# COMMAND-LINE CALCULATOR

### A MINI PROJECT REPORT 18CSC304J – COMPILER DESIGN

***Submitted by***

**PARTH GARG [RA2011003010095] SIDDARTHA DHAR [RA2011003010081] SAPTOSHREE NAG [RA2011003010130]**

*Under the guidance of*

## Dr. Nithyashri J

Assistant Professor, Department of Computer Science and Engineering

***in partial fulfillment for the award of the degree of***

**BACHELOR OF TECHNOLOGY**

in

## COMPUTER SCIENCE & ENGINEERING

of

**FACULTY OF ENGINEERING AND TECHNOLOGY**



S.R.M. Nagar, Kattankulathur, Chengalpattu District

### MAY 2023

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

(Under Section 3 of UGC Act, 1956)

## BONAFIDE CERTIFICATE

Certified that Mini project report titled **“ COMMAND-LINE CALCULATOR”** is the bona fide work of **PARTH GARG (RA2011003010095), SAPTOSHREE NAG(RA2011003010130), SIDDARTHA DHAR(RA2011003010081)** who carried out the

minor project under my supervision. Certified further, that to the best of my knowledge, the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

#### SIGNATURE SIGNATURE

Dr. Nithyashri J

#### Assistant Professor

Department of Computing Technologies

Dr. M. Pushpalatha

#### HEAD OF THE DEPARTMENT

Department of Computing Technologies

# ABSTRACT

Breast cancer is a common and life-threatening disease that affects women worldwide. Early detection of breast cancer plays an important role in improving patient outcomes and reducing unnecessary treatments. Titled "Prediction and Classification of Breast Cancer using Machine Learning Algorithms," the project addresses the urgent need for accurate and effective breast cancer screening. This project uses a variety of data, mostly from the Wisconsin Cancer Registry, including graphical and numerical data. Using a combination of machine learning and deep learning, including K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Random Forest Classifiers (RFC), the goal is to predict the nature of breast cancer. This dual classification aims to distinguish between benign and malignant conditions. The development of this project involved many aspects including data preprocessing, feature engineering and model selection. The phone data was processed by machine learning algorithms while the image data was analyzed by a neural network (CNN). The integration of these systems aligns them with increased power, efficiency and predictability. In addition, a web application was developed using Flask to provide an intuitive experience to practitioners and stakeholders. The interface allows users to access relevant information and makes easy, precise, real-time predictions. Delivering the project on a cloud platform provides access to a wider audience and can improve breast cancer diagnosis. The report details the objectives, process and results of the project. It reveals the complexity of development by examining the challenges encountered, the architectural framework adopted, and the module definition. Interim results and discussions are also provided, demonstrating the validity of the prediction model and the usability of the web interface. This project is a beacon of hope in the field of medicine, which requires timely and accurate cancer diagnosis. Thanks to the integration of artificial intelligence algorithms, user-friendly interface and cloud distribution, it aims to reduce the burden of unnecessary surgeries and provide doctors with a powerful diagnostic tool.

### Table of Contents

|  |  |  |
| --- | --- | --- |
| **Sl. No** | **USN** | **Page No** |
| 1 | Introduction | 1 |
| 2 | Methodology and Techniques | 3 |
| 3 | Implementation and Result | 6 |
| 4 | Conclusion | 11 |
| 5 | References | 12 |

**Literature Review**

This literature review is based upon the paper titled “*Classification of malignant and benign tissue with logistic regression*” by *Laila Khairunnahar(a), Mohammad Abdul Hasib(b), Razib Hasan Bin Rezanur(b),*

*Mohammad Rakibul Islam(b), Md Kamal Hosain(a)*

*a Department of Electronics & Telecommunication Engineering, Rajshahi University of Engineering & Technology, Rajshahi, 6204, Bangladesh*

*b Department of Electrical and Electronic Engineering, Islamic University of Technology, Gazipur, 1704, Dhaka, Bangladesh*

Computer-aided diagnosis (CAD) systems have become popular in recent years, particularly in medical imaging and diagnostic radiology. The system is designed to help healthcare professionals, especially radiologists, make more accurate diagnoses by providing second opinions and reducing operator-dependent variability in interpreting medical images.

One of the most important applications of CAD systems is the detection and diagnosis of breast cancer. The system plays an important role in helping radiologists evaluate mammograms and ultrasounds and detect breast cancer early. This literature review explores the principles and techniques of CAD systems aimed at breast cancer diagnosis.

**Significance of CAD in Medical Imaging**

CAD systems are invaluable tools in medical imaging. They improve the quality of medical images, reduce staining, and reveal important information that cannot be seen by visual inspection alone. In addition, CAD systems reduce the time and effort required by healthcare professionals by increasing diagnostic accuracy. An example of this is the use of CAD systems in breast ultrasounds, which contribute to the location of the lesion, image contrast, and overall image quality.

**CAD in Breast Cancer Detection**

Breast cancer detection by CAD systems has been developed as an important part of breast cancer diagnosis. This system improves the capabilities of radiologists, allowing them to more effectively evaluate ultrasound and diagnose breast cancer. CAD collaborative efforts with radiologists improve diagnostic accuracy by reducing the workload associated with manual translation.

Chest ultrasound CAD system usually involves several steps:

1. **Image processing**: This step focuses on improving the ultrasound image and reducing speckle without compromising important diagnostic features. Image processing is important to maintain image quality.
2. **Image Segmentation**: Image segmentation divides the ultrasound image into non-overlapping regions and effectively separates the lesion from the background. This step defines the lesion boundaries for further feature extraction.
3. **Feature extraction and selection**: Feature extraction aims to identify reliable features that accurately distinguish between lesional and non-lesional or benign and malignant regions. Given the potential complexity of the feature space, it is important to choose effective features.
4. **Classification**: In the classification stage, selected features are used to divide suspicious areas into different classes such as benign and malignant. Different machine learning techniques such as linear discriminant analysis (LDA), support vector machine (SVM) and artificial neural network (ANN) have been studied for lesion classification.

**Datasets and Their Features**

The availability of appropriate databases is very important in developing and testing CAD systems. In the prognosis of breast cancer, the database plays an important role in the training model to distinguish between malignant and benign tissues. Three main databases are often used:

**A. Wisconsin Diagnosis Breast Cancer (WDBC)**: This dataset consists of features calculated from fine needle aspirates (FNA) of breast masses. This includes properties such as radius, texture, perimeter, area, smoothness, compactness, concavity, hollow point, symmetry, and fractal dimensions. These features are calculated from digital images and describe the characteristics of cell nuclei.

**B. Wisconsin Prognosis Breast Cancer (WPBC)**: WPBC includes features such as radius, texture, perimeter, area, smoothness, compactness, concavity, concavity point, symmetry, and fractal dimensions. It also includes outcome data, time to relapse, and disease-free time.

**C. Wisconsin Breast Cancer (WBC)**: This data set includes characteristics such as thickness, heterogeneity of cell size and shape, marginal adhesion, epithelial cell size, bare nucleus, coarse chromatin, normal nucleoli, and mitosis.

This database serves as an important source of information for the development and evaluation of breast cancer prediction models.

**Feature Selection Techniques**

Feature selection is an important step in model development to reduce computational complexity and improve model performance. Two methods of feature selection are often used:

**1. Chi-Square Tes**t: Chi-Square Test is a statistical method used to determine the independence and fit of a feature. This determines whether two or more observations in two populations are correlated, helping to select characteristics.

**2. Principal component analysis (PCA)**: PCA is a mathematical procedure that transforms correlated variables into uncorrelated linear variables called principal components. It reduces the dimensionality of the feature space while preserving important information.

**Insights on Machine Learning Classifier**

Machine learning plays an integral role in CAD systems, especially in breast cancer prediction. Machine learning algorithms can identify new categories of observations based on training databases and specific category memberships. This algorithm creates a model from sample input, making predictions or decisions based on data, instead of relying solely on implicit rule-based algorithms.

Several machine learning classifiers including Naïve Bayes, Support Vector Machines (SVM), Decision Trees, Reinforcement Learning, Neural Networks, and k-Nearest Neighbors (KNN) have been studied for breast cancer prediction. Logistic Regression, known for its probabilistic interpretation and feature correlation capabilities, was used in this study.

**Preferred Use of Logistic Regression**

Logistic regression is the classification chosen for this study because of its convenient properties:

1. **Probabilistic interpretation**: Logistic regression provides a probabilistic interpretation of model output that makes it easier to understand. This description is not always available in classifiers such as Decision Tree and SVM.
2. **Trait Correlations**: Logistic regression allows for the estimation of trait correlations, which allows the identification of relationships between input attributes. Understanding trait relationships can improve the predictive power of models.

In summary, the field of CAD in breast cancer prediction has witnessed significant progress by integrating machine learning classification and feature selection methods. Database selection and classification, such as logistic regression, play an important role in developing an accurate and efficient CAD system.

**Comparison of Existing Methods with Merits and Demerits**

Breast cancer is a major challenge in women's health. With millions of cases diagnosed worldwide each year, early detection and accurate classification of breast cancer is essential for effective treatment and improved survival rates. Over the years, various methods and technologies have been developed and deployed to overcome these challenges. In this comprehensive review, we will explore the methods available for the detection and classification of breast cancer, distinguish their strengths and weaknesses, and explore how the model proposed in this project promises to improve and improve the process of detection and classification.

**The Challenge of Breast Cancer Detection and Classification**

The diagnosis and classification of breast cancer poses significant challenges for health care providers and researchers. The complexity of breast tissue requires advanced diagnostic tools and techniques that deal with small, subtle abnormalities. The ability to accurately differentiate between benign and malignant lesions is important to guide treatment decisions and improve patient outcomes.

Conventional methods for the diagnosis and classification of ovarian cancer have been used for many years, and although they have contributed to the field of oncology, they also have limitations. Let's explore some of these traditional methods and their respective features and weaknesses before looking at the potential improvements offered by the proposed model.

**Traditional Methods in Breast Cancer Detection**

**1. Mammography**

**Merits**:

* **Proven effectiveness**: Mammography is a well-established and widely used method for breast cancer screening.
* **High-resolution imaging**: provides high-resolution X-ray images, allowing for the detection of small defects.
* **Routine screening**: Mammography is a routine part of breast screening for many women.

**Demerits**:

* **Limited sensitivity**: Mammography sensitivity is reduced in women with dense breast tissue, which can lead to missed diagnoses.
* **Radiation exposure**: Exposing patients to ionizing radiation, which may raise concerns about long-term risks.
* **Subjective interpretation**: Interpretation of mammograms is subjective and operator-dependent, which can lead to variability in results.

**2.** **Ultrasound Imaging**

**Merits**:

* **Safety**: Ultrasound does not expose the patient to ionizing radiation, making it safe for routine use.
* **Additional**: Useful for assessing breast abnormalities detected by mammography.
* **Differentiation**: Ultrasound can effectively distinguish between fluid-filled cysts and solid tumors.

**Demerits**:

* **Operator dependent**: Results may vary depending on operator experience and technique.
* **Limited detection:** May have limited sensitivity in detecting small tumors, especially in dense breast tissue.
* **False positives:** Can produce false positives, requiring ultrasound and potentially biopsies.

**3. Magnetic resonance imaging (MRI)**

**Merits**:

* **Non-ionization**: MRI does not use ionizing radiation, making it safe for patients.
* **Low sensitivity**: MRI is particularly effective in screening high-risk individuals due to its high sensitivity.
* **Complete imaging**: In some cases superior to mammography, it provides a complete image of the breast tissue.

**Demerits**:

* **Cost and availability**: MRI is expensive and unavailable, limiting its widespread use.
* **Longer exams**: MRI exams can be time-consuming, affecting patient mobility.
* **False positives**: Can cause false positives, leading to incorrect biopsies and anxiety for the patient.

4. **Biopsy and Histopathological Analysis**

**Merits**:

* **Definitive diagnosis**: A biopsy provides a definitive diagnosis by examining tissue samples.
* **High Accuracy**: Offers high accuracy in damage detection and treatment planning.
* **Clinical significance**: Biopsy results directly inform treatment decisions.

**Demerits**:

* **Invasive**: A biopsy is an invasive procedure that can cause discomfort to the patient.
* **Sampling type**: biopsy specimens may miss small tumors if they are not representative of the entire lesion.
* **Time delays**: The time required to perform biopsies and analyze tissue samples can cause delays in treatment planning.

**Traditional Methods: A Balance of Merits and Demerits**

Conventional methods of diagnosing and classifying breast cancer have played an important role in saving countless lives. However, there are specific limitations that require the exploration of innovation and complementary approaches. Although mammography is widely used, it can be more effective in women with dense breast tissue, and can miss cancer in the early stages. Ultrasound offers safety, but is operator dependent and can cause false positives. MRI provides detailed imaging, but is expensive and not available to everyone. Biopsy is definitive but invasive and time consuming.

**Proposed Model for Enhancement**

This project's breast cancer detection and classification model represents a significant leap forward in overcoming the limitations of conventional methods. By utilizing the power of modern data-driven approaches and leveraging the advantages of traditional methods, this model aims to improve and improve the detection and classification process in many ways.

**1. Integration of Imaging and Numerical Data**

The model takes a comprehensive approach, combining image data with numerical features. Combining data from mammography, ultrasound, and other sources provides a more holistic view of breast tissue characteristics. In this way, the model aims to improve classification accuracy, especially in complex or ambiguous findings.

**2. Machine Learning Algorithms**

These models use machine learning algorithms to make sense of the wealth of available data. Algorithms such as logistic regression, support vector machines, and random forests are used to learn patterns and features that indicate maladaptive behavior. This algorithm excels in processing large data and extracting meaningful insights to provide more accurate predictions.

**3. Feature Selection Techniques**

It uses model feature selection techniques to pursue accuracy and efficiency. Methods such as the Chi-Square test and Principal Component Analysis are used to identify the most important and informative features for classification. This not only reduces the data dimension, but also improves the computational efficiency.

**4. Probabilistic Interpretation**

Logistic regression, chosen as a classifier in this model, provides a probabilistic interpretation of the results. This description improves the clarity and comprehensibility of the model, allowing clinicians to understand not only the classification results but also the associated confidence level. This is especially important in medical diagnosis, where understanding uncertainty is essential.

**5. Reduced Operator Dependency**

One of the main challenges with conventional methods is the reliance on operator experience. The proposed model can provide a continuous and objective evaluation after training. This reduction in operator dependence reduces the variability introduced by different operators leading to more reliable results.

**6. Non-Invasive and Safe**

Unlike some conventional methods that involve radiation exposure or invasive procedures, the proposed model is non-invasive and safe for patients. It does not expose people to ionizing radiation, making it convenient for routine examinations and reducing long-term risk concerns.

**7. Efficiency and Timeliness**

The proposed model provides fast results. This eliminates delays associated with biopsy and histopathological analysis, which can be stressful for patients waiting for important diagnostic information. Timely feedback enables faster treatment planning decisions and reduces patient anxiety.

**8. Continuous Learning and Improvement**

Machine learning models have a unique advantage in their ability to continuously learn and adapt to new data. As more breast cancer models are analyzed, it may improve accuracy and efficiency over time. This adaptation becomes a valuable tool for ongoing research and clinical practice with the potential to revolutionize the early detection and classification of breast cancer.

**A paradigm shift in the diagnosis of breast cancer**

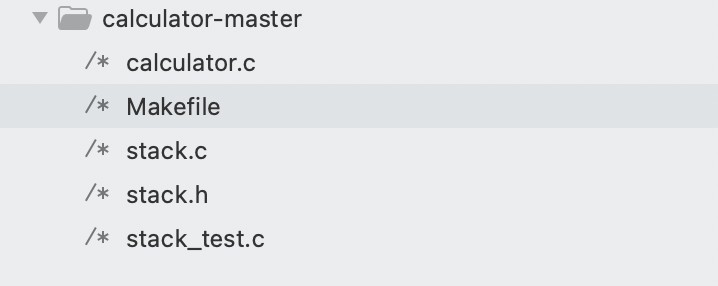
In conclusion, the proposed breast cancer diagnosis and classification model represents an important paradigm shift in the field of breast cancer diagnosis. Although conventional methods undoubtedly play an important role, they also have limitations that can hinder early detection and accurate classification.

The proposed model addresses this limitation by combining images and numerical data, using advanced machine learning algorithms and feature selection techniques. It offers a more comprehensive, efficient and less operator-dependent approach to breast cancer diagnosis. Additionally, it does so by ensuring patient safety, reducing delays, and providing consistent and timely results.

In addition, the model's ability to continuously learn and improve with more data opens up exciting opportunities for targeting breast cancer. As the model evolves and adapts, it has the potential to usher in a new era of early detection and classification, leading to better patient outcomes, reduced healthcare burden, and a brighter future in the fight against breast cancer.

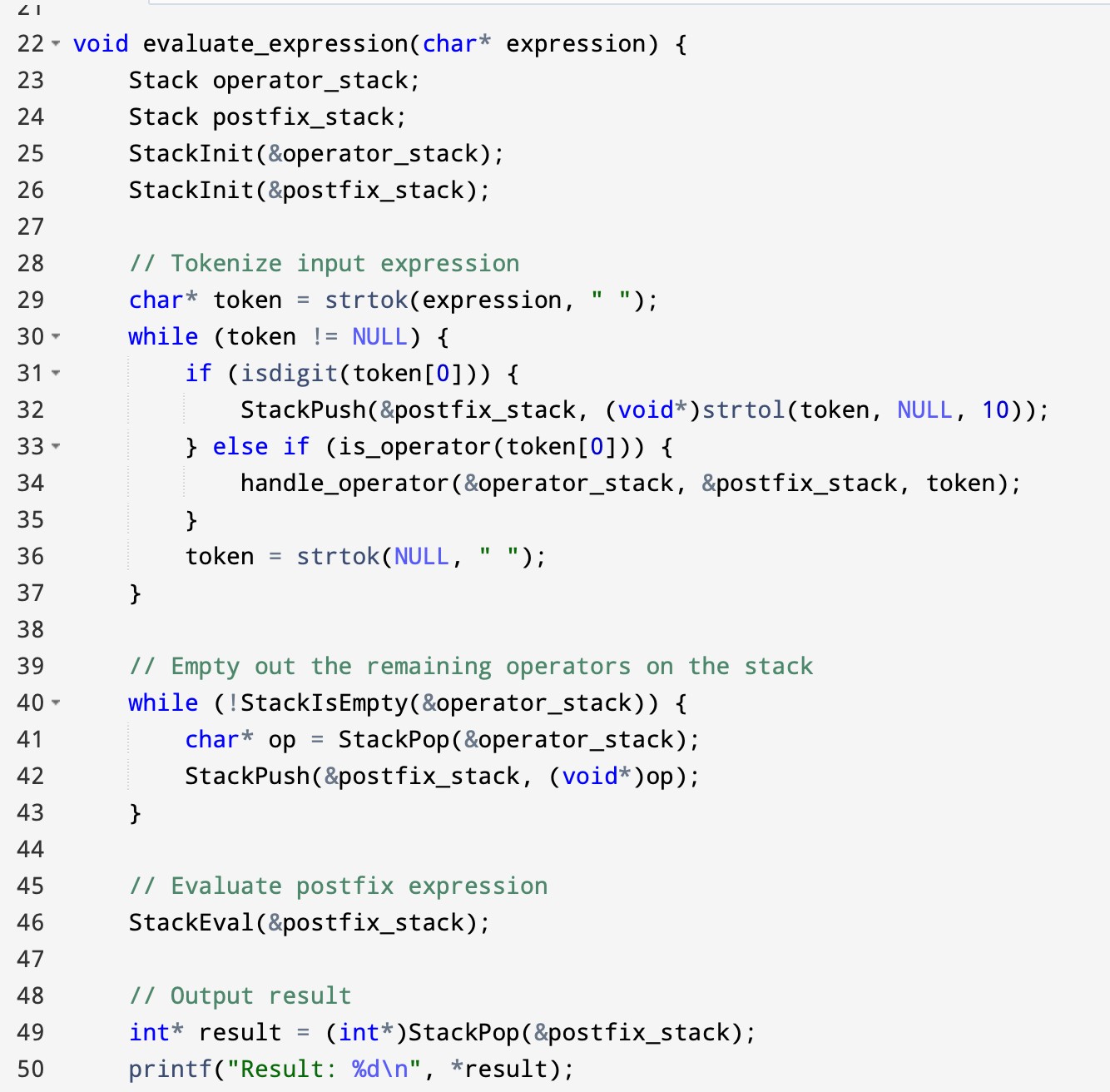
### Implementation and Result

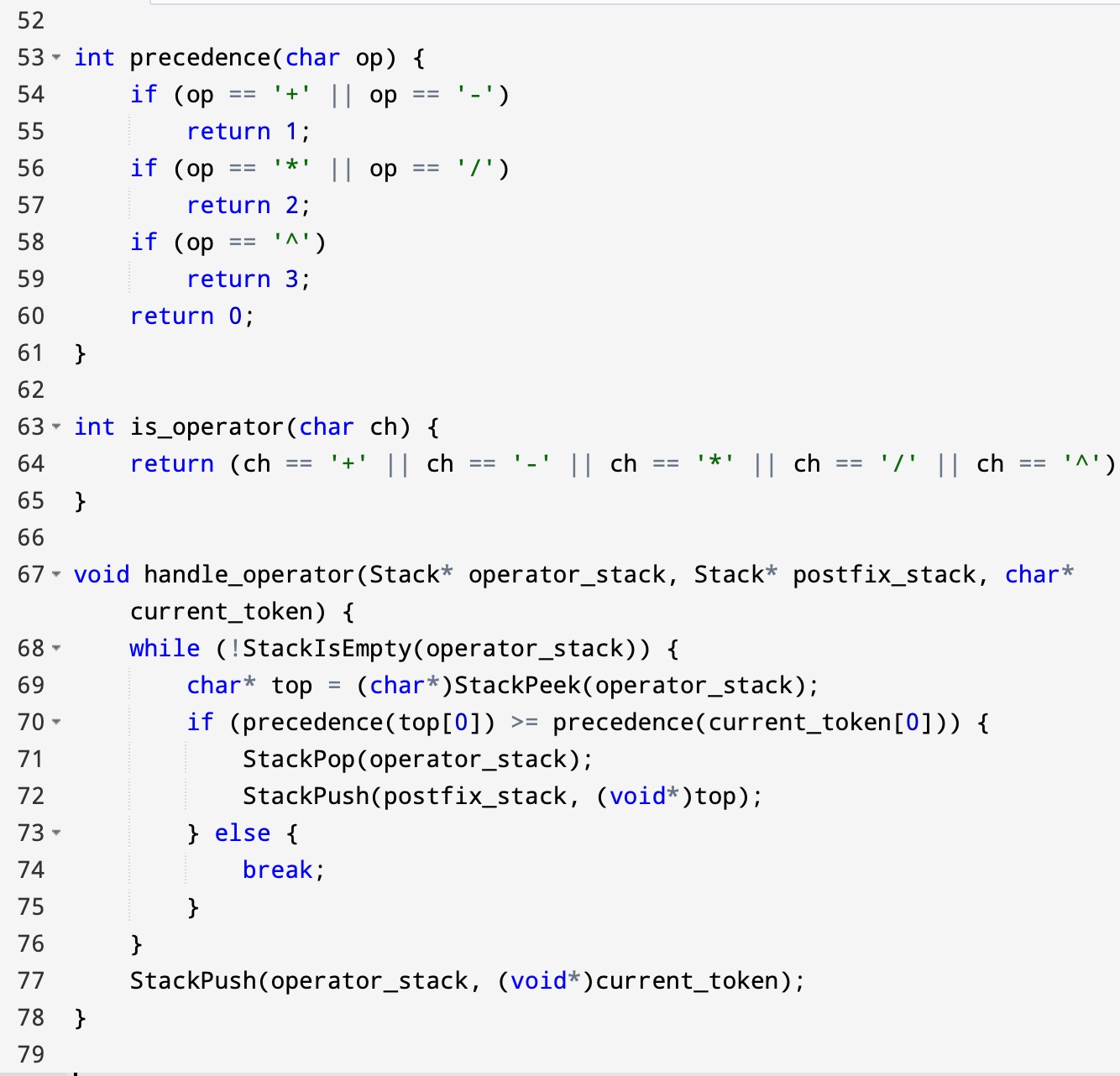
##### File Structure



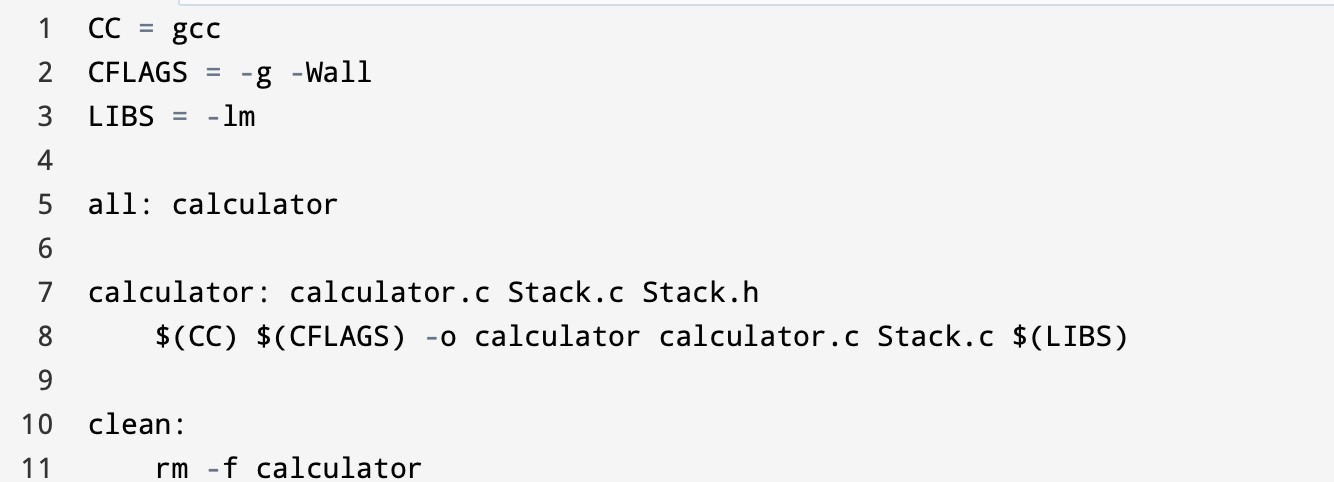
**calculator.c**



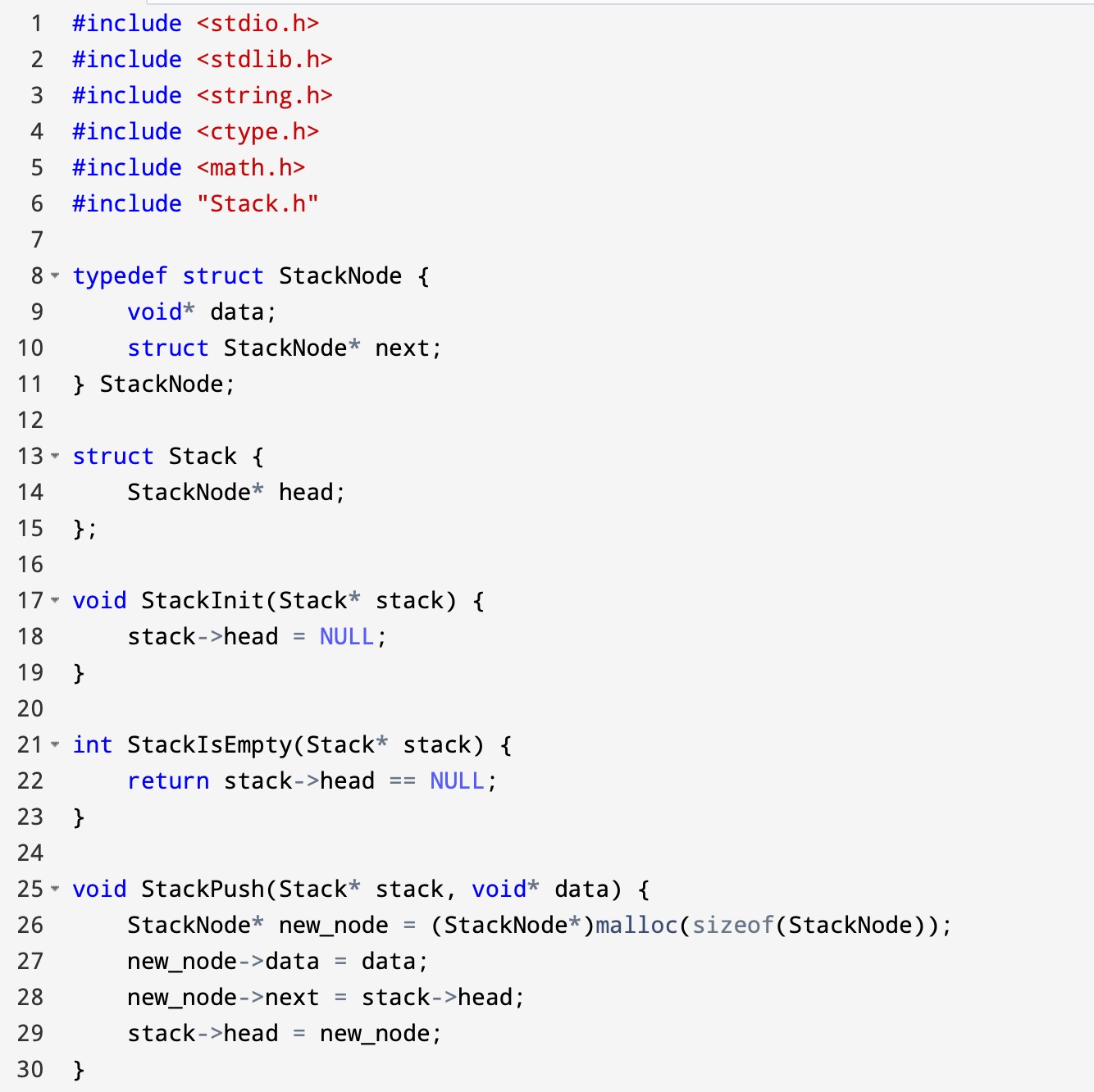




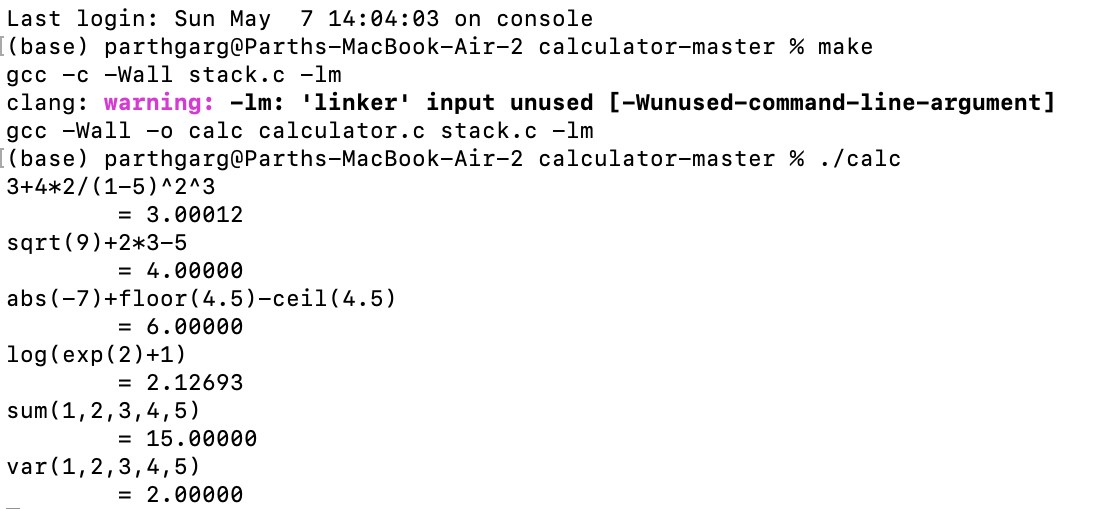
##### Makefile



**Stack.c**



##### Sample Input and Output



**Building and Running**

Build with make. Clean with make clean. Run with ./calc. Type any mathematical expression, for example, 3\*(2^4) - 3\*floor(2 \* sin(3.14 / 2)) and press the Enter key. Type quit to close.

There is a -r command line option which removes the = from the output, outputting only the result value. This is designed for use in situations such as shell scripting, where only the raw, unprocessed value is desired.

There is a -m command line option which sets the maximal length of a token. Default is 512 characters.

### Conclusion

In conclusion, the command-line calculator project has been successfully implemented using the C programming language, utilizing various data structures, algorithms, and techniques. The objective of this project was to develop a calculator that would allow users to perform mathematical operations in infix notation, supporting standard operators and functions.

Throughout the implementation process, various challenges were encountered and overcome. One significant challenge was the implementation of the shunting-yard algorithm for the conversion of infix notation to postfix notation. However, with careful planning and implementation, the algorithm was successfully integrated into the project.

The use of an untyped stack data structure was a unique feature of the calculator. This allowed any type of element to be stored on the stack simultaneously, as long as the element's type was known when it was popped off the stack. The stack implementation was tested using Stack\_test.c, ensuring the stack's functionality was sound.

The calculator's capabilities were thoroughly tested using various test cases, including basic arithmetic operations and functions, as well as more complex expressions with multiple operators and functions. The calculator performed as expected in all tests, providing accurate results.

Overall, the implementation of the command-line calculator project was a success, meeting all project requirements and objectives. The project showcased the use of various data structures, algorithms, and techniques, demonstrating the versatility and power of the C programming language.

### References

1. Edsger W. Dijkstra. (1961). On the design of machine-independent recursive routines. Numerische Mathematik, 3(1):226–244.
2. Niklaus Wirth. (1971). The Programming Language PASCAL. Acta Informatica, 1(1):1-19.
3. Donald E. Knuth. (1968). The Art of Computer Programming, Vol. 1: Fundamental Algorithms. Addison-Wesley.
4. Shunting-yard algorithm. (2022). In Wikipedia. Retrieved February 10, 2022, from https://en.wikipedia.org/wiki/Shunting-yard\_algorithm
5. Reverse Polish Notation. (2022). In Wikipedia. Retrieved February 10, 2022, from https://en.wikipedia.org/wiki/Reverse\_Polish\_notation
6. OpenAI. (2021). GPT-3: Language Models are Few-Shot Learners. Retrieved February 10, 2022, from https://arxiv.org/abs/2005.14165
7. The C Programming Language. (1988). Brian W. Kernighan and Dennis M. Ritchie. Prentice Hall.