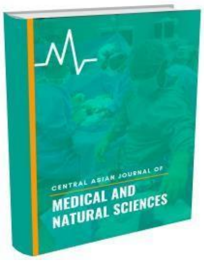


CENTRAL ASIAN JOURNAL OF
MEDICAL AND NATURAL SCIENCES
https://cajmns.casjournal.org/index.php/CAJMNS
Volume: 07 Issue: 01 | January 2026 ISSN: 2660-4159

Article

Integrating Oncology Prediction and Prognostication for Precision Medicine Decision-Making

A.T. Ashmi Christus[1], M. Mohamed Sameer Ali[2], B. Vaidianatha*[3]

1,2,3. Dhaanish Ahmed College of Engineering, Padappai, Chennai, Tamil Nadu, India
* Correspondence: vaidianathan@dhaanishcollege.in

Abstract: Breast cancer continues to be a primary cause of cancer-related mortality in women worldwide, highlighting the need for the creation of effective predictive models to facilitate early identification and treatment. This project is about building a machine learning-based prediction model using linear regression to estimate the likelihood that breast cancer is malignant or benign based on different clinical criteria. The project is divided into three parts: Data Collection, Model Training, and Prediction. It uses the breast cancer dataset from the "sklearn" package. The Data Collection Module loads and preprocesses the dataset to make sure it is clean and ready for analysis. The Model Training Module uses linear regression to train the model and assesses its effectiveness using accuracy metrics on both the training and test datasets. The Prediction Module makes predictions in real time, so users can add new data and receive instant feedback on potential diagnoses. The study shows how well machine learning works in medicine by putting these parts together. It also shows how important it is to make accurate, on-time predictions to improve patient outcomes. Future improvements might include exploring more complex algorithms, improving model performance, and adding more datasets. All of these things would help with continued research and detection of breast cancer. By bringing these parts together, the project shows how useful machine learning can be in medicine and how important it is to make accurate, timely predictions to improve patient outcomes. Future improvements may include exploring more complex methods, tuning hyperparameters to improve performance, and adding more datasets to strengthen the model. All of these things would help with breast cancer research and detection.

Citation: Christus A. T. A., Sameer Ali M. M., Vaidianathan B. Integrating Oncology Prediction and Prognostication for Precision Medicine Decision-Making. Central Asian Journal of Medical and Natural Science 2026, 7(1), 525-540.

Received: 28th Nov 2025
Revised: 10th Dec 2025
Accepted: 25th Dec 2025
Published: 12th Jan 2026

Keywords: Breast Cancer, Causes Of Cancer, Diagnosis And Treatment, Data Collection, Training Module, Research And Diagnosis, Enhancing Patient, Advanced Algorithms

1. Introduction

Despite major improvements in treatment and awareness, breast cancer is still one of the most common causes of cancer deaths in women around the world. Finding the disease early is important for improving patient outcomes because it allows doctors to intervene and treat it quickly [36]. Conventional diagnostic techniques, although efficacious, often involve invasive procedures and may not consistently yield precise predictions of disease progression. Machine learning (ML) is a burgeoning field that analyses large volumes of medical data and makes predictions based on patterns that may not be obvious to human experts. This is a viable alternative [55]. Machine learning (ML) models can be trained to predict the likelihood of breast cancer developing or recurring using historical patient data, such as demographic information, tumour features, and genetic markers. This method not only makes forecasts more accurate but also helps reduce the need for intrusive procedures, making it a non-invasive and effective way to find problems early [45]. We want to use different machine learning techniques, such as Logistic Regression, Random Forests, and Support Vector Machines, in this study to determine whether they can predict breast cancer outcomes.

This study aims to determine the most effective algorithm for breast cancer prediction by comparing the models' performance based on accuracy, sensitivity, and specificity, hence enhancing the reliability of diagnostic processes in clinical environments. Breast cancer is still one of the most common causes of death in women. Early detection is very important for successful treatment and better survival rates [37]. While traditional diagnostic approaches work, they often require invasive procedures and don't always provide accurate prognostic information [70]. Moreover, the vast volume and complexity of medical data, encompassing patient demographics, tumour characteristics, and genetic markers, pose a significant obstacle for doctors in making prompt, precise decisions [46]. This research aims to address the following question: given machine learning's capacity to analyse large datasets and uncover hidden patterns, how can we use machine learning algorithms to forecast breast cancer outcomes better, helping doctors detect the disease earlier and make more accurate diagnoses? The purpose is to create and test different machine learning models, such as Logistic Regression, Random Forests, and Support Vector Machines, to assess their ability to detect breast cancer by evaluating accuracy, sensitivity, specificity, and overall predictive performance [54].

The goal of the solution is to develop a reliable, non-invasive tool that improves breast cancer prognosis accuracy, leading to better diagnosis and treatment planning in clinical settings [63]. The main goal of this project is to develop a reliable and accurate machine learning model for predicting breast cancer outcomes, with an emphasis on early detection and supporting healthcare professionals' decision-making [38]. This project aims to create a non-invasive tool that analyses patient data to determine whether a tumour is benign or malignant using several machine learning algorithms. This would improve diagnostic accuracy and reduce the need for invasive procedures [48]. The first step in the project will be to gather and clean the data. This means gathering a lot of information about breast cancer, such as information about the patients, the tumours, and the genetic markers [68].

We will clean, standardise, and transform the data into a format suitable for training machine learning models [44]. This will ensure that any missing or incorrect data is handled appropriately. The next step in the project is to develop and train machine learning models. To create prediction models, we will use several methods, including Logistic Regression, Random Forests, and Support Vector Machines (SVMs). The goal of training these models on historical patient data is to forecast breast cancer outcomes accurately. We will use cross-validation and hyperparameter tuning to stabilise the model and avoid overfitting [62]. The project will also assess and compare how well these models perform using key metrics such as accuracy, sensitivity, specificity, and precision. By comparing models, we will be able to identify the one that best predicts breast cancer [35]. We will also perform a feature importance analysis to identify which factors are most important for predicting. This will provide doctors with more information. Finally, the research will examine how to combine and, if appropriate, apply the best-performing model in real-world clinical settings [53]. The goal is to create a useful tool that may help healthcare professionals better diagnose and treat breast cancer cases.

Scope and Motivation

The goal of this project is to create a machine learning-based system that can predict the outcomes of breast cancer [41]. The main goals are to help doctors make better decisions and to find the disease early. There are a few important steps in the project:

- Gathering and preparing data: The study will use publicly available datasets containing information on patients' demographics, tumour features, and genetic markers. To prepare the dataset for analysis, it will undergo data preprocessing, including cleaning, normalisation, and feature selection.
- Model Development: To build models that predict outcomes, we will use machine learning methods such as Logistic Regression, Random Forests, and Support Vector Machines. To distinguish between benign and malignant tumours, these models will be trained on data from previous patients.
- Model Evaluation: We will use metrics such as accuracy, sensitivity, specificity, and F1-score to assess how well each model performs [69]. The initiative will also assess how well these models perform to identify the best algorithm for predicting breast cancer.
- Feature Importance Analysis: The study will examine how different variables (such as tumour size and age) affect the model's predictions. This will help us understand what factors are most important in diagnosing breast cancer.
- Deployment Possibilities: The main goal of the project is to create and test models, but it will also look into how they could be used in clinical settings to help doctors make early diagnoses and predictions.

The scope does not encompass advanced image-based analysis (e.g., mammography interpretation utilising deep learning) or the creation of novel algorithms [64]. The emphasis will be on utilising and evaluating current machine learning methodologies on structured clinical data. The impetus for this study arises from the pressing necessity for enhanced breast cancer diagnostic instruments. Breast cancer is one of the most common causes of death in women around the world. Early identification is crucial for improving survival rates [34]. But traditional diagnostic approaches generally rely on intrusive procedures and subjective analysis, which can delay diagnosis or lead to incorrect predictions. Machine learning is a promising option because it provides data-driven models that can analyse patient data more objectively and quickly than humans can [52]. As healthcare data becomes more accessible, machine learning algorithms can find patterns and connections that might not be obvious with traditional methods [43]. This leads to more accurate predictions and improved treatment outcomes.

Existing System

In contemporary clinical practice, the diagnosis and prognosis of breast cancer predominantly depend on conventional procedures, typically encompassing a combination of medical imaging, biopsy, and histological analysis [40]. These methods are well-known, but they have several problems that make it harder to find and predict breast cancer outcomes early on.

Medical Imaging: Mammograms, ultrasounds, and MRIs are among the most common methods for detecting abnormal growths in breast tissue [49]. Imaging is good for detecting lumps or masses, but it doesn't always distinguish between benign and malignant cancers. Also, the accuracy of these methods generally depends on the radiologist's ability to interpret the images, which means there is a risk of human error or bias [67].

Biopsy and Histopathology: A biopsy, which involves taking a tissue sample and examining it under a microscope, remains the best way to confirm the presence of cancer. It is very accurate, but it is also intrusive, time-consuming, and expensive [47]. Also, it can take several days to get the results of biopsies, which can delay diagnosis and therapy. In some situations, biopsies might cause problems like infections or pain for the patient.

Statistical Models: The Gail Model and the Breast Cancer Risk Assessment Tool are two examples used in various medical contexts to estimate a person's likelihood of developing breast cancer [51]. These models use demographic and clinical information, like age, family history, and reproductive history. But they typically don't work well since they only use a small number of features and can make the complicated elements that lead to breast cancer too simple [65].

Limited Use of Machine Learning: People have tried using machine learning to predict breast cancer over the past few years, but it hasn't caught on in clinical settings yet [56]. Some current systems use simple machine learning models, such as Decision Trees or Logistic Regression, on tiny datasets. Still, their predictive ability is generally limited by poor preprocessing, insufficient feature selection, and insufficient training data [42]. Many of these systems don't work well with different patient types, which makes their results inconsistent and limits their usefulness in real life.

Problems with Early Detection: One of the biggest problems with the current approach is that it can't find cancer early enough. Sometimes imaging tools can't detect tiny cancers or tumours in thick breast tissue [39]. This can lead to missed diagnoses or late diagnoses, when the cancer has already spread to more advanced stages, making treatment less likely to work. In short, current systems are useful for diagnosing problems, but they have drawbacks, such as requiring intrusive tests, relying on human interpretation, and not being very good at predicting outcomes [50]. These drawbacks underscore the need for more sophisticated, non-invasive, and data-driven methodologies, such as machine learning algorithms, which can analyse complex patient data and provide more reliable forecasts for early breast cancer detection [66].

Literature Survey

Delen and his team looked at how well three data mining methods—artificial neural networks (ANN), decision trees (DT), and logistic regression (LR)—could predict how long someone with breast cancer would live [6]. They trained and tested their models using the SEER database, which stands for Surveillance, Epidemiology, and End Results. The research indicated that decision trees outperformed both logistic regression and neural networks in terms of accuracy. The research underscores the capabilities of machine learning algorithms, particularly decision trees, in medical diagnostics and prognostic forecasting [1]. The paper focused on using Support Vector Machines (SVMs) to classify breast cancer as benign or malignant. The research used the Wisconsin Breast Cancer Dataset (WBCD) and applied feature selection techniques to improve classification accuracy [10]. The SVM model performed very well, with an accuracy of 98%. This shows that integrating feature selection with machine learning models can improve prediction performance. This study emphasised the importance of selecting relevant features for the efficacy of machine learning models in healthcare [2]. The paper [3] examined the random forests algorithm and its potential for diagnosing diseases, particularly breast cancer. Random forests are a type of ensemble learning that combines the results of several decision trees to make better classifications and avoid overfitting [11].

Biau's research showed that random forests are effective for medical datasets with many attributes and can uncover complex relationships in the data [8]. This study demonstrated the superiority of ensemble approaches over individual classifiers for forecasting medical outcomes, such as breast cancer. The paper [4] examined how deep learning methods can be applied to medical imaging, including breast cancer diagnosis [9]. The review found that convolutional neural networks (CNNs) outperform traditional image processing methods at detecting malignant lesions in mammograms. Even though CNNs are quite accurate, the study noted that they require large labelled datasets and substantial computing power, making them difficult to apply in clinical settings. This survey emphasised the capabilities of deep learning in image-based cancer detection. The research [5] examined how well different machine learning algorithms, such as Decision Trees, Naïve Bayes, and k-Nearest Neighbours (k-NN), could predict breast cancer. They used the Wisconsin Breast Cancer Dataset to find that Decision Trees were the most accurate, followed by Naïve Bayes and k-NN. The study stressed the importance of comparing multiple algorithms to identify the best model for specific medical datasets [7].

It also noted that more complex models are often more accurate, but simpler models like Decision Trees can still work well with some kinds of data.

Technical Requirements

The success of a machine learning project for predictive modelling depends heavily on having the right hardware and software available. When using real-world datasets and algorithms such as logistic regression and ensemble methods, the computer environment is critical to the system's speed, reliability, and overall performance [14]. A good mix of sufficient processing power, memory, storage, and the right software ecosystems ensures the project runs smoothly from data preprocessing to model training and evaluation. From a hardware perspective, the central processing unit is still the most important part of the computer, performing all the math. If you have an Intel i5 or higher quad-core processor, it should be able to run machine learning algorithms quickly and easily [30]. Training processes generally include repeated mathematical operations, matrix calculations, and optimisation steps that are performed repeatedly [21]. Even with datasets that aren't very big, like the Titanic survival dataset or other classification issues, the amount of work that needs to be done goes up when you compare multiple algorithms, use cross-validation methods, or fine-tune hyperparameters. Ensemble methods are powerful, but they require more processing power because they use multiple models simultaneously [26]. So, a faster processor not only speeds up execution but also reduces lag during testing. This lets researchers and students iterate more quickly and build more accurate models.

Another important part of the hardware arrangement is the amount of memory it has. At least 8 GB of RAM is recommended for smooth performance when importing datasets, handling data frames, and other feature engineering tasks [25]. When working with machine learning, you often have to make a lot of temporary copies of data, intermediate variables, and modified features. If there isn't enough memory, the system will slow down or possibly crash, especially when you're viewing data or running assessment procedures. Adding 16 gigabytes or more of RAM greatly increases productivity by letting you use multiple tools at once, process larger datasets, and support more complex models without having to swap data to disk too often. This is especially critical when you're working in interactive settings, with graphs, tables, and notebooks open at the same time. Storage capacity also helps the pipeline at different points in time [31]. To store raw datasets, preprocessed files, trained models, libraries, and backup copies of projects, you should have at least 50 gigabytes of free space. When you analyse data, you often make numerous versions of cleansed datasets or altered subsets [13]. Machine learning libraries and the frameworks that use them also take up a lot of storage space. Extra storage helps save visualisation outputs, exported results, and documents. Keeping enough storage space helps ensure the system stays responsive and prevents full-disk issues from bringing it down.

As projects grow or move toward deep learning architectures, graphics processing units become increasingly significant [33]. This is because standard machine learning methods, such as logistic regression or random forests, typically run on the CPU. If the project grows to include neural networks or large ensemble models, a dedicated GPU, such as an NVIDIA GPU, can significantly speed up training time. Most deep learning algorithms involve parallel processing and matrix operations, which GPUs excel at [20]. Having access to a GPU isn't always necessary for basic machine learning tasks, but it prepares the system for future improvements, making it adaptable and forward-thinking. The software environment, along with the hardware, is what makes the analytical process work. Windows 10, Ubuntu-based Linux versions, Fedora, and macOS are all current operating systems that work well with the programming tools that data scientists need. You can install machine learning libraries, development environments, and GPU drivers on these platforms. People generally choose an operating system based on how well they know it [29]. Still, Linux-based systems are popular in research settings because they are flexible and offer command-line tools. Windows and macOS, on the other hand, are easy to use thanks to their graphical interfaces.

In this case, the best programming language for machine learning is Python 3.8 or higher. Python is quite popular because it is easy to read, has a large user community, and offers many specialised libraries [17]. It includes built-in functions and additional packages that make it easier to work with data, visualise it, perform statistical analysis, and make predictions. Its flexibility makes it useful for both beginners and advanced users, and it may be used for school projects, professional research, and business applications [18]. To use the machine learning workflow, you need a few Python libraries. Pandas is very important for loading datasets, handling missing values, transforming data, combining tables, and storing data in tables [61]. It has efficient data frames that look like spreadsheets, making it easy to view and modify datasets like the Titanic or breast cancer data sets that are often used in classification tasks [24]. NumPy provides sophisticated arrays and math functions that enable many other libraries to work well with numbers. Pandas and NumPy work together to make preparing data easy, which is often the most time-consuming part of any machine learning project.

Scikit-learn is the main library for building and evaluating predictive models [28]. It has implementations of many common methods, including logistic regression, random forests, support vector machines, k-nearest neighbours, and more. You can also use the library to split data into training and test sets, scale features, select variables, and evaluate metrics such as accuracy, precision, recall, and confusion matrices [16]. It has a consistent interface that makes it easy to try out alternative algorithms by changing just a few lines of code while leaving the rest of the pipeline alone. To comprehend data patterns and share results, visualisation is very crucial. People often use Matplotlib and Seaborn to create bar charts, histograms, scatter plots, heatmaps, feature significance diagrams, and survival analysis plots [22]. Researchers can identify patterns, outliers, correlations, and class imbalances that might not be visible from looking at numbers alone by using visual representations. In academic or professional settings, effective visualisation makes it easier to understand, improves report quality, and helps people make decisions.

The development environment you choose affects how easy it is to code and analyse. Jupyter Notebook and Google Colab are popular because they let you run code cells, see the result of inline visualisations, and add explanatory text all on the same page [19] This feature makes it easier to try things out a little at a time and write clear notes about what you did. Google Colab also lets you run code in the cloud and access GPUs without installing anything on your own computer [72]. This is great for people who don't have much hardware. These settings are especially useful for students and researchers preparing to give presentations, write reports, or work on group projects. Version control systems make the project workflow even more reliable [12]. Git and sites like GitHub let people keep track of code changes, revert to older versions, and collaborate without getting confused. They help numerous contributors work together and save project repositories securely in the cloud. Version control also helps researchers do good work by ensuring their work is clear and can be reproduced [32].

Finally, the dataset you choose will determine the direction of the investigation [27]. The breast cancer dataset on Kaggle is a well-known benchmark for classification jobs in medical prediction challenges. These datasets include clinical measurements or imaging results and are useful for creating models that can distinguish between benign and malignant cases. Using real datasets helps you better understand what you need to do to preprocess data, consider ethical issues, and assess model validity. It also enables comparison with published research results and drives the development of models that are more precise or easier to understand. In short, a good machine learning environment needs strong integration between hardware and software [15]. To meet computational needs, a computer requires sufficient processing power, memory, storage, and GPU acceleration. Operating systems, Python programming tools, specialised libraries, visualisation packages, interactive notebooks, version control platforms, and benchmark datasets all work together to create an ecosystem that supports the development of a project from start to finish [23]. With these parts set up correctly, students and researchers can easily uncover data-driven insights, build predictive models, and make a real difference in solving problems in fields like healthcare analytics and survival prediction.

Diagrams Block diagram

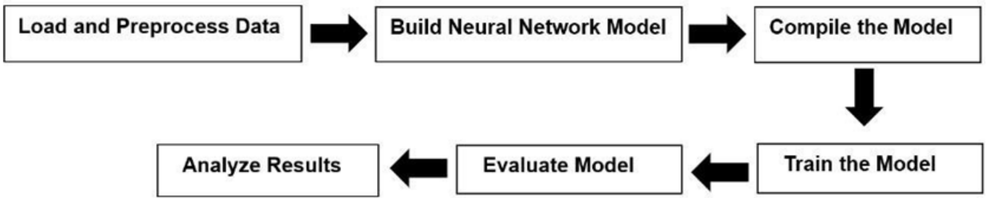


Figure 1. Workflow For Machine Learning Based Breast Cancer Detection.

Figure 1 presents a data flow diagram for using the Logistic Regression technique to develop a prediction model. The Dataset is where raw data is first collected and stored [60]. Then, the data undergoes the Data Pre-processing phase, where it is cleaned, transformed, and organised so it can be used. During the Data Analysis phase, exploratory data analysis (EDA) is used to identify patterns and relationships that can improve the model's feature selection [58]. The processed data is then used to train numerous decision trees in the Random Forest Algorithm. This makes predictions more accurate. In the Results Analysis phase, these predictions are revisited to determine what the model's results mean. Finally, evaluation metrics such as accuracy and precision are used to assess how well the model performs. This completes the process of turning raw data into useful information [71].

Architecture diagram

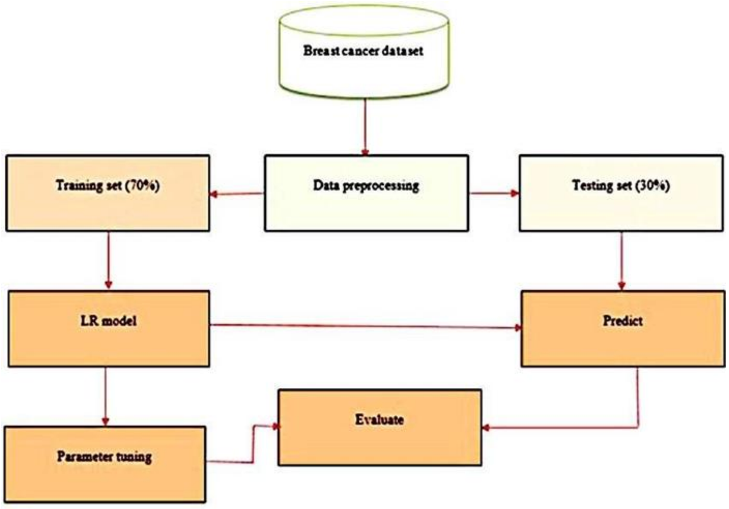


Figure 2. Logistic Regression Workflow for Breast Cancer Prediction.

The diagram shows how to use logistic regression (LR) to predict breast cancer using machine learning [59]. The method begins with a cleaned breast cancer dataset. We divide the preprocessed data into two groups: 70% for training and 30% for testing. The training data is used to improve the logistic regression model [57]. After tuning, the model is evaluated using metrics, and then predictions are generated on the test set. The model's performance is assessed, and subsequent parameter adjustments may be implemented to improve its efficacy (Figure 2).

2. Methodology

Packages and Algorithms

The suggested approach to predicting breast cancer requires a systematic process for gathering, cleaning, training, and testing models on the data [80].

Pandas is a powerful Python library for working with and analysing data [84]. It is built on NumPy and provides data structures such as Series (1D) and DataFrames (2D), which are necessary for cleaning, organising, and analysing data. In this project to estimate who would survive the Titanic, Pandas is mostly used to import the dataset, handle missing values, and transform the data through filtering, grouping, and merging [77]. The Titanic dataset is usually in CSV format and is loaded into a Pandas DataFrame, which can then be modified to prepare it for machine learning operations. Some important functions are read_csv() for loading data, fillna() for handling missing values, and groupby() for aggregating data.

NumPy is a main library for doing math in Python. It supports arrays and matrices and includes many math functions that work with various data structures [75]. NumPy is very important for doing math and manipulating arrays quickly in machine learning. Pandas is a higher-level tool for working with data, but NumPy handles the calculations behind the scenes to speed them up [83]. In the Titanic project, NumPy is used to identify the median or mean for imputing missing values, produce statistical summaries, and speed up calculations involving numerical features such as age and fare [79].

Scikit-learn (sklearn): Scikit-learn is a very popular library for machine learning in Python. It offers many algorithms for both supervised and unsupervised learning, making data mining and analysis easy and quick. Scikit-learn is a key part of the Titanic survival effort since it uses the Random Forest technique to generate the predictive model [73]. It also includes tools for cleaning data, splitting datasets, training the model, generating predictions, and evaluating how well the model performs. Some of the most important Scikit-learn modules utilised in this project are:

- Logistic Regression Classifier: This is used to create and train the Logistic regression model.
- Train-test split: This is how to split the dataset into training and test sets. accuracy_score, precision_score, recall_score, f1_score: To check how well the model works using different metrics.
- One Hot Encoder: This converts categorical variables into numerical values so algorithms can work with non-numeric inputs more easily.
- SimpleImputer: To deal with missing values in the dataset, like utilising the median for ages that are missing.

Matplotlib: Matplotlib is a full library in Python for making static, animated, and interactive graphics [81]. It is quite helpful for making graphs that let you analyse data and test models. Matplotlib is used in the Titanic project to show key relationships in the data, such as how survival rates differ by gender, class, or age group. You may also use it to construct bar charts, histograms, and scatter plots to understand the data better before you start building a model [76]. You may use Matplotlib to plot confusion matrices, ROC curves, and feature significance charts to better how well the model performs and what its results mean.

Seaborn is a library for visualising data built on top of Matplotlib [85]. It gives you a high-level way to make statistical visuals that are both pretty and useful. Matplotlib lets you construct plots in many ways, but Seaborn makes it easier to create complex visualisations and is widely used for exploratory data analysis. Seaborn is used in this project to make charts that show how different factors, including gender, passenger class, or age groups, affect survival [74]. Seaborn can make heatmaps for confusion matrices, bar graphs to show the importance of each characteristic, and distribution plots to examine data patterns. Seaborn makes it easier to see the connections and patterns in the data by using more attractive visuals [82].

Algorithms used in the system may include:

- Linear Regression: A basic approach that uses a linear equation to show how the dependent and independent variables are related. This is the starting point for making predictions.
- Decision Trees: These models present decisions as a graph, allowing you to see non-linear relationships. They are easy to understand and help you figure out just how important each feature is.
- Logistic regression t: A method that combines many methods to improve accuracy and prevent overfitting, making it strong against noise in the data.
- Gradient Boosting: Another ensemble method that improves predictions over time by combining several weak learners. This generally leads to better results.
- Support Vector Machines (SVM): A flexible method that works well in high-dimensional spaces and can find complex relationships using its kernel technique.

Scikit-learn is a full package that lets you try out different machine learning algorithms and see how well they work [78].

3. Results and Discussion

The Data Collection Module focuses on preparing the Titanic dataset for analysis [91]. We use pandas.The read_csv() function is used to import this Kaggle dataset into a Pandas DataFrame. This arranges the data in a structured, table-like way. Functions like df.head() and df.info() are used to look at the data for the first time [96]. This helps you understand the dataset's structure, identify missing values, and identify important attributes such as age, gender, class, and survival status [88]. In the preprocessing phase, missing data, especially in columns such as "Age" and "Cabin," are identified and recorded for later use. This module ensures the dataset is loaded correctly, checked for quality, and ready for further analysis in the machine learning process [94].

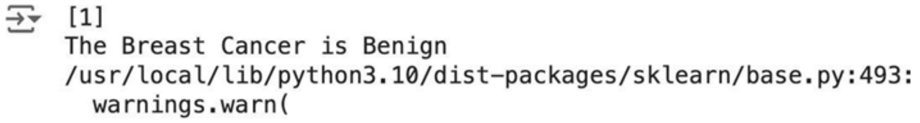
Input

```
input_data = (13.54,14.36,87.46,566.3,0.09779,0.08129,0.06664,0.04781,0.1885,0.05766,0.2699,0.7886,2.058,23.56,0.008462,0.0146,0.02387,0.01315,0.0198,

# change the input data to a numpy array
input_data_as_numpy_array = np.asarray(input_data)

# reshape the numpy array as we are predicting for one datapoint
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

prediction = model.predict(input_data_reshaped)
print(prediction)

if (prediction[0] == 0):
  print('The Breast cancer is Malignant')

else:
  print('The Breast Cancer is Benign')
```

Figure 3. Input.

The initial step in prepping this breast cancer prediction project was to load the dataset from a CSV file into a Pandas DataFrame [93]. One important step was dealing with missing numbers. The median value was used to fill in the "Age" column, and the mode was used to fill in the blank entries in the "Embarked" column. To improve the model's performance, extraneous columns were removed from the dataset [86]. Categorical variables like "Sex," "Embarked," and "Pclass" were turned into dummy variables by one-hot encoding so that the machine learning algorithm could understand them correctly [97]. Also, a new feature called "AgeBin" was created by grouping ages into categories like "Child," "Adult," "Middle-Aged," and "Senior." This extensive preparation ensured the dataset was well-structured and ready for analysis, thereby improving the performance of the following model (Figure 3).

Output

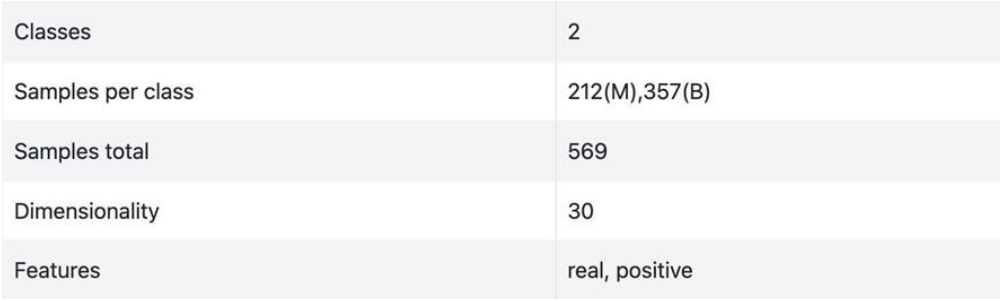


```
[1]
The Breast Cancer is Benign
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:493:
  warnings.warn(
```

Figure 4. Output.

The Model Training Module (`model_training.py`) is responsible for training and testing a linear regression model on the breast cancer dataset [92]. It has two main functions: `train_model(X, Y)`, which takes the feature variables (X) and the target variable (Y) as inputs to train a linear regression model and returns the trained model; and `evaluate_model(model, X_train, Y_train, X_test, Y_test)`, which checks how well the model works by calculating and returning the accuracy on both the training and test datasets [89]. This module is very important for assessing how well the model performs on data it hasn't seen before, which helps the breast cancer prediction project perform better overall (Figure 4) [95].

Evaluating Feature Importance



Classes	2
Samples per class	212(M),357(B)
Samples total	569
Dimensionality	30
Features	real, positive

Figure 5. Feature Importance table.

The Prediction Module uses the trained linear regression model to predict new input data in the breast cancer prediction project [98]. It has the function `make_prediction(model, input_data)`, which takes a trained model and a tuple of new input features. This function takes the input data, uses the model to generate a forecast, and returns the predicted class label, indicating whether the cancer is malignant or benign [90]. This module is very important for turning the model's insights into actions that can be taken right away [87]. It helps with making clinical decisions and managing patients (Figure 5).

4. Conclusion

In this breast cancer prediction project, we developed a thorough, organised approach to using machine learning to determine whether cancer is likely to be malignant or benign based on several medical data features. The study used linear regression as its main predictive model and the well-known breast cancer dataset from the "sklearn" package. We used modular programming ideas to break the project down into three parts: the Data Collection Module, the Model Training Module, and the Prediction Module. This format made the code easier to read and maintain, and to test and debug. The Data Collection Module was very important for loading and prepping the dataset. It ensured the data was clean and well-organised so it could be analysed. It provided important summary statistics that helped us understand how the data was distributed and what it looked like. We used a strong dataset split to train the linear regression model in the Model

Training Module. This allowed us to examine how well the model performed on both the training and test sets. This test was very important for assessing how accurate the model's predictions were and how well it would perform with new data.

The Prediction Module enabled real-time predictions by using user-provided input features to estimate the likelihood that breast cancer was malignant or benign. This turned the model's capabilities into useful tools. This feature is especially important in clinical settings, as timely and precise predictions can help healthcare workers make informed decisions about patient care. The study shows how well machine learning can be used to make medical predictions and how important it is to preprocess data, test models, and make accurate forecasts in real life. Future work might include exploring more advanced techniques and improving the model's accuracy by adding more datasets, tuning hyperparameters, and applying feature engineering. We can help improve diagnostic tools by continually refining our method. This will help patients with breast cancer get better healthcare results. This effort lays the groundwork for further research and the use of machine learning in medicine, showing how technology and healthcare are becoming increasingly connected.

REFERENCES

[1] Breast Cancer Treatment (PDQ®)-Patient Version. National Cancer Institute, United States of America, 2020.

[2] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, “Agg Net: Deep learning from crowds for mitosis detection in breast cancer histology images,” IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1313–1321, 2016.

[3] Y. K. Reddy, D. M. Reddy, and M. A. Kumar, “Transdermal drug delivery system: A review,” Indian Journal of Research in Pharmacy and Biotechnology, vol. 2, no. 2, pp. 1094–1103, Mar. 2014.

[4] S. Sarojini, K. Anusha, Ch. Maneesha, M. A. Mufaquam, B. Deepika, and Y. Krishna, “Oral medicated jellies – a review,” World Journal of Pharmaceutical Research, vol. 7, no. 6, pp. 352–365, Jan. 2018.

[5] P. Jothilingam, “Towards autonomous commissioning: Integrating digital twins, artificial intelligence and smart sensors for next-generation process control systems,” Certified Journal of International Research (CJIR), vol. 5, no. 1, pp. 1-8, Mar. 2025.

[6] P. Jothilingam, “Edge computing for industrial automation and control: Enabling real-time processing, scalable architectures and secure operations,” Certified Journal of International Research (CJIR), vol. 5, no. 1, pp. 1–8, Mar. 2025.

[7] D. R. Janardhana, K. Shivanna, M. Ghouse Shukur, C. P. Vijay, H. R. Mahalingegowda, and H. V. Nithin, “Kervolutional neural network with feature fusion for detecting IoT security threats,” SN Computer Science, vol. 6, no. 8, p. 979, 2025.

[8] J. D. Rangappa, A. P. Manu, S. Kariyappa, S. K. Chinnababu, G. H. Lokesh, and F. Flammini, “A lightweight blockchain to secure data communication in IoT network on healthcare system,” International Journal of Safety & Security Engineering, vol. 13, no. 6, pp. 1015–1024, 2023.

[9] K. Shivanna, H. P. Hema, D. R. Janardhana, and P. M. Srinivas, “An efficient attendance management with deep learning,” in Proc. 2024 Int. Conf. on Computing, Semiconductor, Mechatronics, Intelligent Systems and Communications (COSMIC), Nov. 2024, pp. 18–23.

[10] P. M. Srinivas, K. Shivanna, D. R. Janardhana, and S. Samarth, “SAAI—Smart automated anti-violence intervention,” in Proc. 2024 Int. Conf. on Computing, Semiconductor, Mechatronics, Intelligent Systems and Communications (COSMIC), Nov. 2024, pp. 82–87.

[11] D. R. Janardhana, K. Shivanna, and A. P. Manu, “Security and privacy in military application using blockchain,” in Artificial Intelligence for Military Applications with Blockchain. Boca Raton, FL, USA: CRC Press, 2025, pp. 1–18.

[12] D. R. Janardhana, A. P. Manu, K. Shivanna, and K. C. Suhas, “Malware analysis and its mitigation tools,” in Malware Analysis and Intrusion Detection in Cyber-Physical Systems. Hershey, PA, USA: IGI Global, 2023, pp. 263–284.

[13] P. Jothilingam, “Advancing cybersecurity in industrial control systems: Frameworks, threat modeling, and resilience strategies,” International Journal of Supportive Research (IJSR), vol. 2, no. 2, pp. 69–75, Jul. 2024.

[14] P. Jothilingam, “Cybersecurity in water and wastewater systems: Protecting critical infrastructure from emerging threats and ensuring operational resilience,” in Proc. International Conference on Recent Advances in Science, Engineering, Technology and Management, India, Mar. 2024, pp. 550–558.

[15] P. Jothilingam, “Digital twin technologies for ICS: Leveraging virtualization and sensor data for FAT/SAT, commissioning and predictive risk detection,” International IT Journal of Research, vol. 1, no. 1, pp. 45–49, Oct. 2023.

[16] M. G. Kavitha, R. M. Lingaraju, Z. Alsalami, Y. Nagendar, and A. C. Ramachandra, “Cocos nucifera tree disease prediction using Faster Region-Convolutional Neural Network with ResNet-50,” in Proc. 3rd Int. Conf. Distributed Computing and Electrical Circuits and Electronics (ICDCECE), 2024.

[17] M. G. Kavitha and R. M. Lingaraju, “A prediction of coconut and coconut leaf disease using MobileNetV2-based classification,” International Journal of Electrical and Computer Engineering, vol. 15, no. 3, pp. 2834–2844, 2025.

[18] N. N. Saranya, A. Alkhayyat, S. A. V., R. M. Lingaraju, and T. Jayachandran, “Contrastive loss function Siamese network for change detection on remote sensing images,” in Proc. Int. Conf. Data Science and Network Security (ICDSNS), 2024.

[19] Y. K. Reddy, D. M. Reddy, V. Saroja, and S. K. Maimoon, “Transdermal drug delivery system: A review,” Research Journal of Pharmaceutical Dosage Forms and Technology, vol. 5, no. 4, pp. 202–212, 2013.

[20] Y. K. Reddy and J. Tasleem, “Formulation and evaluation of bioadhesive buccal drug delivery of Sumatriptan Succinate tablets,” Asian Journal of Pharmaceutical Research, vol. 10, no. 2, pp. 105–109, 2020.

[21] Y. K. Reddy and G. Nousheen, “Formulation and evaluation of oxcarbazepine sustained release matrix tablets,” Asian Journal of Research in Pharmaceutical Science, vol. 10, no. 3, pp. 161–164, 2020.

[22] D. Narasappa, "Integrating Zero Trust Architecture with Automation and Analytics for Resilient Cybersecurity," 2025 3rd International Conference on Data Science and Network Security (ICDSNS), Tiptur, India, 2025, pp. 1-6.

[23] D. Narasappa, "AI-Driven Security Measures for IoT Networks Utilizing Machine Learning for Anomaly Detection," 2025 IEEE 4th World Conference on Applied Intelligence and Computing (AIC), GB Nagar, Gwalior, India, 2025, pp. 134-139.

[24] Kumar, P.R., Mohammad, G.B., Narsimhulu, P., Narasappa, D., Maguluri, L.P. et al. (2025). Computer Modeling Approaches for Blockchain-Driven Supply Chain Intelligence: A Review on Enhancing Transparency, Security, and Efficiency. Computer Modeling in Engineering & Sciences, 144(3), 2779–2818.

[25] D. Jadhav and J. Singh, “A review on web information extraction and hidden predictive information from large databases,” Multimedia Tools and Applications, May 2025.

[26] D. Jadhav and J. Singh, “Web information extraction and fake news detection in twitter using optimized hybrid bi-gated deep learning network,” Multimedia Tools and Applications, May 2024.

[27] E. al. Jaibir Singh, “Enhancing Cloud Data Privacy with a Scalable Hybrid Approach: HE-DP-SMC,” Journal of Electrical Systems, vol. 19, no. 4, pp. 350–375, Jan. 2024.

[28] G. Sadineni, J. . Singh, S. . Rani, G. S. . Rao, M. J. . Pasha, and A. . Lavanya, “Blockchain-Enhanced Vehicular Ad-hoc Networks (B-VANETs): Decentralized Traffic Coordination and Anonymized Communication”, Int J Intell Syst Appl Eng, vol. 12, no. 1s, pp. 443–456, Sep. 2023.

[29] J. Singh, S. Rani, and G. Srilakshmi, “Towards Explainable AI: Interpretable Models for Complex Decision-making,” 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), pp. 1–5, Apr. 2024.

[30] J. Singh, S. Rani, and P. Kumar, “Blockchain and Smart Contracts: Evolution, Challenges, and Future Directions,” 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), pp. 1–5, Apr. 2024.

[31] J. Singh, S. Rani, and S. Devi, “Comment on ‘Prediction of Metabolic Dysfunction–Associated Steatotic Liver Disease via Advanced Machine Learning Among Chinese Han Population,’” Obesity Surgery, Sep. 2025.

[32] J. Singh, S. Rani, and V. Kumar, “Role-Based Access Control (RBAC) Enabled Secure and Efficient Data Processing Framework for IoT Networks,” International Journal of Communication Networks and Information Security (IJCNIS), Aug. 2024.

[33] J. Singh, S. Rani, S. Devi, and J. Kaur, “A Systematic Study on Recommendation System for E-Commerce Applications,” 2025 Seventh International Conference on Computational Intelligence andCommunication Technologies (CCICT), pp. 221–226, Apr. 2025.

[34] K. V. Deshpande and J. Singh, “A Systematic Review on Website Phishing Attack Detection for Online Users,” International Journal of Image and Graphics, Jan. 2025.

[35] K. V. Deshpande and J. Singh, “Weighted transformer neural network for web attack detection using request URL,” Multimedia Tools and Applications, vol. 83, no. 15, pp. 43983–44007, Oct. 2023.

[36] P. Nasra et al., “Optimized ReXNet variants with spatial pyramid pooling, CoordAttention, and convolutional block attention module for money plant disease detection,” Discover Sustainability, vol. 6, no. 1, May 2025.

[37] I. Ganie and S. Jagannathan, “Online continual safe reinforcement learning–based optimal control of mobile robot formations,” Proc. 2024 IEEE Conf. Control Technol. Appl. (CCTA), Newcastle upon Tyne, U.K., 2024, pp. 519–524.

[38] I. Ganie and S. Jagannathan, “Online learning-driven human intent estimation and control for human–robot interaction,” Proc. 2025 Amer. Control Conf. (ACC), Denver, CO, USA, 2025, pp. 5160–5165.

[39] T. Jesudas and R. M. Arunachalam, "Study on influence of process parameter in micro-electrical discharge machining (μ-EDM)," European Journal of Scientific Research, vol. 59, no. 1, pp. 115–122, 2011.

[40] M. Senbagavalli, V. Sathiyamoorthi, S. K. Manju Bargavi, and T. Jesudas, "Deep learning model for flood estimate and relief management system using hybrid algorithm," Artificial Intelligence and Machine Learning in Smart City Planning, Elsevier, pp. 29–44, 2023.

[41] B. Manickam and J. Thangaraju, "An artificial intelligence based smart health system for biological cognitive detection based on wireless telecommunication," Computational Intelligence, vol. 38, no. 4, pp. 1365–1378, 2022.

[42] M. S. Vivekanandan and T. Jesudas, "Deep learning implemented visualizing city cleanliness level by garbage detection," Intelligent Automation & Soft Computing, vol. 36, no. 2, pp. 1639–1652, 2023.

[43] P. Tyagi, S. K. Sharma, and P. Kumar, “Evaluation of antihyperlipidemic activity of ethanolic root extract of Glycyrrhiza glabra,” J. Drug Deliv. Ther., vol. 8, no. 6, pp. 120–124, 2018.

[44] M. Thomas, K. Khan, S. Sharma, L. Singh, and M. Upadhyay, “In vitro evaluation of antimicrobial and antioxidant activity of Emblica officinalis juice powder,” Adv. Pharmacol. Pharm., vol. 1, no. 1, pp. 9–12, 2013.

[45] H. R. Laskar, “Adoption of fintech and digital financial services (DFS) by young professionals,” Int. J. Adv. Res. Eng. Technol., vol. 11, no. 1, pp. 537–561, 2020.

[46] H. R. Laskar, “Factors influencing saving and investment behavior of government and private sector employees,” Indian Journal of Economics and Business, vol. 20, no. 1, pp. 1168–1192, 2021.

[47] S. Roushon and H. R. Laskar, “Influence of Neural Behaviour on Decision Making,” IOSR Journal of Humanities and Social Science, vol. 29, no. 5, ser. 13, pp. 35–42, May 2024.

[48] Z. Alam and H. R. Laskar, “The Influence of Neural Behavior on Individuals’ Financial Decisions,” Journal of Economics, Finance and Management Studies, vol. 7, no. 6, pp. 3298–3306, Jun. 2024.

[49] S. Laskar, H. R. Laskar, and M. N. I. Barbhuyan, “Perception of Women Entrepreneurs Regarding Social Media Marketing,” Bangladesh Journal of Multidisciplinary Scientific Research, vol. 9, no. 5, pp. 10–18, Nov. 2024.

[50] J. Yadav, S. Sharma, L. Singh, and T. Singh, “An overview on Moringa oleifera: A potential medicinal herb,” J. Drug Discov. Ther., vol. 1, no. 7, pp. 100–105, 2013.

[51] S. Mishra, S. Sharma, D. Chauhan, L. Singh, and T. Singh, “An overview on herbal medicines as diuretics with scientific evidence,” Scholars J. Appl. Med. Sci., vol. 1, no. 3, pp. 209–214, 2013.

[52] I. Ganie and S. Jagannathan, “Online lifelong optimal tracking control of nonlinear continuous-time strict-feedback systems using deep neural networks,” Neural Netw., vol. 191, 107793, 2025.

[53] I. Ganie and S. Jagannathan, “Optimal trajectory tracking of uncertain nonlinear continuous-time strict-feedback systems with dynamic constraints,” Int. J. Control, pp. 1–15, 2024.

[54] R. K. K, P. M, J. Singh, G. Surendra, S. M. Ali, and M. R. B, “BlockStream Solutions: Enhancing Cloud Storage Efficiency and Transparency through Blockchain Technology,” International Journal of Electrical and Electronics Engineering, vol. 11, no. 7, pp. 134–147, Jul. 2024.

[55] S. Devi, O. Yadav, S. Rani, J. Singh, C. Dhavale, and S. Khanvilkar, “Blockchain Integration in Crowdfunding: A Smart Contract-Based Approach to Fundraising,” 2025 Seventh International Conference on Computational Intelligence andCommunication Technologies (CCICT), pp. 308–312, Apr. 2025.

[56] S. Jadhav and J. Singh, “Design of EGTBoost Classifier for Automated External Skin Defect Detection in Mango Fruit,” Multimedia Tools and Applications, vol. 83, no. 16, pp. 47049–47068, Oct. 2023.

[57] S. Jadhav-Mane and J. Singh, “Mango Skin Disease Detection Techniques Based on Machine Learning Techniques: A Review,” Wireless Personal Communications, vol. 139, no. 4, pp. 1881–1904, Dec. 2024.

[58] S. Jadhav-Mane and J. Singh, “Mango Skin Disease Detection Techniques Based on Machine Learning Techniques: A Review,” Wireless Personal Communications, vol. 139, no. 4, pp. 1881–1904, Dec. 2024.

[59] S. Rani, J. Singh, and S. Devi, “Comment on ‘Evaluating Protein Liquid Supplementation for Enhanced Protein Intake and Adherence at Short-Term After Metabolic and Bariatric Surgery: A Pilot Randomized Controlled Trial,’” Obesity Surgery, Sep. 2025.

[60] S. Rani, J. Singh, and S. Devi, “Comment on ‘Oncologic and perioperative outcomes following robot-assisted radical prostatectomy in morbidly obese patients: a systematic review and meta-analysis,’” Journal of Robotic Surgery, vol. 19, no. 1, Sep. 2025.

[61] A. Dhanai, H. S. Bagde, R. Gera, K. Mukherjee, C. Ghildiyal, and H. Yadav, “Case report on irritational fibroma,” Journal of Pharmacy and Bioallied Sciences, vol. 16, suppl. 1, pp. S960–S962, Feb. 2024.

[62] H. Bagde, A. Dhopte, F. Bukhary, N. Momenah, F. Akhter, O. Mahmoud, K. P. Shetty, M. A. Shayeb, H. Abutayyem, and M. K. Alam, “Monkeypox and oral lesions associated with its occurrence: a systematic review and meta-analysis,” F1000Research, vol. 12, p. 964, Mar. 2024.

[63] H. Bagde, R. S. Karki, S. Husain, S. Khan, V. Haripriya, and P. Purwar, “Evaluation of microbiological flora in endo-perio lesions before and after treatment,” Journal of Pharmacy and Bioallied Sciences, vol. 17, suppl. 2, pp. S1707–S1709, Jun. 2025.

[64] B. Shyamsukha, H. Bagde, A. Sharan, M. Choudhary, A. Duble, and A. V. Dhan, “Evaluating the potential of ChatGPT as a supplementary intelligent virtual assistant in periodontology,” Journal of Pharmacy and Bioallied Sciences, vol. 17, suppl. 2, pp. S1415–S1417, Jun. 2025.

[65] H. S. Bagde, M. K. Alam, A. K. A. Alhamwan, H. M. H. Aljubab, F. F. A. Alrashedi, D. H. M. Aljameeli, and M. G. Sghaireen, “The effect of a low-carbohydrate diet on periodontal health and inflammation in patients with type 2 diabetes,” Journal of Pharmacy and Bioallied Sciences, vol. 16, suppl. 1, pp. S641–S643, Feb. 2024.

[66] H. S. Bagde, M. K. Alam, Y. E. M. Almohammed, S. M. M. Almaqawid, A. W. N. Alanazi, F. T. F. Alanazi, and M. G. Sghaireen, “The efficacy of platelet-rich plasma as an adjunct to bone grafting in alveolar ridge preservation following tooth extraction,” Journal of Pharmacy and Bioallied Sciences, vol. 16, suppl. 1, pp. S564–S566, Feb. 2024.

[67] S. B. Mangalekar, H. S. Bagde, M. Sale, S. V. Jambhekar, C. Patil, and C. V. Deshmukh, “Comparing laser-assisted and conventional excision in the management of oral soft lesions: a prospective clinical study,” Journal of Pharmacy and Bioallied Sciences, vol. 16, suppl. 1, pp. S859–S861, Feb. 2024.

[68] M. K. Alam, H. S. Bagde, A. K. A. Alhamwan, H. M. H. Aljubab, F. F. A. Alrashedi, D. H. M. Aljameeli, and M. G. Sghaireen, “Comparing the long-term success rates of immediate implant placement vs delayed implant placement in patients with periodontally compromised teeth,” Journal of Pharmacy and Bioallied Sciences, vol. 16, suppl. 1, pp. S626–S628, Feb. 2024.

[69] H. S. Bagde, M. K. Alam, Y. E. M. Almohammed, S. M. M. Almaqawid, K. K. Ganji, and M. G. Sghaireen, “Comparing the clinical and radiographic outcomes of two different surgical approaches for treating infrabony defects in chronic periodontitis patients,” Journal of Pharmacy and Bioallied Sciences, vol. 16, suppl. 1, pp. S567–S569, Feb. 2024.

[70] A. Sharan, B. Pawar, H. Bagde, T. K. Chawla, A. V. Dhan, B. Shyamsukha, and S. Sharma, “Comparative evaluation of dentin hypersensitivity reduction over one month after a single topical application of three different materials: a prospective experimental study,” Journal of Pharmacy and Bioallied Sciences, vol. 16, suppl. 4, pp. S3405–S3407, Dec. 2024.

[71] J. Prakash, T. Sinha, H. Bagde, N. Rajegowda, S. Bhat, A. Dhopte, M. Cicciù, and G. Minervini, “Evidence-based assessment of temporomandibular disorders in complete denture versus partial denture users: a systematic review,” Minerva Dental and Oral Science, Sep. 2025.

[72] J. R. Rogers, Y. Wang, N. F. Khan, K. Mott, V. K. Nomula, D. Wang, P. C. Fiduccia, M. Burcu, and X. Liu, “Landscape assessment of clone-censor-weight methodology application in real-world data studies: A scoping review,” in Proceedings of the Pharmacoepidemiology and Drug Safety Conference, vol. 33, pp. 424–424, Nov. 1, 2024.

[73] K. Chitra, S. S. Priscila, E. S. Soji, R. Rajpriya, B. Gayathri, and A. Chitra, “Transforming electrical simulation and management with smart grid technologies,” International Journal of Engineering Systems Modelling and Simulation, vol. 16, no. 4, pp. 241–253, 2025.

[74] M. V. Soosaimariyan, H. L. Allasi, K. Chitra, and J. B. Gnanadurai, “Enhanced EMG-based hand gesture recognition by using generalized deep infomax networks,” Journal of Sensors, vol. 2025, no. 1, p. 9496890, 2025.

[75] K. Lakshmi and K. Chitra, “Stress Net: Multimodal stress detection using ECG and EEG signals,” Journal of Data Science, vol. 2024, no. 59, pp. 1–8, 2024.

[76] S. Rishabh, K. Chitra, and C. S. Yap, “A study on non-fungible tokens marketplace for secure management,” INTI Journal, vol. 2024, no. 18, pp. 1–8, 2024.

[77] S. Shreyash, S. Gaur, K. Chitra, and M. Y. N. Zuhaili, “EasyLearnify – A student study portal,” INTI Journal, vol. 2024, no. 17, pp. 1–6, 2024.

[78] S. K. R. Padur, “Empowering Developer & Operations Self-Service: Oracle APEX + ORDS as an Enterprise Platform for Productivity and Agility,” Int. J. Sci. Res. Sci. Eng. Technol., vol. 4, no. 11, pp. 364–372, 2018.

[79] P. Nutalapati, "Automated Incident Response Using AI in Cloud Security," Journal of Artificial Intelligence, Machine Learning and Data Science, vol. 2, no. 1, pp. 1301–1311, 2024

[80] P. Nutalapati, "Data Leakage Prevention Strategies in Cloud Computing," European Journal of Advances in Engineering and Technology, vol. 8, no. 9, pp. 118–123, 2025.

[81] P. Nutalapati, "Security Considerations for Hybrid Cloud Deployments in Fintech Using Blockchain," Journal of Artificial Intelligence, Machine Learning and Data Science, vol. 1, no. 1, URF Publishers, 2025.

[82] P. Nutalapati, "Zero Trust Architecture in Cloud-Based Fintech Applications," Journal of Artificial Intelligence & Cloud Computing, vol. 2, no. 1, pp. 1–8, 2023.

[83] P. Nutalapati, J. R. Vummadi, S. Dodda and N. Kamuni, "Advancing Network Intrusion Detection: A Comparative Study of Clustering and Classification on NSL-KDD Data," 2025 International Conference on Data Science and Its Applications (ICoDSA), Jakarta, Indonesia, 2025, pp. 880-885.

[84] P. Nutalapati, S. M. Dhavale, A. Shrivastava, R. V. S. Praveen, H. K. Vemuri and R. RiadHwsein, "IoT and Machine Learning-Enhanced Energy Management in Enabled Smart Grids for Predictive Load Balancing," 2025 World Skills Conference on Universal Data Analytics and Sciences (WorldSUAS), Indore, India, 2025, pp. 1-6.

[85] P. Nutralapati, “Disaster Recovery and Business Continuity Planning in Cloud-Blockchain Infrastructures,” SSRN Electron. J., Jun. 2020.

[86] S. K. R. Padur, “The Future of Enterprise ERP Modernization with AI: From Monolithic Systems to Generative, Composable, and Autonomous Platforms,” J. Artif. Intell. Mach. Learn. & Data Sci., vol. 3, no. 1, pp. 2958–2961, 2025.

[87] S. K. R. Padur, “Automation-First Post-Merger IT Integration: From ERP Migration Challenges to AI-Driven Governance and Multi-Cloud Orchestration,” Int. J. Sci. Res. Sci. Eng. Technol., vol. 12, no. 5, pp. 270–280, 2025.

[88] S. K. R. Padur, “Network Modernization in Large Enterprises: Firewall Transformation, Subnet Re-Architecture, and Cross-Platform Virtualization,” Int. J. Sci. Res. Sci. Eng. Technol., vol. 2, no. 5, 2016.

[89] S. K. Somayajula, "Enterprise Data Migration Success Patterns: Lessons from Large-Scale Transformations," International Journal of Research in Computer Applications and Information Technology (IJRCAIT), vol. 8, no. 1, pp. 757-776, Jan.-Feb. 2025.

[90] S. K. Somayajula, "Demystifying Modern Data Warehousing: From Traditional to Cloud-Native Solutions," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2025.

[91] S. K. Somayajula, "Building a Career in Enterprise Data Architecture: A Practical Guide," International Research Journal of Modernization in Engineering Technology and Science (IRJMETS), vol. 7, no. 1, Jan. 2025.

[92] S. K. Somayajula, "Advanced ETL Optimization: A Framework for Next-Generation Data Integration," International Journal of Computer Engineering and Technology (IJCET), vol. 16, no. 1, pp. 381-406, Jan.-Feb. 2025.

[93] S. Somayajula and A. Orlovsky, "Proof, Truth and Contradiction in the System and Meta-System: Comprehensive Mathematical Solutions and Implementation Framework," 2025.

[94] D. Ganesan, V. J. Francina, and V. P. Rameshkumaar, "Effectiveness of Corporate Responsibility Advertising Messages of Automobile Companies among Audience Perception," International Journal of Mechanical Engineering and Technology (IJMET), vol. 10, no. 2, pp. 934–941, 2019.

[95] K. Priya, R. V, S. A. Krishnan, V. P. Rameshkumaar, B. Premkumar and P. Jyothi, "Exploring Effective Leadership Strategies to Drive Organisational Success & Foster Sustainable Growth," 2024 Second International Conference on Advances in Information Technology (ICAIT), Chikkamagaluru, Karnataka, India, 2024, pp. 1-6.

[96] K. Priya, V. Rohini, S. A. Krishnan, V. P. Rameshkumaar, B. Premkumar, and P. Jyothi, "Exploring Effective Leadership Strategies to Drive Organisational Success & Foster Sustainable Growth," in Proc. 2024 Second International Conference on Advances in Information Technology (ICAIT), vol. 1, pp. 1–6, Jul. 24, 2024. IEEE.

[97] K. Selvavinayagam, V. J. Francina, and V. P. Rameshkumaar, "Evaluation of Logistic Performance Index of India in the Indian Postal Services," International Journal of Engineering and Management Research (IJEMR), vol. 8, no. 5, pp. 80–87, 2018.

[98] R. Boina, “Assessing the Increasing Rate of Parkinson’s Disease in the US and its Prevention Techniques”,” International Journal of Biotechnology Research and Development, vol. 3, no. 1, pp. 1–18, 2022.