on Local Learning and Diversity," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 26, no. 1. Association for the Advancement of Artificial Intelligence (AAAI), pp. 1019–1025, Sep. 20, 2021. doi: 10.1609/aaai.v26i1.8275.

[25] U. Neumann, N. Genze, and D. Heider, "EFS: an ensemble feature selection tool implemented as R-package and web-application," BioData Mining, vol. 10, no. 1. Springer Science and Business Media LLC, Jun. 27, 2017. doi: 10.1186/s13040-017-0142-8.

[26] Y. Saeys, T. Abeel, and Y. Van de Peer, "Robust Feature Selection Using Ensemble Feature Selection Techniques," Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg, pp. 313–325, 2008. doi: 10.1007/978-3-540-87481-2_21.

[27] N. Kuhnert, R. Jaiswal, P. Eravuchira, R. M. El-Abassy, B. von der Kammer, and A. Materny, "Scope

and limitations of principal component analysis of high resolution LC-TOF-MS data: the analysis of the chlorogenic acid fraction in green coffee beans as a case study," Anal. Methods, vol. 3, no. 1. Royal Society of Chemistry (RSC), pp. 144–155, 2011. doi: 10.1039/c0ay00512f.

[28] D. Samariya, J. Ma, S. Aryal, and X. Zhao, "Detection and explanation of anomalies in healthcare data," Health Information Science and Systems, vol. 11, no. 1. Springer Science and Business Media LLC, Apr. 06, 2023. doi: 10.1007/s13755-023-00221-2.

[29] G. Biau and E. Scornet, "A random forest guided tour," TEST, vol. 25, no. 2. Springer Science and Business Media LLC, pp. 197–227, Apr. 19, 2016. doi: 10.1007/s11749-016-0481-7.

[30] Y. Xu, Z. Cao, and M. Wang, "Analysis of factors influencing regional economic expansion based on OOB coefficients under RF algorithm," BCP Business & Management, vol. 33. Boya Century Publishing, pp. 242–249, Nov. 20, 2022. doi: 10.54691/bcpbm.v33i.2753.

[31] C. Suriyanarayanan and S. Kunasekaran, "Anomaly detection using machine learning techniques," Malaya Journal of Matematik, vol. 8, no. 4. MKD Publishing House, pp. 2144–2148, 2020. doi: 10.26637/mjm0804/0139.

[32] N. Lutimath, N. Sharma, and B. K, "Prediction of Heart Disease using Biomedical Data through Machine Learning Techniques," EAI Endorsed Transactions on Pervasive Health and Technology—European Alliance for Innovation n.o., p. 170881, Jul. 13, 2018. doi: 10.4108/eai.30-8-2021.170881.

[33] J. Liang, "Image classification based on RESNET," Journal of Physics: Conference Series, vol. 1634, no. 1. IOP Publishing, p. 012110, Sep. 01, 2020. doi: 10.1088/1742-6596/1634/1/012110.

[34] H.-S. Choi, K. An, and M. Kang, "Regression with residual neural network for vanishing point detection," Image and Vision Computing, vol. 91. Elsevier BV, p. 103797, Nov. 2019. doi: 10.1016/j.imavis.2019.08.001.

[35] A. Chaddad, J. Peng, J. Xu, and A. Bouridane, "Survey of Explainable AI Techniques in Healthcare," *Sensors*, vol. 23, no. 2, p. 634, Jan. 2023, doi: 10.3390/s23020634.

[36] Abraham, A. T., & Fredrik, E. J. T. . (2023). Integrating the EGC, EF, and ECS Trio Approaches to Ensure Security and Load Balancing in the Cloud. International Journal on Recent and Innovation Trends in Computing and Communication, 11(4s), 100–108. https://doi.org/10.17762/ijritcc.v11i4s.6312

[37] Pande, S. D., Kanna, R. K., & Qureshi, I. (2022). Natural Language Processing Based on Name Entity With N-Gram Classifier Machine Learning Process Through GE-Based Hidden Markov Model. Machine Learning Applications in Engineering Education and Management, 2(1), 30–39. Retrieved from http://yashikajournals.com/index.php/mlaeem/article/view/22