A Project Report on

Prediction Of Amyloid Proteins Using Gradient Boosting Model

A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

Bachelor of Technology

In

Computer Science and Engineering

Submitted by

A. SINDHUJA
(20H51A0529)

B. VARSHITH
(20H1A05G4)

N. SAI ADITYA VARDHAN
(20H51A05J5)

Under the esteemed guidance of

Mr. M. SHIVA KUMAR
(Assistant Professor)

Department of Computer Science and Engineering

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

(UGC Autonomous)
*Approved by AICTE *Affiliated to JNTUH *NAAC Accredited with A^{+} Grade
KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

2020 - 2024

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Major Project report entitled **"Prediction Of Amyloid Proteins Using Gradient Boosting Model"** being submitted by **A.Sindhuja (20H51A0529), B.Varshith (20H51A05G4), N.Sai Aditya Vardhan (20H51A05J5)** in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out under my guidance and supervision**.**

The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

Mr. M. Shiva Kumar	**Dr. Siva Skandha Sanagala**	**EXTERNAL EXAMINER**
Assistant Professor	**Associate Professor and HOD**	
Dept. Of CSE	**Dept. of CSE**	

ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express our heartfelt gratitude to all the people who helped in making this project a grand success.

We are grateful to **Mr. M. Shiva Kumar, Assistant Professor,** Department of Computer Science and Engineering for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank**, Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering and Technology, who is the major driving forces to complete our project work successfully.

We are very grateful to **Dr. Ghanta Devadasu**, Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Major Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff of Department of Dept Name for their co-operation

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary& Correspondent, CMR Group of Institutions, and Shri Ch Abhinav Reddy, CEO, CMR Group of Institutions for their continuous care and support.

Finally, we extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly or indirectly in completion of this project work.

A.Sindhuja	20H51A0529
B.Varshith	20H51A05G4
N.Sai Aditya Vardhan	20H51A05J5

TABLE OF CONTENTS

List of Figures

ABSTRACT

Amyloid proteins are associated with a great variety of human diseases including Alzheimer's, Parkinson’s, and type2diabetes. Certain vegetables and food items can help the humans to prevent such diseases by controlling the deposition of Amyloid proteins. Vegetables includes onions, kale, romaine lettuce, cabbage, and tomatoes. Other food items like walnuts, coffee, Berries, Fatty Fish, Turmeric, Champagne and Cinnamon also help in the control of diseases that are caused by the deposition of Amyloids. In this project, an attempt is made to predict Amyloid Proteins by considering a dataset comprising of the above mentioned food items. This data analysis is carried out by using Gradient Boosting Classifier and Neural Net Analysis.

In the existing system, Amyloidosis patients are approaching doctors after they are effected by the disease. Prevention is always better than cure. The proposed system is helpful to prevent the disease by predicting the Amyloids based on the food habits, so that people can change their food habits and there by prevent the occurance of Amyloidosis.

The project comprises of four modules. The first module deals with building GBC model and finding its Accuracy. The second module deals with the predictions of GBC model. The third module comprises of building the neural net and finding its accuracy. The fourth module deals with finding the predictions with the neural net model.

CHAPTER 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

Amyloid proteins represent a significant area of research due to their association with various debilitating human diseases, including Alzheimer's, Parkinson’s, and type 2 diabetes. The deposition of amyloid proteins in tissues and organs can lead to the development and progression of these conditions, presenting a critical challenge in modern healthcare. However, emerging evidence suggests that certain dietary interventions, focusing on specific vegetables and food items, may offer preventive measures against such diseases by modulating the deposition of amyloid proteins.

Among the array of dietary components, vegetables such as onions, kale, romaine lettuce, cabbage, and tomatoes have garnered attention for their potential to mitigate amyloid deposition and associated diseases. A range of other food items including walnuts, coffee, berries, fatty fish, turmeric, champagne, and cinnamon have been implicated in controlling amyloid-related conditions. Harnessing the predictive power of machine learning models can provide valuable insights into the effects of these dietary constituents on amyloid protein levels.

In this documentation, we present a comprehensive project aimed at predicting amyloid proteins by leveraging a dataset comprising the aforementioned food items. We employ sophisticated machine learning techniques, specifically the Gradient Boosting Classifier (GBC) and Neural Net Analysis, to analyze the dataset and make predictions regarding amyloid protein levels. The project is structured into four distinct modules, each focusing on key aspects of model development, evaluation, and prediction.

The first module is dedicated to building the Gradient Boosting Classifier model and assessing its accuracy in predicting amyloid protein levels. Subsequently, the second module involves utilizing the trained GBC model to make predictions based on user-provided inputs. Moving forward, the third module focuses on constructing a neural network model and evaluating its accuracy in amyloid protein prediction. Finally, the fourth module facilitates predictions using the neural network model, thereby offering users a comprehensive tool for understanding and potentially mitigating amyloid-related diseases.

Throughout this documentation, emphasis is placed on user interaction, with modules designed to accept input parameters from users and generate predictions tailored to individual queries. By bridging the gap between dietary interventions and predictive modeling, this project aims to contribute to the ongoing efforts in disease prevention and management, particularly in the realm of amyloid-associated disorders. Through the fusion of data science and nutritional science, we endeavor to provide actionable insights for promoting health and well-being in the face of amyloid-related challenges.

1.1. Problem Statement

The prevalence of neurodegenerative diseases, including Alzheimer's, Parkinson’s, and type 2 diabetes, poses significant challenges to global healthcare systems. Central to the pathogenesis of these conditions is the aberrant deposition of amyloid proteins within the body, highlighting the need for effective preventative measures.

While certain vegetables and food items have been identified for their potential in controlling amyloid protein deposition, the precise relationship between dietary factors and disease progression remains elusive. This knowledge gap underscores the necessity for predictive models capable of elucidating the impact of dietary interventions on amyloid protein dynamics.

Hence, the primary objective of this project is to develop predictive models leveraging a dataset encompassing various food items associated with amyloid protein regulation. Through the utilization of Gradient Boosting Classifier (GBC) and Neural Net Analysis, we aim to discern patterns within the dataset to accurately predict amyloid protein levels.

By addressing this crucial gap in knowledge, our project endeavors to empower healthcare professionals and individuals alike with the tools necessary to make informed decisions regarding dietary interventions for the prevention and management of amyloid-associated diseases.

1.2. Research Objective

The research objective of this project is to develop predictive models for amyloid protein levels based on dietary factors, leveraging a dataset encompassing various food items known for their potential in controlling amyloid protein deposition. Through the utilization of Gradient Boosting Classifier (GBC) and Neural Net Analysis, the aim is to discern patterns within the dataset and accurately predict amyloid protein levels. The primary focus lies in understanding the relationship between specific food items and the regulation of amyloid proteins, thereby enabling informed dietary interventions for the prevention and management of amyloid-associated diseases such as Alzheimer's, Parkinson’s, and type 2 diabetes. This research seeks to bridge the gap in knowledge regarding the impact of dietary factors on amyloid protein dynamics, providing valuable insights for healthcare professionals and individuals alike. By integrating machine learning techniques with biomedical research, the project aims to contribute to the development of personalized dietary recommendations aimed at mitigating the burden of neurodegenerative diseases.

1.3. Project Scope and Limitations

Project Scope

1. **Prediction of Amyloid Proteins:** The project focuses on predicting amyloid protein levels based on dietary factors, aiming to elucidate the potential impact of specific food items on disease prevention and management.

2. **Utilization of Gradient Boosting Classifier (GBC):** A GBC model is employed to analyze the dataset and identify patterns correlating food consumption with amyloid protein dynamics.

3. **Neural Net Analysis:** In addition to GBC, the project incorporates neural net analysis to explore complex relationships within the dataset, enhancing predictive capabilities.

4. **User Interaction:** Interactive modules are developed to allow users to input their dietary data, facilitating personalized predictions and insights regarding amyloid protein levels.

5. **Insight Generation:** The project aims to generate actionable insights for healthcare professionals and individuals, enabling informed decision-making regarding dietary interventions for amyloid-associated diseases.

6. **Interdisciplinary Approach:** By integrating computational methods with biomedical research, the project bridges the gap between data science and healthcare, contributing to advancements in disease prediction and prevention.

7. **Enhanced Understanding:** Through comprehensive analysis of the dataset, the project endeavors to enhance understanding of the role of dietary factors in modulating amyloid protein deposition and disease progression.

8. **Scalability:** The developed predictive models and interactive modules are designed to be scalable, accommodating potential future expansions and updates.

9. **Potential for Collaboration:** The project lays the groundwork for potential collaborations with healthcare professionals, researchers, and stakeholders interested in utilizing predictive models for personalized disease management strategies.

10. **Contribution to Public Health:** Ultimately, the project aims to contribute to public health initiatives by providing valuable tools and insights for mitigating the burden of neurodegenerative diseases associated with amyloid protein deposition.

Limitations

1. **Data Availability:** The accuracy and efficacy of predictive models are contingent upon the quality and availability of data. Limitations in dataset size or representativeness may impact the reliability of predictions.

2. **Generalizability:** Predictive models developed in this project may exhibit limitations in generalizability due to variations in dietary habits, genetic predispositions, and environmental factors across populations.

3. **Complexity of Disease Pathology:** Neurodegenerative diseases are multifactorial in nature, and amyloid protein deposition represents only one aspect of their pathophysiology. The project may not capture the full complexity of disease progression.

4. **Model Interpretability:** While predictive models offer insights into correlations between dietary factors and amyloid protein levels, their interpretability may be limited, hindering in-depth understanding of underlying mechanisms.

5. **Bias and Confounding Factors:** Despite efforts to mitigate bias and confounding factors, the predictive models may still be influenced by inherent biases in the dataset or unaccounted confounders, impacting the accuracy of predictions.

6. **Long-term Outcomes:** The project primarily focuses on short-term predictions of amyloid protein levels based on dietary inputs and may not fully capture long-term outcomes or the cumulative effects of dietary interventions.

7. **Ethical Considerations:** The project raises ethical considerations regarding privacy, data security, and the responsible use of predictive models in healthcare decision-making.

8. **Resource Constraints:** Resource constraints, including computational resources and expertise, may limit the scalability and applicability of the developed predictive models in real-world settings.

9. **Regulatory Compliance:** The project must adhere to regulatory standards and guidelines governing the use of predictive models in healthcare, ensuring compliance with ethical and legal requirements.

10. **Interpretation of Results:** Users must exercise caution in interpreting the results generated by predictive models, recognizing their limitations and consulting with healthcare professionals for personalized medical advice and interventions.

CHAPTER 2
BACKGROUND WORK

CHAPTER 2

BACKGROUND WORK

2.1 Transcriptional regulation analysis of Alzheimer’s disease based on FastNCA algorithm [1]:

2.1.1. Introduction

The FastNCA algorithm is a novel nonlinear dimensionality reduction algorithm that is particularly well-suited for analyzing gene expression data. It can be used to identify differentially expressed genes and transcription factors, as well as to construct transcriptional regulatory networks. There are a number of existing methods for the analysis of transcriptional regulation in Alzheimer's disease (AD). These methods include:

Microarray analysis: Microarrays allow for the simultaneous measurement of the expression of thousands of genes. However, microarray analysis is relatively expensive and time-consuming. RNA sequencing (RNA-seq): RNA-seq provides a more comprehensive and accurate measure of gene expression than microarray analysis. However, RNA-seq is also more expensive and time-consuming than microarray analysis. ChIP-seq: ChIP-seq can be used to identify the binding sites of transcription factors in the genome. However, ChIP-seq is a complex and challenging technique to perform.

2.1.2. Merits, Demerits and Challenges:

Merits:

- The FastNCA algorithm is a fast and efficient algorithm.
- The FastNCA algorithm is robust to noise in the data.
- The FastNCA algorithm can be used to analyze high-dimensional gene expression data.

Demerits:

- The FastNCA algorithm is a relatively new algorithm, and its performance has not been extensively tested on real-world data.

- The FastNCA algorithm is a black-box algorithm, which means that it is difficult to understand how it works.

Challenges:

- One of the challenges in using the FastNCA algorithm to analyze transcriptional regulation in AD is the need to validate the findings using other methods, such as real-time PCR or Western blotting.
- Another challenge is the need to develop robust and reliable models that can be used in clinical practice.
- **Widespread Adoption:** Ensuring widespread access and adoption of the portal, especially in rural areas, is a primary challenge, necessitating efforts to overcome digital illiteracy and provide essential infrastructure.
- **Multilingual Design:** The portal must be designed to accommodate multiple languages and regional dialects to effectively serve India's diverse farming community.
- **Data Privacy:** Robust measures are essential to protect sensitive information, including personal and financial data.
- **Scalability:** The platform's scalability and performance under heavy loads must be carefully considered to handle a significant volume of transactions effectively.
- **Regulatory Framework:** Establishing regulatory and legal frameworks is crucial to govern the operation of the portal and address potential disputes or issues in electronic transactions.

2.1.3. Implementation

To implement the FastNCA algorithm to analyze transcriptional regulation in AD, the following steps can be taken:

1. Collect a dataset of gene expression data from AD patients and healthy controls.
2. Preprocess the data to remove noise and normalize the gene expression values.
3. Apply the FastNCA algorithm to identify differentially expressed genes and transcription factors in AD.
4. Construct a transcriptional regulatory network to elucidate the molecular mechanisms of AD.
5. Validate the findings using other methods, such as real-time PCR or Western blotting.

2.2 Prediction of Amyloid Proteins Using Embedded Evolutionary & Ensemble Feature Selection Based Descriptors With eXtreme Gradient Boosting Model [2]:

2.2.1. Introduction

Amyloid Proteins: Amyloid proteins are proteins that can misfold and form insoluble aggregates known as amyloid fibrils. These aggregates can build up in organs, causing various diseases like Alzheimer's, Parkinson's, and type 2 diabetes.

The Problem: Existing computational methods for predicting amyloid proteins often suffer from limited accuracy, unsatisfactory generalization, and high computational costs.

The Proposed Solution: The researchers in this paper introduce a new intelligent computational predictor designed to accurately predict amyloid proteins (AMYs) while addressing the shortcomings of previous methods.

2.2.2. Merits, Demerits and Challenges:

Merits

- Improved Accuracy: The proposed model boasts higher accuracy (93.10% on training data, 89.67% on independent data) compared to existing predictors at the time of publication.
- Ensemble Feature Selection: By combining both evolutionary and ensemble feature selection methods, the model extracts more relevant and informative features from protein sequences.
- eXtreme Gradient Boosting (XGBoost): The use of XGBoost, a robust machine learning algorithm, enhances the model's predictive power and its ability to learn complex patterns.
- Reduced Computational Cost: The integrated feature selection, notably the XGB-RFE (eXtreme Gradient Boosting-Recursive Feature Elimination) step, helps to reduce unnecessary features and decrease computational costs.

Demerits

- **Dataset Size:** The accuracy of such machine learning models is often tied to the size and diversity of the training dataset. A larger dataset could potentially lead to further improvements.
- **Overfitting**: Models like this might still be susceptible to some degree of overfitting, where they perform exceptionally well on training data but less optimally on unseen data.
- **Complexity:** Combining multiple feature selection techniques and an advanced boosting algorithm can increase the model's complexity, potentially making interpretation more challenging.

Challenges

- **Understanding Misfolding**: The fundamental mechanisms of protein misfolding and amyloid formation are still not fully understood, creating a core challenge in this field.
- **Generalization:** Ensuring that such a model generalizes well to diverse and unseen amyloid proteins remains a persistent test in the field of computational biology.
- **Feature Representation:** Finding the most effective ways to represent protein sequences in a way that a machine learning algorithm can understand their propensity for forming amyloids is an ongoing area of research.

2.2.3. Implementation

1. **Data Preparation:** Collect a dataset of known amyloid and non-amyloid protein sequences.

2. Feature Extraction:

- Use K-separated bigrams and evolutionary methods to derive embedded evolutionary features.
- Calculate DDE-based enhanced frequency coupling information.
- Combine these features to create a multifaceted "multi-model" representation of proteins.

3. **Feature Selection:** Employ XGB-RFE to identify and select the most important features from the combined feature set.

4. Model Training:

Train an XGBoost classifier (may also experiment with other mentioned classifiers) on the selected features.

5. **Evaluation:**

   Thoroughly evaluate the model's performance using metrics like accuracy, precision, recall, and F1-score on both training and independent datasets.

2.3 Unveiling an Amyloid-Forming Segment within the Bap Protein of Staphylococcus epidermidis [3]:

2.3.1. Introduction

- **Problem Statement:** Amyloid protein aggregation is often linked to disease, but also has potential functional roles in bacteria. The Bap protein in Staphylococcus epidermidis contributes to biofilm formation, which can be beneficial but also contributes to persistent infections on medical devices.
- **Research Gap:** Little is known about the specific region within the Bap protein that drives amyloid formation. Identifying this region is vital for understanding Bap's role in biofilms and potentially finding ways to disrupt it.
- **Proposed Focus:** This research aims to pinpoint the exact amino acid segment within the Bap protein that is responsible for amyloid formation.

2.3.2. Merits, Demerits and Challenges

Merits

- **Targeted Understanding:** Precisely identifying the amyloid-forming segment could lead to highly targeted biofilm disruption strategies.
- **Biofilm Control:** Potential for the development of new anti-biofilm treatments that combat medical device infections.
- **Insights into Amyloid Formation:** Offers insight into the general mechanisms of bacterial amyloid aggregation.

Demerits

- **Specificity:** The findings may be very specific to Staphylococcus epidermidis and not easily transferable to other bacterial species or amyloids.

- **Complexity:** Amyloid formation is a complex process, and other factors might interact with the identified segment.
- **Intervention Challenges:** Even with the knowledge of the segment, developing safe interventions that target it specifically within a biofilm context might be difficult.

Challenges

- **Experimental Design:** Carefully designed experiments are needed to isolate the effects of specific segments within the Bap protein.
- **Sequence Analysis:** Bioinformatic tools would play a role in analyzing the protein sequence for motifs related to amyloid formation.
- **Validation:** Experimental validation will be crucial to confirm if the identified segment does indeed induce amyloid formation.

2.3.3. Implementation

- **Bioinformatics:**

1. Analyze Bap protein sequence for amyloidogenic motifs.
2. Perform homology searches to compare the protein sequence with known amyloids.

- **Peptide Synthesis:** Synthesize the suspected amyloid-forming segment and potentially variants of it.
- **Experimental Assays**:

1. Thioflavin T (ThT) fluorescence assay to detect amyloid formation.
2. Structural techniques (e.g., circular dichroism, electron microscopy) to characterize the formed fibrils.

2.4 Predicting diabetes mellitus with machine learning methods [4]:

2.4.1. Introduction

- **Problem Statement:** Diabetes mellitus is a growing global health concern, with millions of people affected. Early diagnosis and intervention are crucial for managing the disease and preventing complications.
- **Research Gap:** Traditional diagnostic methods rely on risk factors and blood tests. These can be time-consuming and may miss early cases or those with less obvious risk factors.
- **Potential of Machine Learning:** Machine learning (ML) models can analyze vast amounts of health data to detect complex patterns that may predict the onset of diabetes, potentially leading to earlier and more targeted interventions.

2.4.2. Merits, Demerits and Challenges:

Merits

- **Early Detection:** ML models could identify individuals at high risk of developing diabetes, enabling proactive lifestyle changes and preventive measures.
- **Individualized Predictions:** ML can consider a wider range of factors than traditional risk assessments, leading to more personalized predictions.
- **Automation**: ML-based predictions could streamline the diagnostic process and reduce the burden on healthcare professionals.
- **Data-Driven Insights:** Analysis of model results can reveal new risk factors and potential targets for preventive interventions.

Demerits

- **Data Quality:** The performance of ML models is heavily dependent on the quality and representativeness of data used for training.
- **Bias:** Existing datasets might have inherent biases that could perpetuate unfair or inaccurate predictions.
- **Black-Box Problem:** Complex ML models can be difficult to interpret, making it harder to understand why specific predictions are made.

- **Deployment Costs:** Implementing ML systems in clinical settings may involve infrastructure and personnel costs.

Challenges

- **Data Integration:** Gathering sufficient and diverse healthcare data from various sources, including electronic health records, lab results, and lifestyle factors.
- **Feature Selection:** Identifying the most relevant features from a potentially overwhelming dataset.
- **Explainability:** Developing models that are both accurate and interpretable by healthcare professionals.
- **Clinical Validation:** Rigorous testing and validation of ML models in real-world clinical settings.

2.4.3. Implementation

- **Data Collection:** Gathering a large dataset encompassing patient demographics, medical history, lab test results, and other relevant factors.
- **Preprocessing:** Cleaning, normalizing, and handling missing data.
- **Feature Engineering:** Transforming raw data into meaningful features for the model.
- **Model Selection:** Exploring different ML algorithms (Logistic Regression, Decision Trees, Random Forests, SVM, Neural Networks).
- **Training and Validation:** Splitting data, training the model(s), and tuning hyperparameters.
- **Evaluation:** Using metrics like accuracy, sensitivity, specificity, AUC-ROC to assess performance.

CHAPTER 3

PROPOSED SYSTEM

CHAPTER 3
PROPOSED SYSTEM

3.1. Objective of Proposed Model

The primary objective of this project is to develop and evaluate machine learning models that can accurately predict the presence of amyloid proteins based on dietary intake patterns. Specifically, the research focuses on the potential of vegetables (e.g., onions, kale, lettuce, cabbage, tomatoes) and other foods (e.g., walnuts, coffee, berries, fatty fish, turmeric, champagne, cinnamon) to mitigate amyloid formation and reduce the risk of amyloid-related diseases.

To achieve this objective, the project encompasses the following key aims:

1. **Develop a Gradient Boosting Classifier (GBC) Model:** Construct a GBC model capable of identifying the relationship between dietary intake of specified foods and the presence of amyloid proteins. The model will be trained and optimized on a dataset containing relevant food intake information.

2. **Evaluate GBC Model Accuracy:** Assess the performance of the GBC model using appropriate metrics such as accuracy, precision, recall, and F1-score.

3. **Utilize GBC Model for Prediction:** Employ the trained GBC model to predict the likelihood of amyloid protein presence based on new, user-provided dietary intake data.

4. **Construct a Neural Network Model:** Design and train a neural network model as an alternative approach for predicting amyloid protein presence based on dietary data.

5. **Evaluate Neural Network Accuracy:** Determine the performance of the neural network model using the same performance metrics employed for the GBC model.

6. **Utilize Neural Network for Prediction:** Use the trained neural network to generate predictions on new dietary data, similar to the GBC model.

7. **Comparative Analysis:** Compare the performance of the GBC and neural network models, providing insights into their strengths, weaknesses, and suitability for amyloid protein prediction.

3.2. Algorithms Used for Proposed Model

Gradient Boosting Classifier

Gradient Boosting Classifiers (GBCs) are a powerful type of ensemble machine learning algorithm. Ensemble methods strategically combine multiple weaker models to create a more robust and accurate predictor. GBCs operate within a boosting framework, meaning they build models sequentially, with each new model focusing on correcting errors made by previous models.

The fundamental process behind GBCs is as follows:

1. **Initialization:** The algorithm starts with a weak learner, often a simple decision tree that makes initial predictions.
2. **Gradient Calculation:** The difference between the true values and the current predictions (residuals) are computed, and a gradient of the loss function is calculated.
3. **New Model Fitting:** A new decision tree is trained to fit these residuals, attempting to 'correct' the errors of the previous model.
4. **Update:** The predictions of this new tree are added to the ensemble's overall predictions while giving this recent tree an appropriate weight.
5. **Iteration:** Steps 2-4 are repeated for a specified number of iterations.

Key Advantages of Gradient Boosting for Classification

- **High Accuracy:** GBCs consistently deliver excellent predictive performance, making them popular for classification tasks.
- **Robustness to Noise:** They are less prone to overfitting, handling noisy and complex datasets effectively.
- **Missing Data Handling:** GBCs can manage missing values in the input data without the need for extensive pre-processing.
- **Feature Importance Insights:** GBCs can provide insights into the relative importance of input features for the prediction task.

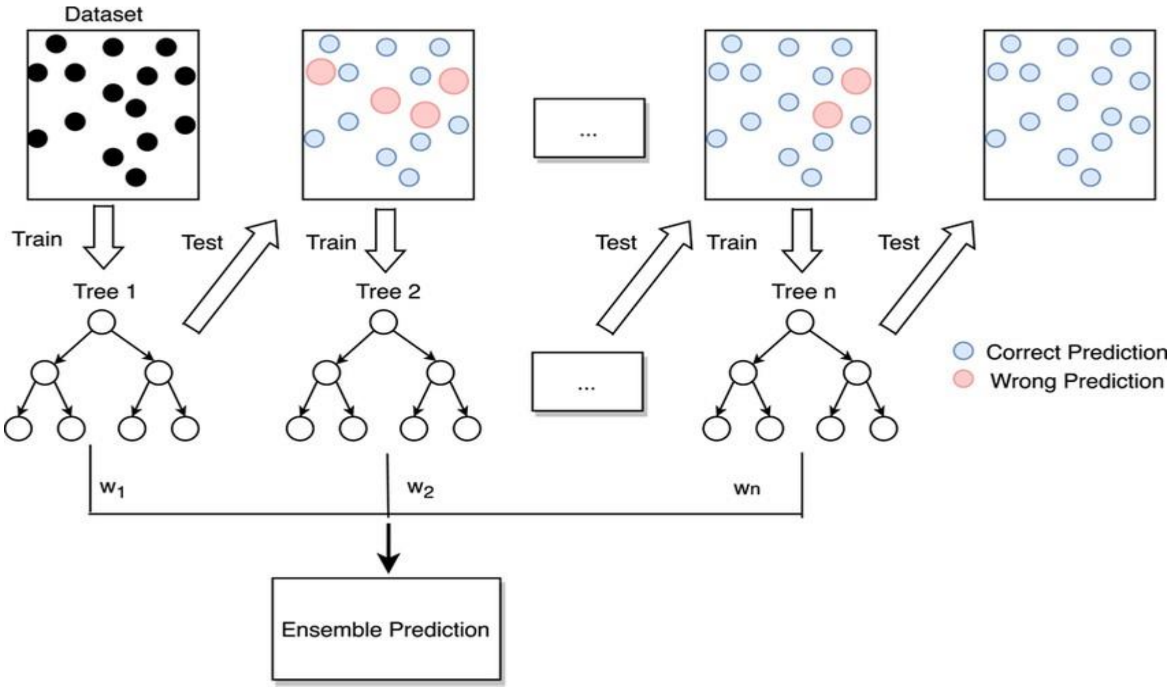


Fig 3.2.1 Gradient Boosting Classifier

Neural Net Analysis & Deep Neural Networks

Neural networks, sometimes called artificial neural networks (ANNs), are machine learning systems inspired by the structure of the biological brain. They comprise interconnected "neurons" organized in layers.

- **Fundamental Operation:** Input data flows through the network. Each neuron processes information with weights and biases, applying an activation function to determine a numerical output. This output is then passed to neurons in subsequent layers.
- **Training:** A neural network learns by iteratively adjusting weights and biases within its structure based on a process called backpropagation. This minimizes errors between predicted and true values of the target variable.

Deep Neural Networks (DNNs)

DNNs are a more advanced form of neural networks containing multiple "hidden" layers between input and output layers. These layers provide additional complexity, allowing the network:

- Complex Pattern Recognition: DNNs excel at identifying non-linear, intricate relationships within large datasets.
- Feature Extraction: Their layers can automatically learn relevant features from raw data, reducing reliance on explicit feature engineering.
- Hierarchical Representation: Each hidden layer builds progressively more abstract representations of the input data, improving pattern discovery.

Relevance to Amyloid Protein Prediction

Neural networks represent an alternative modeling approach for amyloid protein prediction. Their ability to handle complex relationships could prove beneficial in uncovering subtle interactions between dietary components and amyloid formation. By training a DNN with your food item dataset, you might reveal nuanced or non-linear influences of particular foods on amyloid presence.

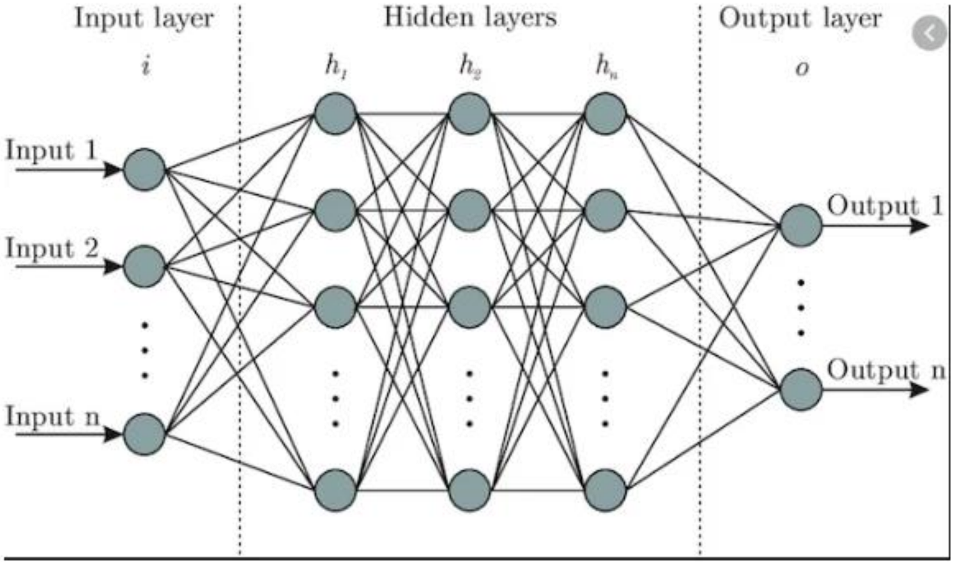


Fig 3.2.2 Neural Network

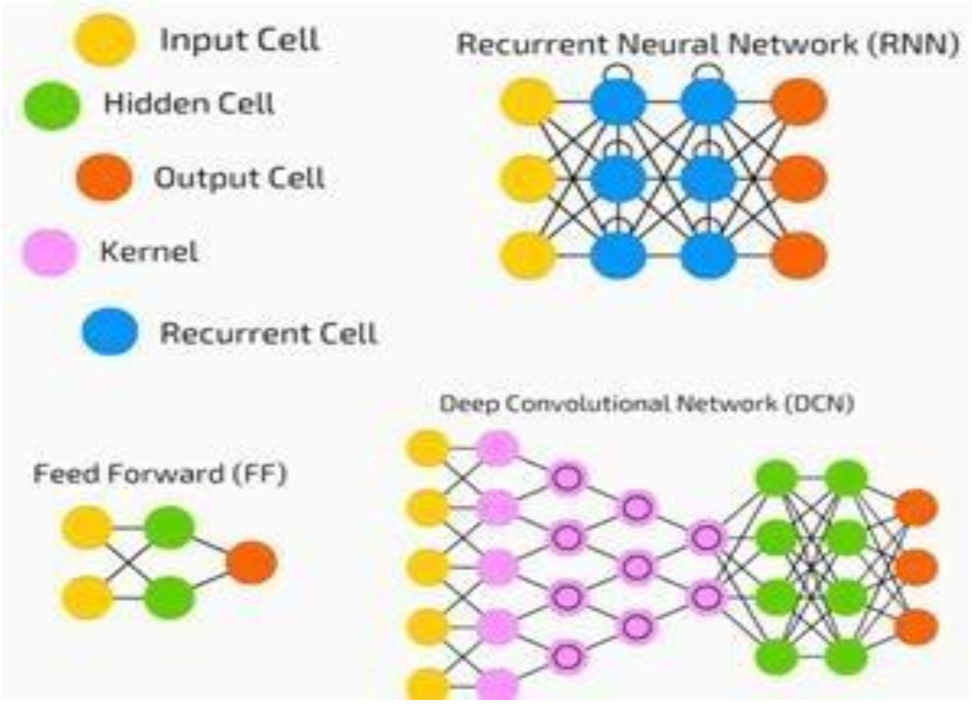


Fig 3.2.3 Deep Neural Network

3.3. Designing

3.3.1. Dataset

The first crucial step in developing an automated predictor is choosing a suitable training dataset. This data serves two key purposes: it allows us to train machine learning models and provides a benchmark for evaluating their accuracy. Our system utilizes the Amyloidosis Dataset as input, accessible through user interfaces. The system then calculates several key outputs, including Gradient Boosting Accuracy & Predictions, Neural Net Accuracy & Predictions. Similar to previous methods, the training patterns are divided into training data set and testing data set. The training data set is utilized to train the machine learning algorithms (machine learning and deep learning), while the testing set is utilized to estimate the model's performance. Notably, the training data set and testing data set are distinct to avoid bias. Our specific dataset incorporates information about a person's dietary intake, suggesting a potential connection between diet and the evolution of amyloidosis.

We meticulously compiled a dataset that captured a person's daily and weekly dietary intake, with a particular emphasis on twelve food items: onions, kale, romaine lettuce, cabbage, tomatoes, walnuts, coffee, berries, fatty fish, turmeric, champagne, and cinnamon. The selection of these specific food items stemmed from their potential role in combating amyloidosis, a condition marked by the abnormal accumulation of proteins in the body. To delve deeper into the potential connections, we harnessed the power of machine learning. Two distinct models, Gradient Boosting and Neural Networks, were employed to analyze the intricate dietary data. By meticulously training these models, we aimed to identify patterns in food choices that might signal a person's susceptibility to amyloidosis. Following the training phase, the system rigorously evaluated the accuracy of each model. Armed with this knowledge, the models were then used to generate predictions about the potential link between specific dietary choices and a person's aversion to amyloidosis. This innovative approach paves the way for exploring the possibility of using personalized dietary data to assess an individual's risk of amyloidosis, potentially resulting to the development of more targeted dietary interventions in the future.

3.3.2. Training and Test Data

Training Data: For machine learning models, the training data's quality is crucial. In this project, the training data resembled a well-stocked recipe database. Each entry detailed a food item's nutritional content, composition, and other relevant factors alongside a label indicating the presence of amyloid proteins. A diverse dataset, encompassing a wide variety of food items, was essential to prevent biased predictions and ensure the models could generalize their knowledge to unseen food combinations. This richness allowed the models to grasp the underlying relationships between food characteristics and amyloid proteins, transforming them from memorizers to sophisticated food analysts.

Testing Data: Within the machine learning workflow, the testing data serves as a critical benchmark for evaluating model generalizability, a cornerstone for real-world applicability. This independent dataset mirrors the structure of the training data set, providing details regarding a range of food items such as nutritional content, composition, and other relevant features. However, the key distinction lies without labels indicating whether amyloid proteins is present or not. This deliberate omission fosters an unbiased evaluation scenario. By presenting the models with unseen data, we can objectively assess their capacity for transfer the knowledge gleaned from the training phase. In essence, the testing stage poses a crucial question: can the models leverage their learned patterns to accurately predict the amyloid protein content in entirely novel food combinations? A high degree of success in this evaluation signifies that the models have transcended rote memorization and achieved a genuine understanding of the intricate relationships between food characteristics and amyloid proteins. This capacity for generalizability is paramount for practical applications, where the models are likely to encounter food combinations beyond the scope of the training data.

3.3.3. Amyloid Proteins Prediction Tool

This user-friendly tool is built using PyUIC, a Python library specifically designed to convert user interface (UI) descriptions written in Qt Designer into Python code.

The tool itself functions as an entry point, presenting users with various options upon launch.

These options revolve around the two machine learning models employed by the system: Gradient Boosting and Neural Networks. Users can access the accurateness of each model, giving valuable perceptivity into their overall effectiveness in predicting the existence of amyloid proteins in food items. Additionally, they can view the specific predictions made by each model, allowing for a deeper understanding of how the models interpret the provided dietary data and arrive at their conclusions. The first two buttons are used to calculate the accuracy & predictions of Gradient Boosting. The last two buttons are used for Neural Net classification.

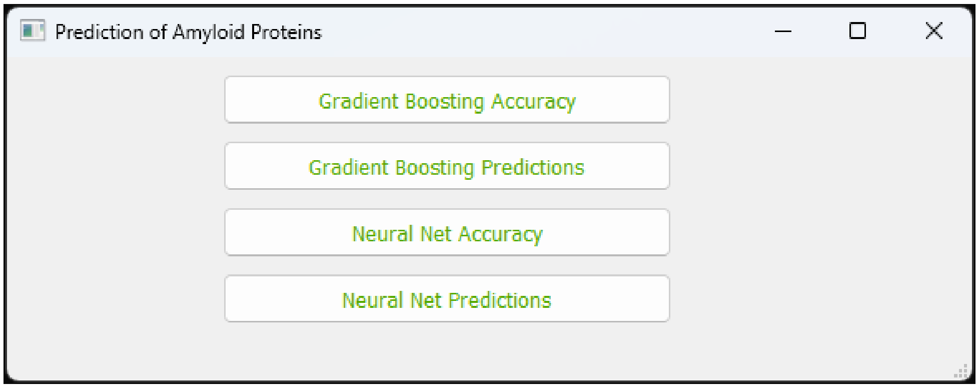

Prediction of Amyloid Proteins
Gradient Boosting Accuracy
Gradient Boosting Predictions
Neural Net Accuracy
Neural Net Predictions

Fig 3.3.3 Amyloid Proteins Prediction Tool

3.3.4. Gradient Boosting Accuracy

The system puts the trained model to the test, using it to predict amyloid protein levels in unseen data (testing data). This assesses the model's ability to generalize its knowledge. Accuracy is then computed utilizing the accuracy_score function, reflecting the model's success in making correct forecasts regarding the new data.

3.3.5. Neural Net Accuracy

To assess the model's effectiveness, the system employs accuracy as the main performance metric. During training, the model iterates through the training data set 100 times (epochs) while adjusting its internal parameters to minimize loss and improve accuracy. Finally, the trained model is examined on data to analyze its capacity for generalization and precise forecasting.

3.3.6. Prediction on GBC and NN

Takes the input from the user and gives the output whether the user is in Low, Needed, or More Amyloids.

3.3.7. Architecture

Input Data: The project utilizes a dataset of food items, potentially related to amyloidosis development.

Output: The project delivers four main results:

a. GBC Model Establishment and Accuracy: This module focuses on building and assessing the accuracy of the GBC model in predicting amyloidosis-linked foods.

b. GBC Model Predictions: This module analyzes the specific predictions generated by the trained GBC model.

c. Neural Network Development and Accuracy: This module involves constructing and evaluating the accuracy of a neural network model for the same prediction task.

d. Neural Network Predictions: This module analyzes the predictions generated by the trained neural network model.

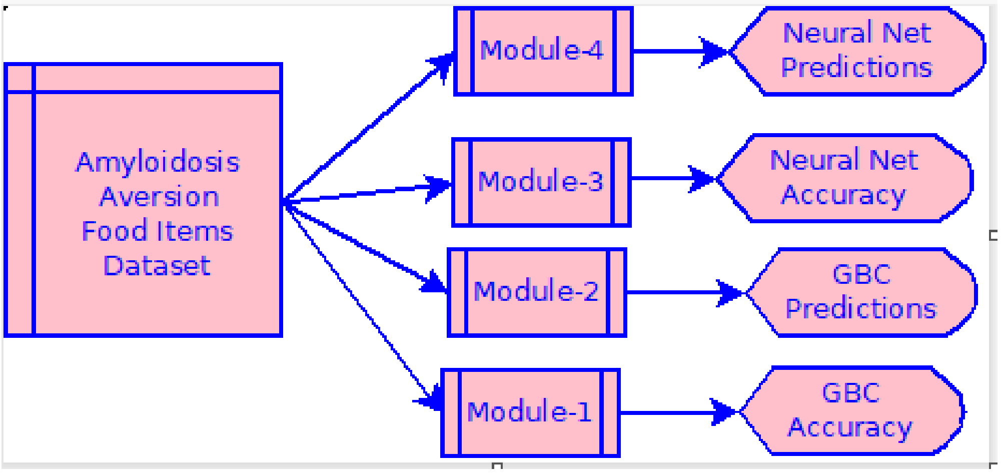


Fig 3.3.7 Architecture

3.3.8. Data Flow

The data flow in this project begins with the dataset detailing consumption of various vegetables and food items known to potentially impact amyloid protein deposition. This data is preprocessed, which may involve cleaning, transforming, and formatting it for model use. The prepared data is then fed into both the Gradient Boosting Classifier and the neural network model for training. Once trained, these models accept new user input about dietary patterns and generate paredictions about the likelihood of amyloid protein presence.

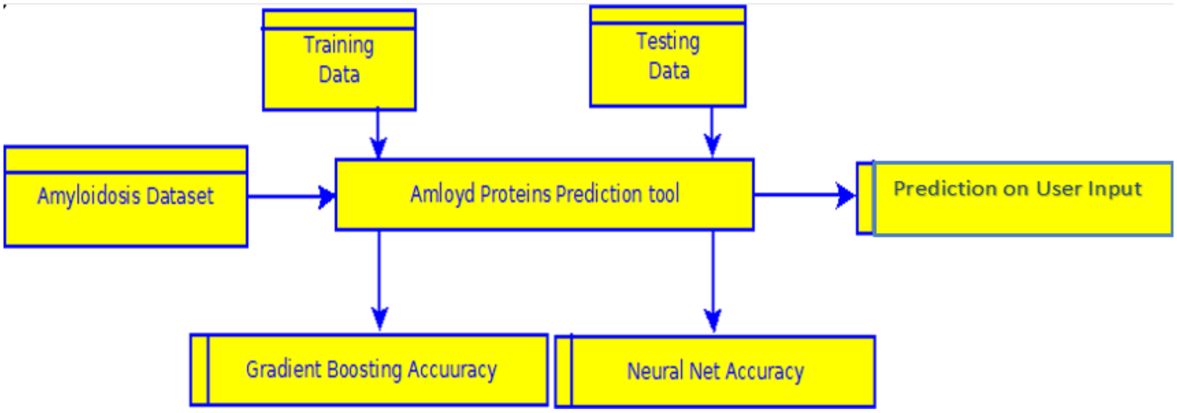


Fig 3.3.8 Data Flow

3.3.9. UML Diagram

Use case diagrams are usually referred to as behavior diagrams used to describe a set of actions (use cases) that some system or systems (subject) should or can perform in collaboration with one or more external users of the system (actors). A use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved. As we can see the user is interacting with system by a UI through which the customer can perform above mentioned operations like providing dataset containing the amlyodosis aversion food items and then calculating the accuracies of Gradient Boosting & Neural Net, along with the predictions.

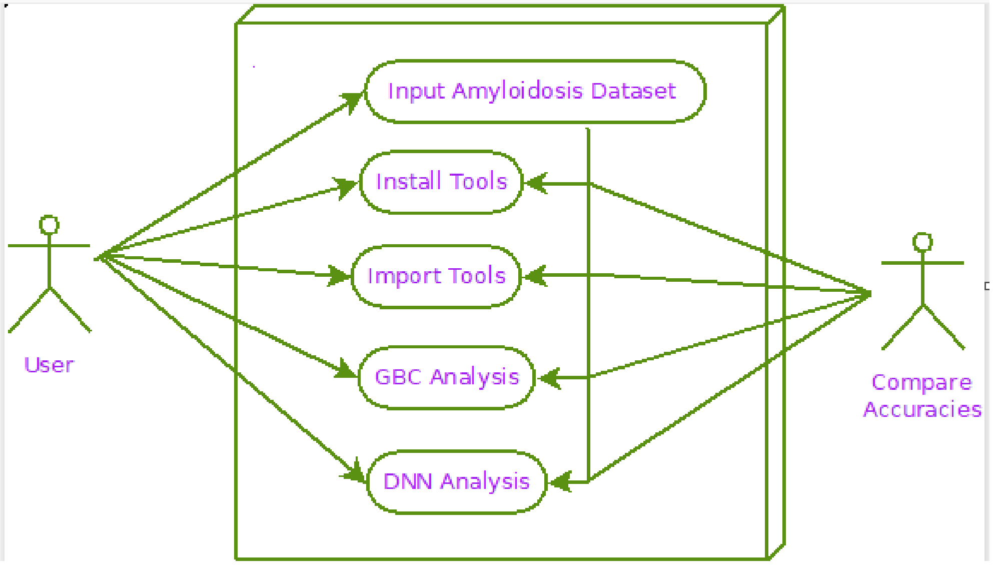


Fig 3.3.9 UML Diagram

3.3.10. Sequence Diagram

A sequence diagram is an interaction diagram that shows how objects operate with one another and in what order. It is a construct of a message sequence chart. A sequence diagram shows object interactions arranged in time sequence. From above mentioned sequence diagram we have to go in sequence: Enter the needed details as shown in the above figure, provide the Amyloidosis data set containing the aversion food items and then calculate the accuracies of Gradient Boosting & Neural Net along with the predictions.

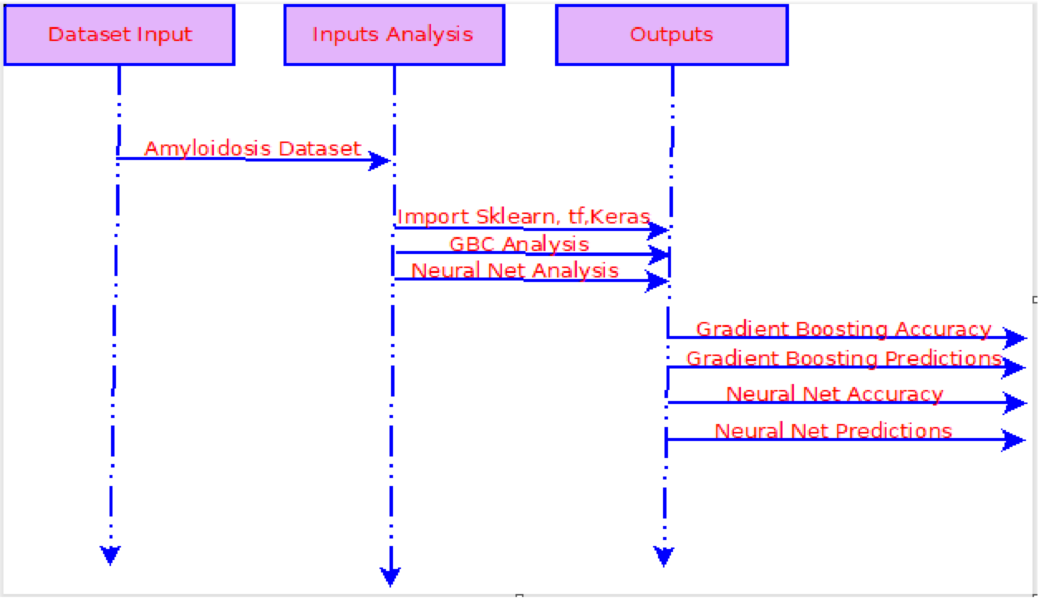


Fig 3.3.10 Sequence Diagram

3.3.11. Activity Diagram

Activity diagram is another important diagram in UML to describe dynamic aspects of the system. Activity diagram is basically a flow chart to represent the flow from one activity to another activity. The activity can be described as an operation of the system. So, the control flow is drawn from one operation to another. In activity diagram we can see that first provide the Amyloidosis dataset comprising of the aversion food items, and then train the model, test the model and calculate the accuracies along with the predictions of Gradient Boosting & Neural Net.

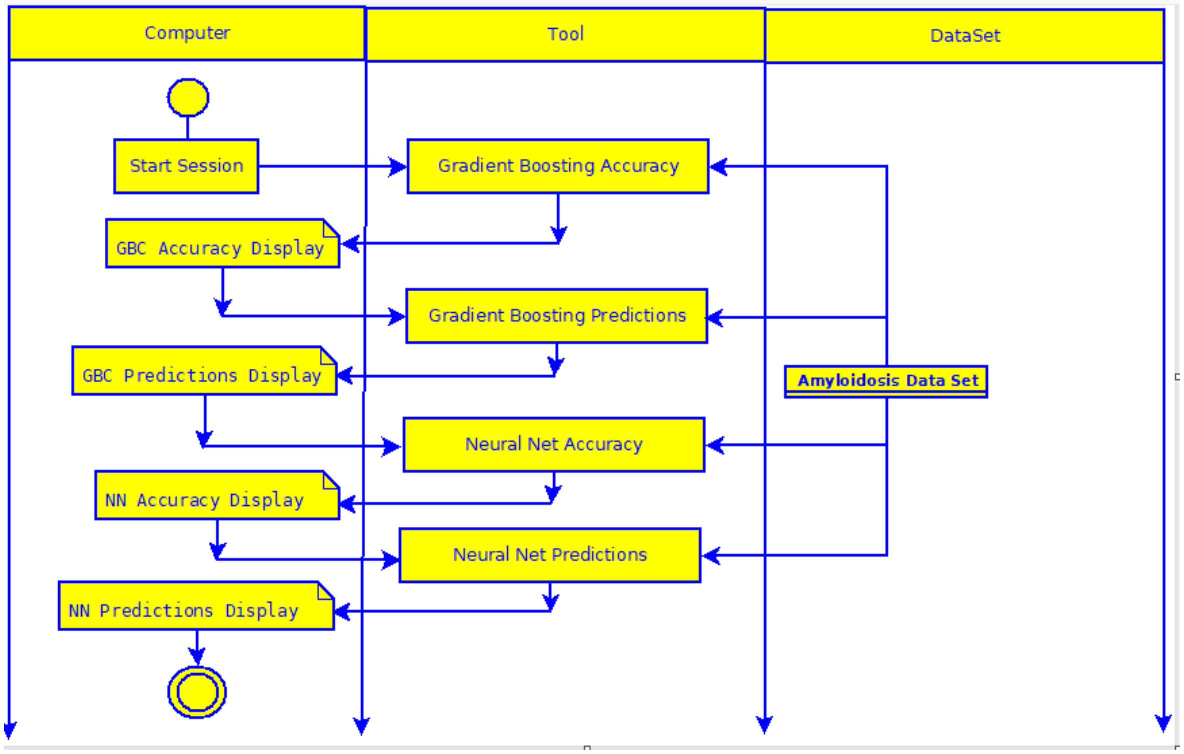


Fig 3.3.11 Activity Diagram

3.4. Stepwise Implementation and Code

GBC Accuracy

- Implements the Gradient Boosting Classifier (GBC) model.
- Preprocesses the data (mapping food items to numerical values).
- Splits the data into training and testing sets.
- Trains the GBC model on the training data.
- Evaluates the model's accuracy on the testing data.

GBC Predictions

- Loads the trained GBC model.
- Provides a function to get user input about food intake.
- Preprocesses the user input.
- Uses the trained model to make a prediction about amyloid protein levels.

Neural Network Accuracy

- Implements a neural network model using TensorFlow/Keras.
- Preprocesses the data (similar to GBC).
- Splits the data into training and testing sets.
- Defines the neural network architecture (layers, activation functions).
- Compiles the model (specifies loss function, optimizer, and metrics).
- Trains the neural network on the training data.
- Evaluates the model's accuracy on the testing data.

Neural Network Predictions

- Loads the trained neural network model.
- Provides functions to get and preprocess user input.
- Uses the trained model to make a prediction about amyloid protein levels.

Data Preprocessing (GBC & NN Accuracy)

- Import necessary libraries (NumPy, Pandas, scikit-learn, TensorFlow/Keras).
- Load the amloydset.csv dataset into a Pandas DataFrame.
- Map the categorical food items to numerical values (e.g., "LessThan6Servings" -> 1).
- Separate the input features (LeafyGreens to Berries) from the target variable (AmloydProteins).

Model Building

1. Gradient Boosting Classifier Accuracy

- Create a GradientBoostingClassifier object.
- Train the model using the fit method on the training data.
- Calculate accuracy using the .score method on the testing data.

2. Neural Network Accuracy

- Define a sequential model using tf.keras.models.Sequential.
- Add layers:
  - ➢ Flatten (if input isn't already 1D)
  - ➢ Dense layers with appropriate units and activation functions (e.g., 'relu')
  - ➢ A final Dense layer with units equal to the number of output classes and 'softmax' activation.
- Compile the model with an optimizer (e.g., 'adam'), loss function (e.g., 'sparse_categorical_crossentropy'), and metrics (e.g., 'accuracy').
- Train using fit on the training data.
- Evaluate accuracy on testing data.

3. Predictions (GBC & NN)

- Load the trained models.
- Define functions to:
  - ➢ Take food item intake as input from the user.
  - ➢ Preprocess the user input into the correct format.

- Use the model.predict method to generate predictions.
- Map the numerical prediction back to a label (e.g., 0 -> "LowAmloyds").

4. GUI

- Use PyQt5 to design the interface with buttons.
- Connect button clicks to functions that execute the other code files.

GUI CODE

```
import sys
import os
from amlyprotein import *
from PyQt5 import QtWidgets, QtGui, QtCore
class MyForm(QtWidgets.QMainWindow):
  def __init__(self,parent=None):
     QtWidgets.QWidget.__init__(self,parent)
     self.ui = Ui_MainWindow()
     self.ui.setupUi(self)
     self.ui.pushButton.clicked.connect(self.gbcacc)
     self.ui.pushButton_2.clicked.connect(self.nnacc)
     self.ui.pushButton_3.clicked.connect(self.nnpred)
     self.ui.pushButton_4.clicked.connect(self.gbcpred)
  def gbcacc(self):
    os.system("python -W ignore gbc1.py")
  def nnacc(self):
    os.system("python -W ignore nn1.py")
  def nnpred(self):
    os.system("python -W ignore nn2.py")
def gbcpred(self):
    os.system("python -W ignore gbc2.py")
if __name__ == "__main__":
    app = QtWidgets.QApplication(sys.argv)

    myapp = MyForm()
    myapp.show()
    sys.exit(app.exec_())
```

GRADIENT BOOSTING CLASSIFIER ACCURACY & PREDICTION CODE

```
import numpy as np
import pandas as pd
from sklearn import *
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.metrics import accuracy_score
df = pd.read_csv('amloydset.csv')
df["LeafyGreens"] = df["LeafyGreens"].map({'LessThan6Servings':1
,'6Servings':2,'MoreThan6Servings':3})
df["Champagne"] = df["Champagne"].map({'NoChampagne':1
,'Around3Glasses':2,'MoreThan6Glasses':3})
df["Vegetables"] = df["Vegetables"].map({'Rarely':1 ,'Regularly':2,'Daily':3})
df["Walnuts"] = df["Walnuts"].map({'LessThanHalfCup':1
,'HalfCup':2,'MoreThanHalfCup':3})
df["Onions"] = df["Onions"].map({'NotSpecific':1 ,'Regularly':2,'Compulsorily':3})
df["Coffee"] = df["Coffee"].map({'NoToLowCoffee':1 ,'Around2cups':2,'MoreThan2cups':3})
df["Turmeric"] = df["Turmeric"].map({'NotSpecific':1 ,'RegularUsage':2,'DailyUsage':3})
df["Cinnamon"] = df["Cinnamon"].map({'NoCinnamon':1
,'LowCinnamon':2,'CinnamonDaily':3})
df["FattyFish"] = df["FattyFish"].map({'NoToLow':1 ,'3to5Servings':2,'MoreThan5':3})
df["Berries"] = df["Berries"].map({'LessThan2cups':1 ,'3or4cups':2,'Above4cups':3})
df["AmloydProteins"] = df["AmloydProteins"].map({'LowAmloyds':0
,'NeededAmloyds':1,'MoreAmloyds':2})
data =
df[["LeafyGreens","Champagne","Vegetables","Walnuts","Onions","Coffee","Turmeric","Cin
namon","FattyFish","Berries","AmloydProteins"]].to_numpy()
inputs = data[:,:-1]
```

```
outputs = data[:, -1]
training_inputs = inputs[:600]
training_outputs = outputs[:600]
testing_inputs = inputs[600:]
testing_outputs = outputs[600:]
classifier = GradientBoostingClassifier()
classifier.fit(training_inputs, training_outputs)
predictions = classifier.predict(testing_inputs)
accuracy = accuracy_score(testing_outputs, predictions)
def get_user_input():
    food_data = {}
    # Prompt user for each food item
food_data["LeafyGreens"] = input("Enter Leafy Greens intake (1 - Less than 6 servings, 2 - 6 servings, 3 - More than 6 servings): ")
    food_data["Champagne"] = input("Enter Champagne intake (1 - No Champagne, 2 - Around 3 Glasses, 3 - More than 6 Glasses): ")
    food_data["Vegetables"] = input("Enter Vegetables intake (1 - Rarely, 2 - Regularly, 3 - Daily): ")
    food_data["Walnuts"] = input("Enter Walnuts intake (1 - LessThanHalfCup, 2 - HalfCup, 3 - MoreThanHalfCup): ")
    food_data["Onions"] = input("Enter Onions intake (1 - NotSpecific, 2 - Regularly, 3 - Compulsorily): ")
    food_data["Coffee"] = input("Enter Coffee intake (1 - NoToLowCoffee, 2 - Around2cups, 3 - MoreThan2cups): ")
    food_data["Turmeric"] = input("Enter Turmeric intake (1 - NotSpecific, 2 - RegularUsage, 3 - DailyUsage): ")
    food_data["Cinnamon"] = input("Enter Cinnamon intake (1 - NoCinnamon, 2 - LowCinnamon, 3 - CinnamonDaily): ")
    food_data["FattyFish"] = input("Enter FattyFish intake (1 - NoToLow, 2 - 3to5Servings, 3 - MoreThan5): ")
```

```
food_data["Berries"] = input("Enter Berries intake (1 - LessThan2cups, 2 - 3or4cups, 3 -
Above4cups): ")
    return food_data
food_data["LeafyGreens"] = input("Enter Leafy Greens intake (1 - Less than 6 servings, 2 - 6
servings, 3 - More than 6 servings): ")
    food_data["Champagne"] = input("Enter Champagne intake (1 - No Champagne, 2 -
Around 3 Glasses, 3 - More than 6 Glasses): ")
    food_data["Vegetables"] = input("Enter Vegetables intake (1 - Rarely, 2 - Regularly, 3 -
Daily): ")
    food_data["Walnuts"] = input("Enter Walnuts intake (1 - LessThanHalfCup, 2 - HalfCup, 3
- MoreThanHalfCup): ")
    food_data["Onions"] = input("Enter Onions intake (1 - NotSpecific, 2 - Regularly, 3 -
Compulsorily): ")
    food_data["Coffee"] = input("Enter Coffee intake (1 - NoToLowCoffee, 2 - Around2cups, 3
- MoreThan2cups): ")
    food_data["Turmeric"] = input("Enter Turmeric intake (1 - NotSpecific, 2 - RegularUsage,
3 - DailyUsage): ")
    food_data["Cinnamon"] = input("Enter Cinnamon intake (1 - NoCinnamon, 2 -
LowCinnamon, 3 - CinnamonDaily): ")
    food_data["FattyFish"] = input("Enter FattyFish intake (1 - NoToLow, 2 - 3to5Servings, 3 -
MoreThan5): ")
    food_data["Berries"] = input("Enter Berries intake (1 - LessThan2cups, 2 - 3or4cups, 3 -
Above4cups): ")
    return food_data
def get_user_input_as_datapoint(user_data):
    mapping_dictionary = {
        "LeafyGreens": {"1": 1, "2": 2, "3": 3},  # Direct numerical mapping
        "Champagne": {"1": 1, "2": 2, "3": 3},
        "Vegetables": {'1': 1, '2': 2, '3': 3},
        "Walnuts": {'1': 1, '2': 2, '3': 3},
        "Onions": {'1': 1, '2': 2, '3': 3},
        "Coffee": {'1': 1, '2': 2, '3': 3},
        "Turmeric": {'1': 1, '2': 2, '3': 3},
        "Cinnamon": {'1': 1, '2': 2, '3': 3},
        "FattyFish": {'1': 1, '2': 2, '3': 3},
        "Berries": {'1': 1, '2': 2, '3': 3},
    }
    numerical_data = []
    for key, value in user_data.items():
        numerical_data.append(mapping_dictionary[key][value])
    return np.array([numerical_data])
def predict_gbc(user_data):
    datapoint = get_user_input_as_datapoint(user_data)
    prediction = classifier.predict(datapoint)[0]
    amloyds_mapping = {0: "LowAmloyds", 1: "NeededAmloyds", 2: "MoreAmloyds"}
    prediction_label = amloyds_mapping[prediction]
    print("GBC Model Prediction:", prediction_label)
if __name__ == "__main__":
    classifier.fit(training_inputs, training_outputs)
    predictions = classifier.predict(testing_inputs)
    accuracy = accuracy_score(testing_outputs, predictions)
    print("Gradient Boosting Classifier Accuracy:", accuracy)
    user_data = get_user_input()
    predict_gbc(user_data)
```

NEURAL NET CLASSIFIER ACCURACY & PREDICTION CODE

```
import numpy as np
import pandas as pd
import tensorflow as tf
import tensorflow.keras.backend as K
print(tf.__version__)
import warnings
```

```
warnings.filterwarnings("ignore")
df = pd.read_csv('amloydset.csv')
df["LeafyGreens"] = df["LeafyGreens"].map({'LessThan6Servings': 1, '6Servings': 2,
'MoreThan6Servings': 3})
df["Champagne"] = df["Champagne"].map({'NoChampagne': 1, 'Around3Glasses': 2,
'MoreThan6Glasses': 3})
df["Vegetables"] = df["Vegetables"].map({'Rarely': 1, 'Regularly': 2, 'Daily': 3})
df["Walnuts"] = df["Walnuts"].map({'LessThanHalfCup': 1, 'HalfCup': 2, 'MoreThanHalfCup':
3})
df["Onions"] = df["Onions"].map({'NotSpecific': 1, 'Regularly': 2, 'Compulsorily': 3})
df["Coffee"] = df["Coffee"].map({'NoToLowCoffee': 1, 'Around2cups': 2, 'MoreThan2cups':
3})
df["Turmeric"] = df["Turmeric"].map({'NotSpecific': 1, 'RegularUsage': 2, 'DailyUsage': 3})
df["Cinnamon"] = df["Cinnamon"].map({'NoCinnamon': 1, 'LowCinnamon': 2,
'CinnamonDaily': 3})
df["FattyFish"] = df["FattyFish"].map({'NoToLow': 1, '3to5Servings': 2, 'MoreThan5': 3})
df["Berries"] = df["Berries"].map({'LessThan2cups': 1, '3or4cups': 2, 'Above4cups': 3})
df["AmloydProteins"] = df["AmloydProteins"].map({'LowAmloyds': 0, 'NeededAmloyds': 1,
'MoreAmloyds': 2})
data = df[["LeafyGreens", "Champagne", "Vegetables", "Walnuts", "Onions", "Coffee",
"Turmeric", "Cinnamon", "FattyFish", "Berries", "AmloydProteins"]].to_numpy()
outputs = data[:, -1]
training_data = inputs[:600]
training_labels = outputs[:600]
test_data = inputs[600:]
test_labels = outputs[600:]
tf.keras.backend.clear_session()
model = tf.keras.models.Sequential([
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(128, activation=tf.nn.relu),
    tf.keras.layers.Dense(64, activation=tf.nn.relu),
    tf.keras.layers.Dense(32, activation=tf.nn.relu),
    tf.keras.layers.Dense(10, activation=tf.nn.softmax)
])
model.compile(loss='sparse_categorical_crossentropy', optimizer='adam',
metrics=['accuracy'])
model.fit(training_data, training_labels, epochs=150)
def get_user_input():
    food_data = {}
    food_data["LeafyGreens"] = input("Enter Leafy Greens intake (1 - Less than 6 servings, 2 -
6 servings, 3 - More than 6 servings): ")
    food_data["Champagne"] = input("Enter Champagne intake (1 - No Champagne, 2 -
Around 3 Glasses, 3 - More than 6 Glasses): ")
    food_data["Vegetables"] = input("Enter Vegetables intake (1 - Rarely, 2 - Regularly, 3 -
Daily): ")
    food_data["Walnuts"] = input("Enter Walnuts intake (1 - LessThanHalfCup, 2 - HalfCup, 3
- MoreThanHalfCup): ")
    food_data["Onions"] = input("Enter Onions intake (1 - NotSpecific, 2 - Regularly, 3 -
Compulsorily): ")
    food_data["Coffee"] = input("Enter Coffee intake (1 - NoToLowCoffee, 2 - Around2cups, 3
- MoreThan2cups): ")
    food_data["Turmeric"] = input("Enter Turmeric intake (1 - NotSpecific, 2 - RegularUsage,
3 - DailyUsage): ")
    food_data["Cinnamon"] = input("Enter Cinnamon intake (1 - NoCinnamon, 2 -
LowCinnamon, 3 - CinnamonDaily): ")
    food_data["FattyFish"] = input("Enter FattyFish intake (1 - NoToLow, 2 - 3to5Servings, 3 -
MoreThan5): ")
    food_data["Berries"] = input("Enter Berries intake (1 - LessThan2cups, 2 - 3or4cups, 3 -
Above4cups): ")
    return food_data
def get_user_input_as_datapoint(user_data):
    mapping_dictionary = {
        "LeafyGreens": {"1": 1, "2": 2, "3": 3},
        "Champagne": {"1": 1, "2": 2, "3": 3},
        "Vegetables": {'1': 1, '2': 2, '3': 3},
        "Walnuts": {'1': 1, '2': 2, '3': 3},
        "Onions": {'1': 1, '2': 2, '3': 3},
        "Coffee": {'1': 1, '2': 2, '3': 3},
        "Turmeric": {'1': 1, '2': 2, '3': 3},
        "Cinnamon": {'1': 1, '2': 2, '3': 3},
        "FattyFish": {'1': 1, '2': 2, '3': 3},
        "Berries": {'1': 1, '2': 2, '3': 3},
    }
    numerical_data = []
    for key, value in user_data.items():
        numerical_data.append(mapping_dictionary[key][value])
    return np.array([numerical_data])
def predict_nn(user_data):
    datapoint = get_user_input_as_datapoint(user_data)
    prediction = np.argmax(model.predict(datapoint), axis=-1)[0]
    amloyds_mapping = {0: "LowAmloyds", 1: "NeededAmloyds", 2: "MoreAmloyds"}
    prediction_label = amloyds_mapping[prediction]
    print("Neural Network Model Prediction:", prediction_label)
if __name__ == "__main__":
    test_loss, test_acc = model.evaluate(test_data, test_labels)
    print("\nNeural Network Test Accuracy:", test_acc)
    user_data = get_user_input()
    predict_nn(user_data)
```

CHAPTER 4

RESULTS AND DISCUSSION

CHAPTER 4

RESULTS AND DISCUSSION

This section showcases our project's functionality through a series of steps and accompanying visuals.

Executing the Application:

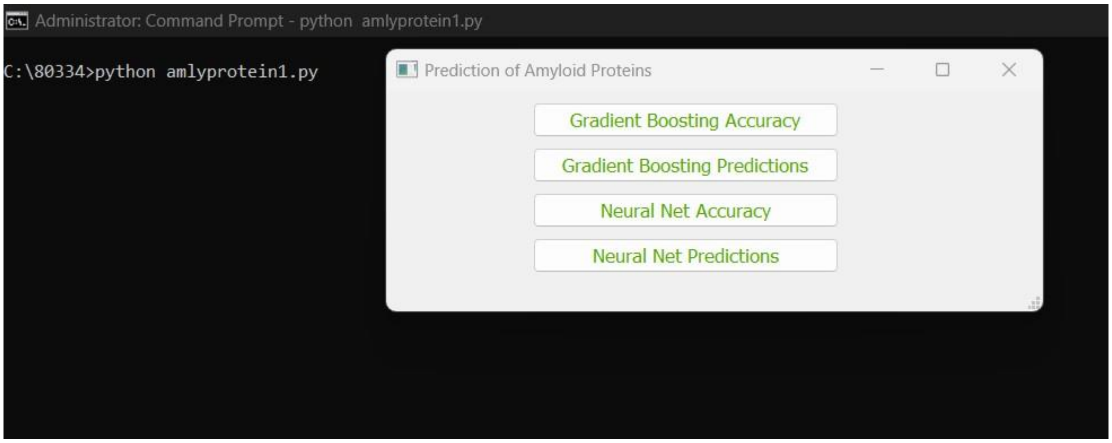
1. Launch the Command Prompt as Administrative.

2. Navigate to the Project Directory: This step instructs the program on location of the necessary files.

3. Run the File amlyproteinl1.py: Executing this file initiates the application.

Upon program execution, the "Amyloid Proteins Prediction Tool" window launches as shown in Fig:4.0.1. This interface comprises four distinct modules:

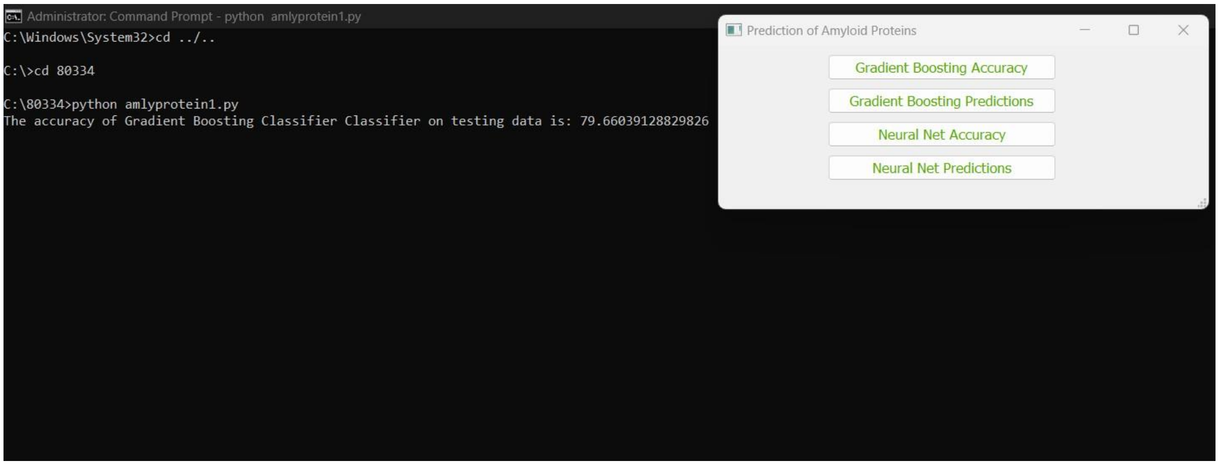
- **Gradient Boosting Accuracy:** This section displays the model's overall accuracy in predicting amyloid protein level as shown in Fig: 4.0.2.
- **Gradient Boosting Prediction:** This module showcases the model's predicted amyloid protein levels of user as shown in Fig: 4.0.3.
- **Neural Network Accuracy:** This section presents the accuracy of the neural network model in predicting amyloid protein levels as shown in Fig: 4.0.4.
- **Neural Network Prediction:** This module displays the neural network model's predicted amyloid protein levels of user as shown in Fig: 4.0.5.



```
Administrator: Command Prompt - python  amlyprotein1.py
C:\80334>python amlyprotein1.py
```

Prediction of Amyloid Proteins
Gradient Boosting Accuracy
Gradient Boosting Predictions
Neural Net Accuracy
Neural Net Predictions

Fig 4.0.1 Amyloid Protein Prediction Tool



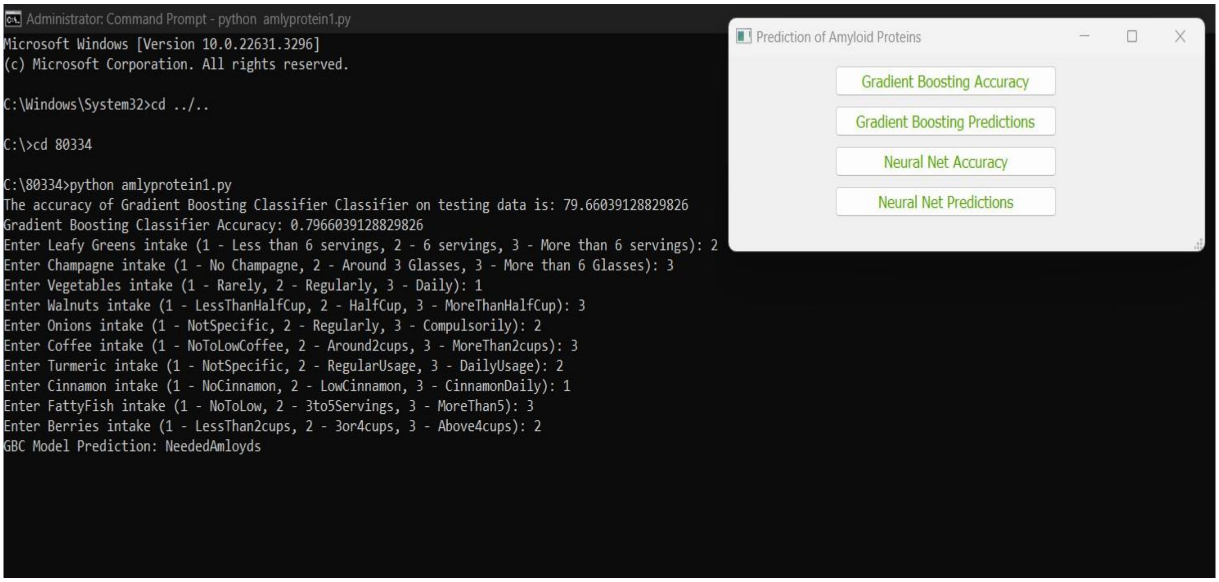
Administrator: Command Prompt - python amlyprotein1.py
Prediction of Amyloid Proteins
Gradient Boosting Accuracy
Gradient Boosting Predictions
Neural Net Accuracy
Neural Net Predictions

```
C:\Windows\System32>cd ../..

C:\>cd 80334

C:\80334>python amlyprotein1.py
The accuracy of Gradient Boosting Classifier Classifier on testing data is: 79.66039128829826
```

Fig 4.0.2 GBC Accuracy



Administrator: Command Prompt - python amlyprotein1.py
Prediction of Amyloid Proteins
Gradient Boosting Accuracy
Gradient Boosting Predictions
Neural Net Accuracy
Neural Net Predictions

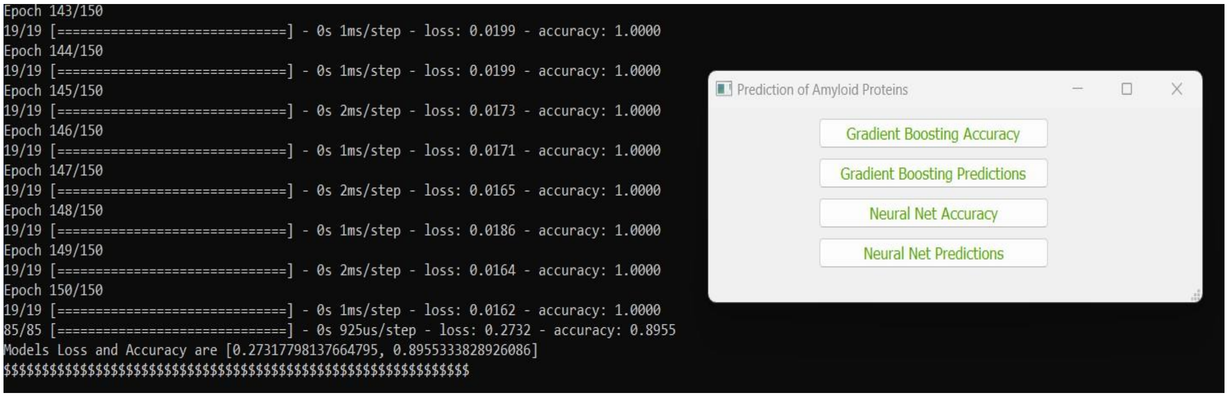
```
Microsoft Windows [Version 10.0.22631.3296]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\System32>cd ../..

C:\>cd 80334

C:\80334>python amlyprotein1.py
The accuracy of Gradient Boosting Classifier Classifier on testing data is: 79.66039128829826
Gradient Boosting Classifier Accuracy: 0.7966039128829826
Enter Leafy Greens intake (1 - Less than 6 servings, 2 - 6 servings, 3 - More than 6 servings): 2
Enter Champagne intake (1 - No Champagne, 2 - Around 3 Glasses, 3 - More than 6 Glasses): 3
Enter Vegetables intake (1 - Rarely, 2 - Regularly, 3 - Daily): 1
Enter Walnuts intake (1 - LessThanHalfCup, 2 - HalfCup, 3 - MoreThanHalfCup): 3
Enter Onions intake (1 - NotSpecific, 2 - Regularly, 3 - Compulsorily): 2
Enter Coffee intake (1 - NoToLowCoffee, 2 - Around2cups, 3 - MoreThan2cups): 3
Enter Turmeric intake (1 - NotSpecific, 2 - RegularUsage, 3 - DailyUsage): 2
Enter Cinnamon intake (1 - NoCinnamon, 2 - LowCinnamon, 3 - CinnamonDaily): 1
Enter FattyFish intake (1 - NoToLow, 2 - 3to5Servings, 3 - MoreThan5): 3
Enter Berries intake (1 - LessThan2cups, 2 - 3or4cups, 3 - Above4cups): 2
GBC Model Prediction: NeededAmloyds
```

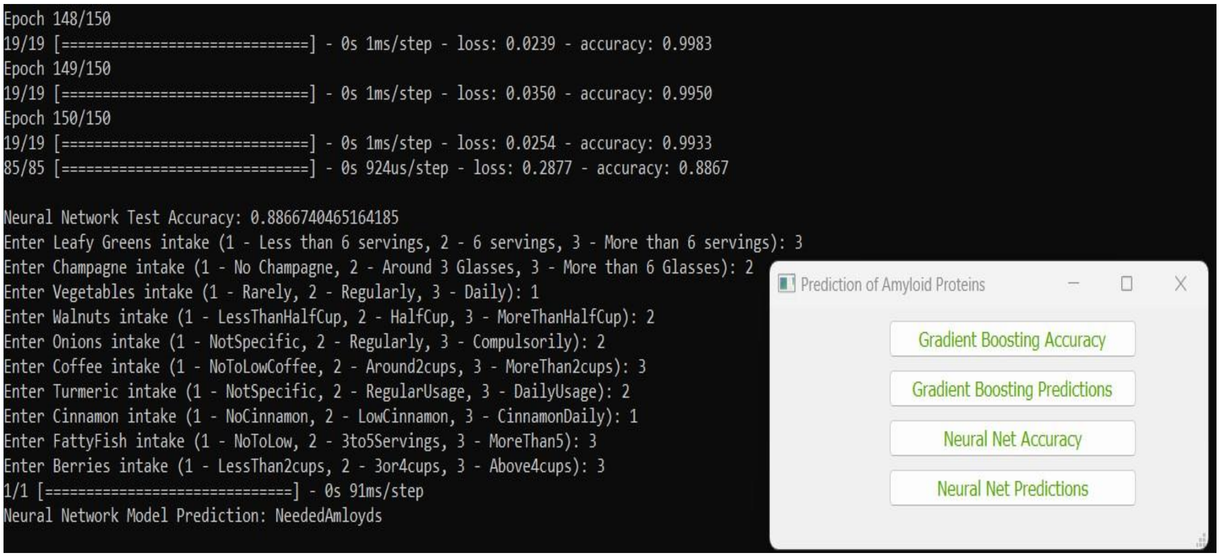
Fig 4.0.3 GBC Prediction



Prediction of Amyloid Proteins
Gradient Boosting Accuracy
Gradient Boosting Predictions
Neural Net Accuracy
Neural Net Predictions

```
Epoch 143/150
19/19 [==============================] - 0s 1ms/step - loss: 0.0199 - accuracy: 1.0000
Epoch 144/150
19/19 [==============================] - 0s 1ms/step - loss: 0.0199 - accuracy: 1.0000
Epoch 145/150
19/19 [==============================] - 0s 2ms/step - loss: 0.0173 - accuracy: 1.0000
Epoch 146/150
19/19 [==============================] - 0s 1ms/step - loss: 0.0171 - accuracy: 1.0000
Epoch 147/150
19/19 [==============================] - 0s 2ms/step - loss: 0.0165 - accuracy: 1.0000
Epoch 148/150
19/19 [==============================] - 0s 1ms/step - loss: 0.0186 - accuracy: 1.0000
Epoch 149/150
19/19 [==============================] - 0s 2ms/step - loss: 0.0164 - accuracy: 1.0000
Epoch 150/150
19/19 [==============================] - 0s 1ms/step - loss: 0.0162 - accuracy: 1.0000
85/85 [==============================] - 0s 925us/step - loss: 0.2732 - accuracy: 0.8955
Models Loss and Accuracy are [0.27317798137664795, 0.8955333828926086]
$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
```

Fig 4.0.4 Neural Net Accuracy



```
Epoch 148/150
19/19 [==============================] - 0s 1ms/step - loss: 0.0239 - accuracy: 0.9983
Epoch 149/150
19/19 [==============================] - 0s 1ms/step - loss: 0.0350 - accuracy: 0.9950
Epoch 150/150
19/19 [==============================] - 0s 1ms/step - loss: 0.0254 - accuracy: 0.9933
85/85 [==============================] - 0s 924us/step - loss: 0.2877 - accuracy: 0.8867

Neural Network Test Accuracy: 0.8866740465164185
Enter Leafy Greens intake (1 - Less than 6 servings, 2 - 6 servings, 3 - More than 6 servings): 3
Enter Champagne intake (1 - No Champagne, 2 - Around 3 Glasses, 3 - More than 6 Glasses): 2
Enter Vegetables intake (1 - Rarely, 2 - Regularly, 3 - Daily): 1
Enter Walnuts intake (1 - LessThanHalfCup, 2 - HalfCup, 3 - MoreThanHalfCup): 2
Enter Onions intake (1 - NotSpecific, 2 - Regularly, 3 - Compulsorily): 2
Enter Coffee intake (1 - NoToLowCoffee, 2 - Around2cups, 3 - MoreThan2cups): 3
Enter Turmeric intake (1 - NotSpecific, 2 - RegularUsage, 3 - DailyUsage): 2
Enter Cinnamon intake (1 - NoCinnamon, 2 - LowCinnamon, 3 - CinnamonDaily): 1
Enter FattyFish intake (1 - NoToLow, 2 - 3to5Servings, 3 - MoreThan5): 3
Enter Berries intake (1 - LessThan2cups, 2 - 3or4cups, 3 - Above4cups): 3
1/1 [==============================] - 0s 91ms/step
Neural Network Model Prediction: NeededAmloyds
```

Prediction of Amyloid Proteins
Gradient Boosting Accuracy
Gradient Boosting Predictions
Neural Net Accuracy
Neural Net Predictions

Fig 4.0.5 Neural Net Prediction

4.1. Performance Metrics

This section will evaluate the effectiveness of **Neural Network (NN) model** in predicting amyloid protein levels. The following metrics are used:

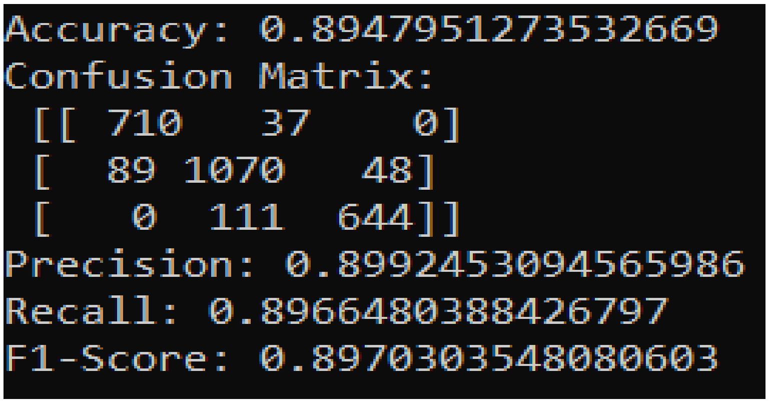
- **Accuracy:** 0.8948 (89.48%) - This indicates that model is approximately 89.48% accurate in its predictions on the testing data.

- **Confusion Matrix:** The confusion matrix shows the counts of true positive, true negative, false positive, and false negative predictions. A table that breaks down correct and incorrect predictions by class (LowAmloyds, NeededAmloyds, MoreAmloyds). The confusion matrix highlights specific areas where either model may be struggling.

- **Precision:** 0.8992 (89.92%) - Precision is the ratio of correctly predicted positive observations to the total predicted positives. It's a measure of the accuracy of the positive predictions. A precision of 0.8992 means that about 89.92% of the predictions labeled as positive are actually correct.

- **Recall:** 0.8966 (89.66%) - Recall, also known as sensitivity or true positive rate, measures the ratio of correctly predicted positive observations to all actual positives. It indicates the model's ability to find all relevant cases within a dataset. A recall of 0.8966 means that about 89.66% of actual positives were correctly identified by the model.

- **F1-Score:** 0.8970 (89.70%) - The F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall and is often used as a single metric to evaluate a model's performance. A higher F1-score indicates better model performance. In this case, your model achieved an F1-score of 0.8970, which is quite good.



```
Accuracy: 0.8947951273532669
Confusion Matrix:
 [[ 710   37    0]
 [  89 1070   48]
 [   0  111  644]]
Precision: 0.8992453094565986
Recall: 0.8966480388426797
F1-Score: 0.8970303548080603
```

Fig 4.1.1 Performance Metrics of Neural Net

This section will evaluate the effectiveness of **Gradient Boosting Classifier (GBC) model** in predicting amyloid protein levels. The following metrics are used:

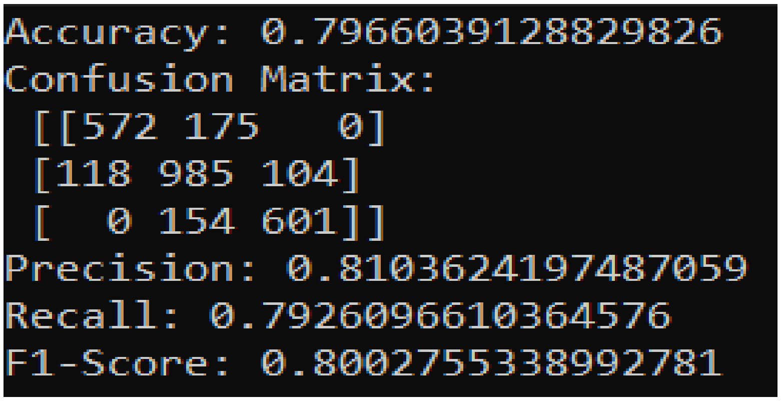
- **Accuracy:** 0.7966 (79.66%) - This indicates that the model is approximately 79.66% accurate in its predictions on the testing data.

- **Confusion Matrix:** The confusion matrix shows the counts of true positive, true negative, false positive, and false negative predictions. It breaks down correct and incorrect predictions by class (LowAmloyds, NeededAmloyds, MoreAmloyds). The matrix highlights specific areas where either the model may be struggling.

- **Precision:** 0.8104 (81.04%) - Precision is the ratio of correctly predicted positive observations to the total predicted positives. It's a measure of the accuracy of the positive predictions. A precision of 0.8104 means that about 81.04% of the predictions labeled as positive are actually correct.

- **Recall:** 0.7926 (79.26%) - Recall, also known as sensitivity or true positive rate, measures the ratio of correctly predicted positive observations to all actual positives. It indicates the model's ability to find all relevant cases within a dataset. A recall of 0.7926 means that about 79.26% of actual positives were correctly identified by the model.

- **F1-Score:** 0.8003 (80.03%) - The F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall and is often used as a single metric to evaluate a model's performance. A higher F1-score indicates better model performance. In this case, your model achieved an F1-score of 0.8003, which is quite good.



```
Accuracy: 0.7966039128829826
Confusion Matrix:
 [[572 175   0]
 [118 985 104]
 [  0 154 601]]
Precision: 0.8103624197487059
Recall: 0.7926096610364576
F1-Score: 0.8002755338992781
```

Fig 4.1.2 Performance of Gradient Boosting Classifier

4.2. Comparison of accuracies with other classifiers

What is Accuracy?

Accuracy is a crucial metric in evaluating the performance and reliability of predictive models, especially in endeavors like the prediction of amyloid proteins using gradient boosting classifiers (GBC) and neural networks. In this context, accuracy refers to the degree of agreement between the predictions made by the models and the actual outcomes or ground truth. It serves as a fundamental indicator of how effectively the models are able to discern patterns and make correct classifications based on the provided data.

For the prediction of amyloid proteins, accuracy holds significant importance due to the severe implications of amyloid-related diseases such as Alzheimer's, Parkinson’s, and type 2 diabetes. Understanding the accuracy of predictive models aids in assessing their potential utility in identifying dietary components that may influence the deposition of amyloid proteins, thereby contributing to preventive healthcare strategies.

In the context of this project, accuracy is measured through various methodologies, primarily focusing on the performance of the gradient boosting classifier (GBC) and neural network models. The GBC model is evaluated in terms of its ability to correctly classify instances of amyloid proteins based on features extracted from a dataset comprising various food items known to influence amyloid deposition.

Accuracy assessment involves comparing the model's predictions against known outcomes, typically using techniques such as cross-validation or holdout validation, to ensure robustness and generalizability. A high accuracy score indicates that the model can effectively differentiate between amyloid-associated and non-associated instances, providing confidence in its predictive capabilities.

However, it's essential to interpret accuracy in the context of the specific domain and dataset. In the prediction of amyloid proteins, the complexity of biological processes and the multifaceted interactions between dietary factors and disease pathways may present challenges in achieving high accuracy.

Furthermore, accuracy alone may not provide a comprehensive evaluation of model performance, especially in scenarios where class imbalances or misclassification costs are prevalent. Therefore, complementary metrics such as precision, recall, F1-score, and receiver operating characteristic (ROC) curves are often employed to provide a more nuanced understanding of model behavior.

In summary, accuracy serves as a foundational metric in assessing the efficacy of predictive models for amyloid protein prediction. Through rigorous evaluation and interpretation, it facilitates informed decision-making regarding the suitability of these models for practical applications in preventive healthcare and disease management.

Based on the below graph, which includes a list of classifiers with our model compared to the other existing classifiers. Accuracy is a common metric used to evaluate the performance of a classifier. It is defined as the ratio of the number of correct predictions to the total number of input samples. The accuracy score of each classifier is given for a range of scores from 0.0 to 1.0. The classifier with the highest accuracy point is the classifier considered to be most accurate that is, Gradient Boosting classifier is highest among all the classifiers.

Mathematically, the formula for accuracy can be expressed as:

Accuracy = Number of Correct Predictions / Total Number of Predictions

Where:

- Number of Correct Predictions: The total count of correctly predicted words in all captions.
- Total Number of Predictions: The total count of words in all captions

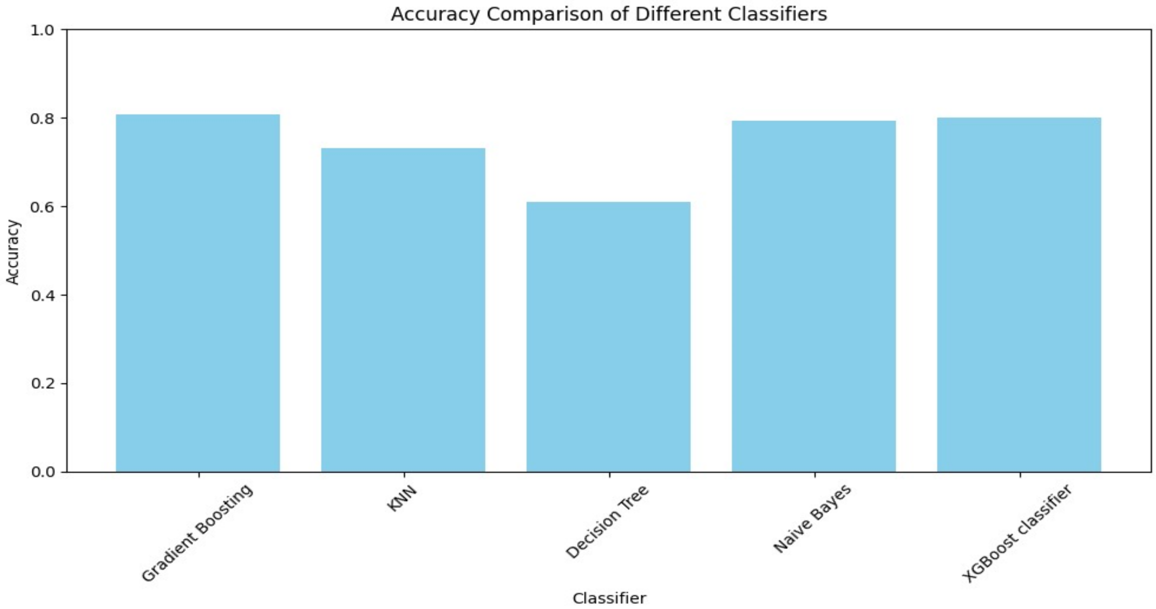
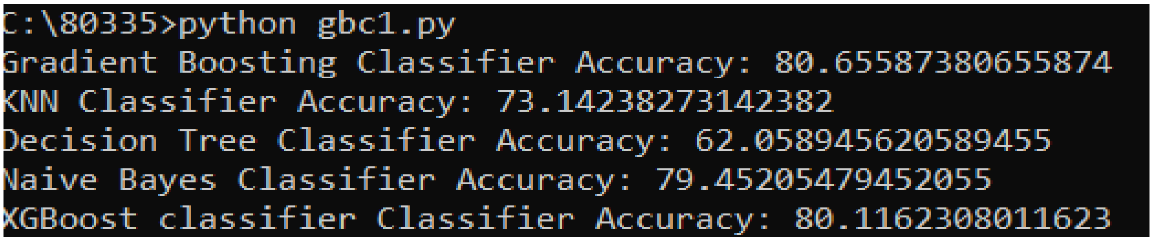


Fig 4.2.2 Accuracy Comparison Graph



```
C:\80335>python gbc1.py
Gradient Boosting Classifier Accuracy: 80.65587380655874
KNN Classifier Accuracy: 73.14238273142382
Decision Tree Classifier Accuracy: 62.058945620589455
Naive Bayes Classifier Accuracy: 79.45205479452055
XGBoost classifier Classifier Accuracy: 80.1162308011623
```

Fig 4.2.3 Accuracy comparison

CHAPTER 5
CONCLUSION

CHAPTER 5
CONCLUSION

Our study successfully developed and evaluated two machine learning models: Gradient Boosting Classifier (GBC) and a Neural Network. These models aim to predict the quantity of amyloid protein in various foods, including common ones like onions, kale, and berries. Both models performed well on entirely new data, suggesting they might possess the ability to identify foods linked to various levels of amyloid proteins.

The GBC model attained a precision of 79.6%, while the Neural Network reached 89%. Notably, both models could predict the amyloid protein levels in unseen food items.

Importantly, the models performed well on previously unseen food items. This suggests they have learned to identify broader dietary trends linked to amyloid protein levels, rather than simply memorizing specific examples. Our work highlights the value of exploring machine learning approaches within the context of disease prevention and health promotion.

Overall, this project contributes to the field of using ML (machine learning) to analyze dietary patterns. It makes room for more research into the connection between specific foods and the accumulation of amyloid proteins. This research possesses the potential to educate dietary strategies for individuals worried about health problems associated with amyloid buildup.

5.1 Future Enhancement

- **Larger Dataset:** Expanding and diversifying your dataset will improve model robustness and ability to generalize to new situations. Consider acquiring more data points or using data augmentation techniques.

- **Feature Engineering:**
  - Explore additional features relevant to amyloid protein deposition (e.g., age, genetic markers, other lifestyle factors).
  - Experiment with feature selection techniques to identify the most impactful food items and combinations.

- **Hyperparameter Tuning:** Optimize the performance of both Gradient Boosting Classifier and Neural Network models by fine-tuning hyperparameters (e.g., learning rate, number of trees, number of neurons). Methods like Grid Search or Random Search can help with this task.

- **Alternative Algorithms:** Investigate other advanced machine learning algorithms such as:
  - Random Forest for potentially improved accuracy.
  - Support Vector Machines (SVM) for handling smaller datasets.

- **Ensemble Methods:** Combine predictions from multiple models (e.g., GBC, NN, others) using voting or averaging techniques. This often leads to better overall results.

- **Interpretability:** Employ techniques like SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-Agnostic Explanations) to understand how each food item impacts the model's predictions. This is crucial for providing actionable dietary advice.

- **User Interface Improvement:** Develop a more sophisticated and user-friendly interface for your application. This could include:
    - ➢ Data visualizations representing intake patterns and outcomes.
    - ➢ Personalized dietary recommendations based on predictions.

- **Clinical Validation:** Collaborate with healthcare professionals to validate model findings from a clinical perspective and test the potential for real-world application in disease prevention.

REFERENCES

REFERENCES

[1] Q.Sun, W. Kong, X. Mou, and S. Wang, ‘‘Transcriptional regulation analysis of Alzheimer’s disease based on FastNCA algorithm,’’ Current Bioinf., vol. (14), no. 8, pp. 771–782, Dec. 2019.

[2] Ishahid Akbar, Hashim Ali, Ashfaq Ahmad, Mahidur R. Sarker,Aamir Saeed, Ely Salwana, Sarah Gul, Ahmad Khan1 And Farman AL: Prediction of Amyloid Proteins Using Embedded Evolutionary & Ensemble Feature Selection Based Descriptors With eXtreme Gradient Boosting Model, IEEE Access

[3] P. Lembré, C. Vendrely, and P. Martino, ‘‘Identification of an amyloidogenic peptide from the Bap protein of Staphylococcus epidermidis,’’ Protein Peptide Lett., vol. 21, no. (1), pp. 75–79

[4] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, ‘‘Predicting diabetes mellitus (dm) with ML(machine learning) methods,’’ Frontiers Genet., vol. 9,p. 515..

[5] https://www.python.org/

[6] https://github.com/baoboa/pyqt5/blob/master/pyuic/uic/pyuic.py

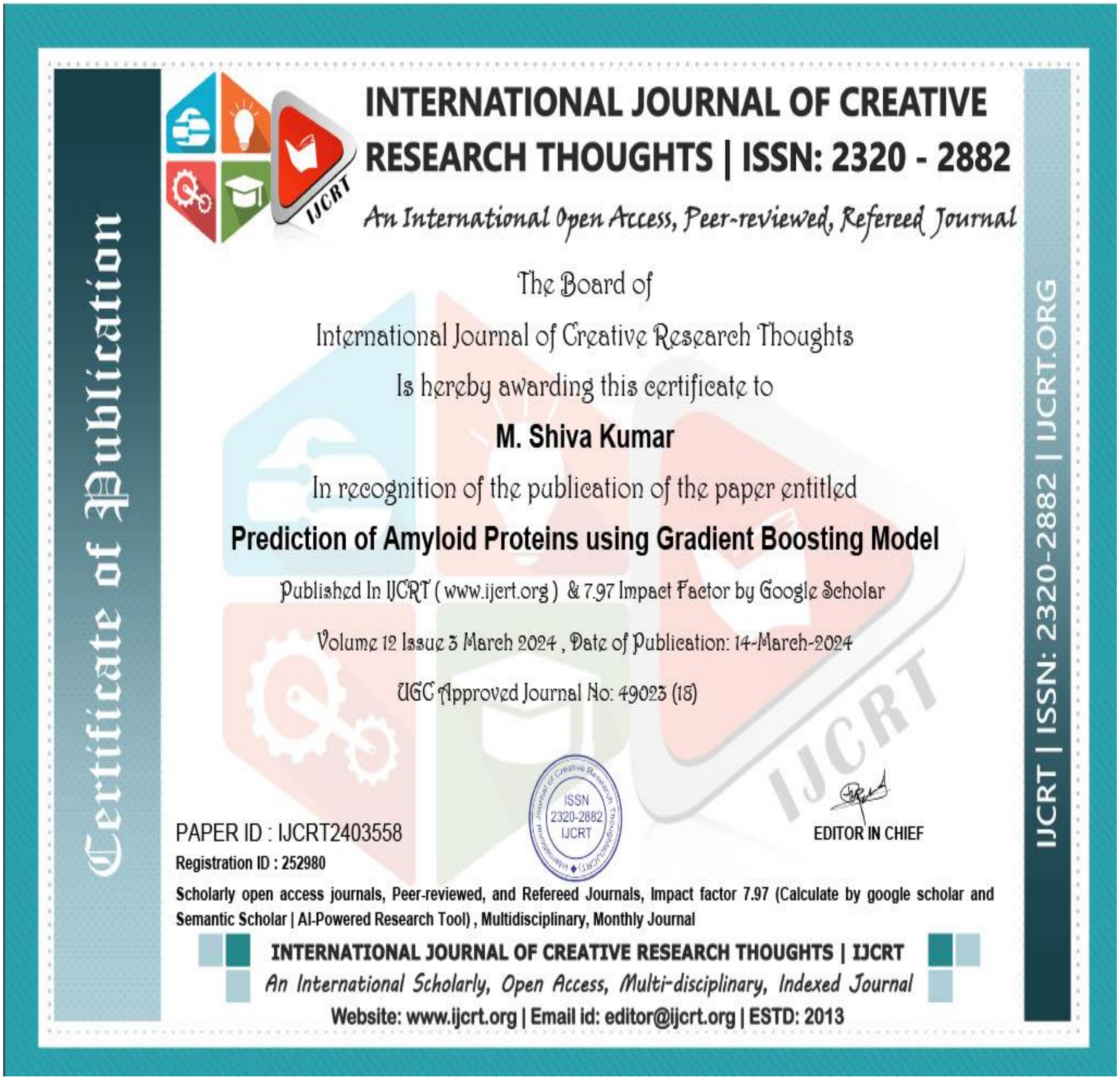
[7] https://www.numpy.org/

[8] https://riverbankcomputing.com/software/pyqt/intro

GITHUB LINK

https://github.com/AdityaVardhan754/Major-Project-Phase-2

PUBLICATION CERTIFICATE



Certificate of Publication

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS | ISSN: 2320 - 2882

An International Open Access, Peer-reviewed, Refereed Journal

The Board of

International Journal of Creative Research Thoughts

Is hereby awarding this certificate to

M. Shiva Kumar

In recognition of the publication of the paper entitled

Prediction of Amyloid Proteins using Gradient Boosting Model

Published In IJCRT (www.ijcrt.org) & 7.97 Impact Factor by Google Scholar

Volume 12 Issue 3 March 2024 , Date of Publication: 14-March-2024

UGC Approved Journal No: 49023 (18)

PAPER ID : IJCRT2403558

Registration ID : 252980

International Journal of Creative Research Thoughts (IJCRT)

ISSN 2320-2882 IJCRT

EDITOR IN CHIEF

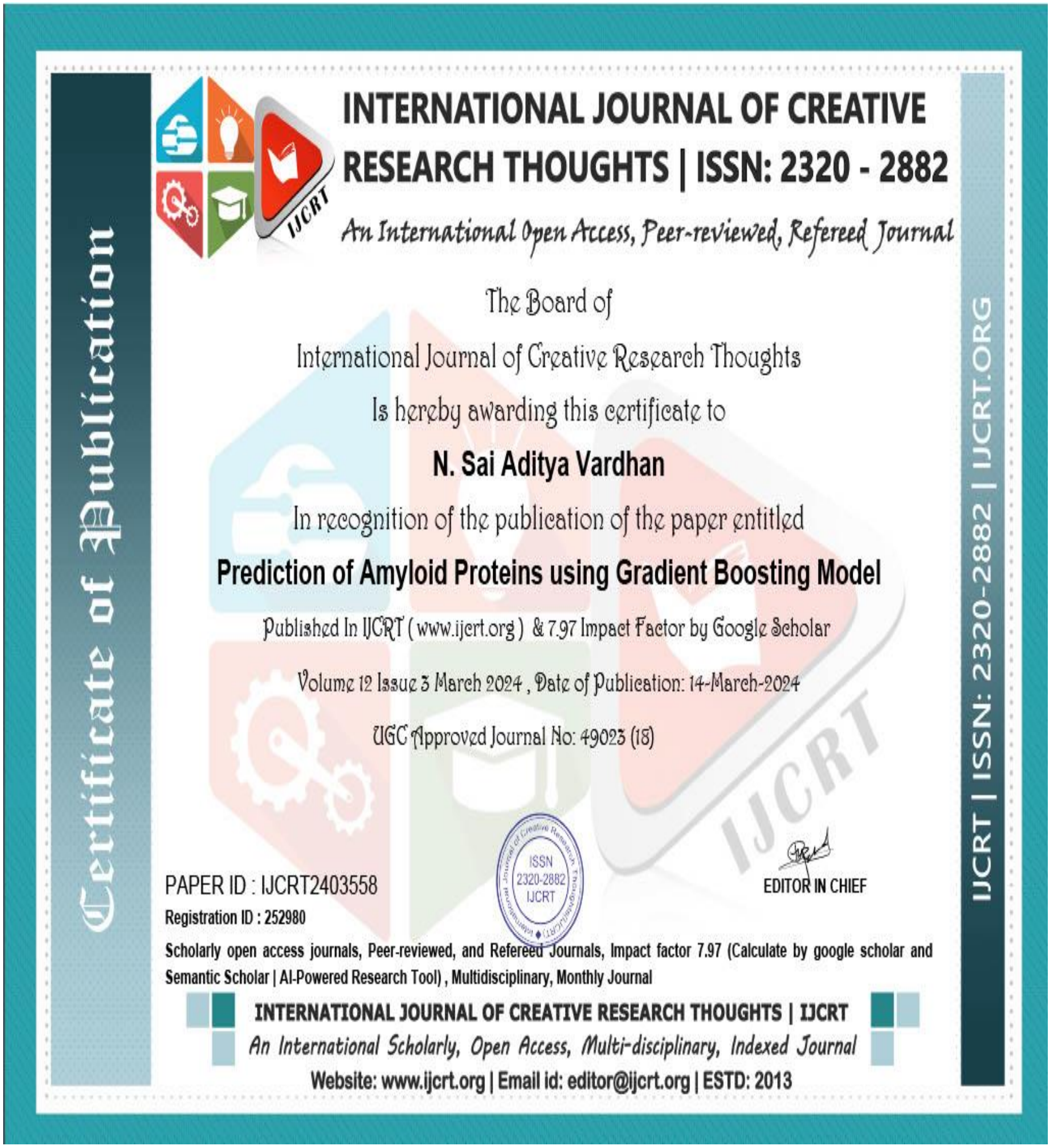
Scholarly open access journals, Peer-reviewed, and Refereed Journals, Impact factor 7.97 (Calculate by google scholar and Semantic Scholar | AI-Powered Research Tool) , Multidisciplinary, Monthly Journal

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS | IJCRT

An International Scholarly, Open Access, Multi-disciplinary, Indexed Journal

Website: www.ijcrt.org | Email id: editor@ijcrt.org | ESTD: 2013

IJCRT | ISSN: 2320-2882 | IJCRT.ORG



Certificate of Publication

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS | ISSN: 2320 - 2882

An International Open Access, Peer-reviewed, Refereed Journal

The Board of

International Journal of Creative Research Thoughts

Is hereby awarding this certificate to

N. Sai Aditya Vardhan

In recognition of the publication of the paper entitled

Prediction of Amyloid Proteins using Gradient Boosting Model

Published In IJCRT (www.ijcrt.org) & 7.97 Impact Factor by Google Scholar

Volume 12 Issue 3 March 2024 , Date of Publication: 14-March-2024

UGC Approved Journal No: 49023 (18)

PAPER ID : IJCRT2403558

Registration ID : 252980

International Journal of Creative Research Thoughts (IJCRT)
ISSN
2320-2882
IJCRT

EDITOR IN CHIEF

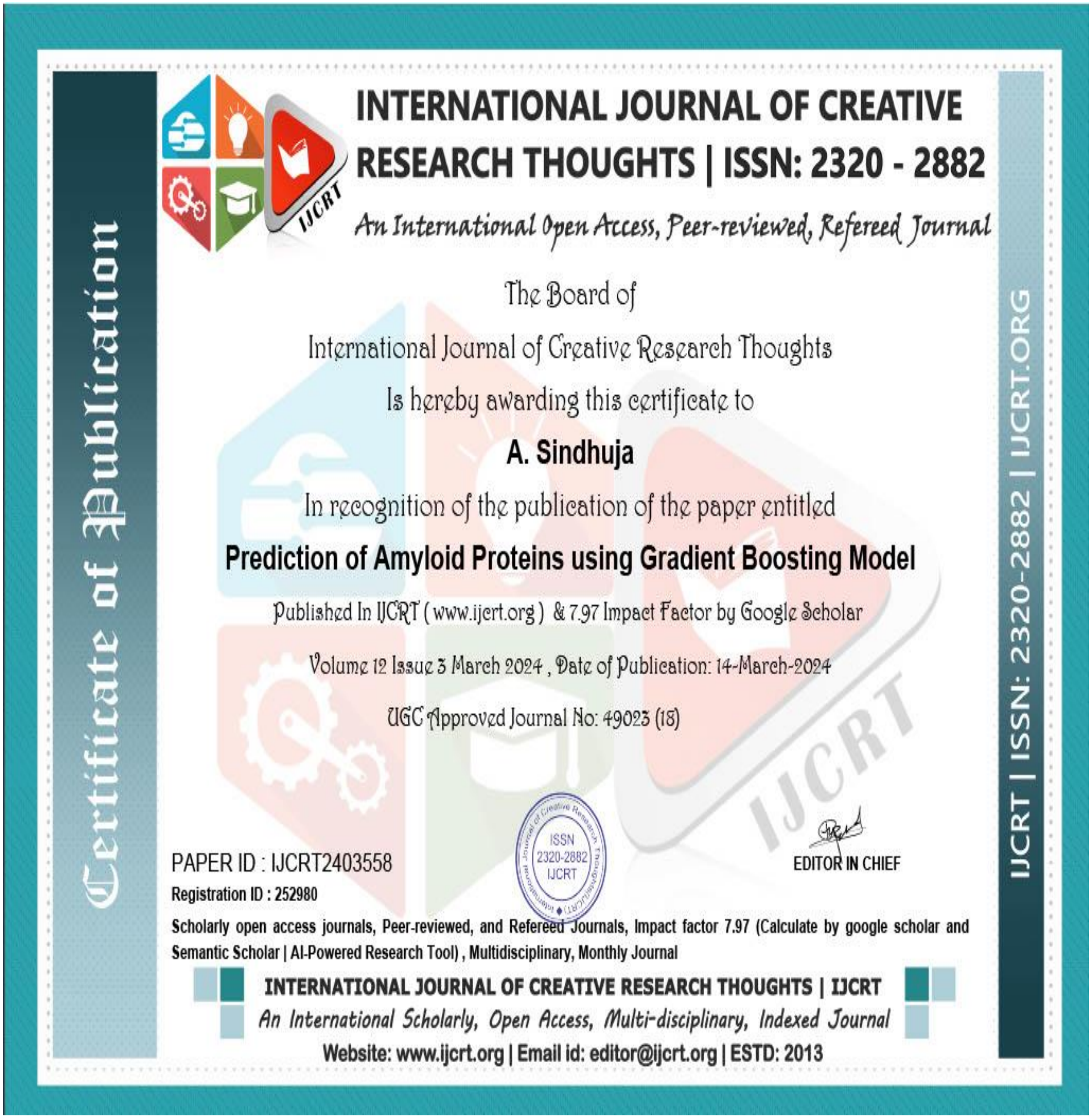
Scholarly open access journals, Peer-reviewed, and Refereed Journals, Impact factor 7.97 (Calculate by google scholar and Semantic Scholar | AI-Powered Research Tool) , Multidisciplinary, Monthly Journal

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS | IJCRT

An International Scholarly, Open Access, Multi-disciplinary, Indexed Journal

Website: www.ijcrt.org | Email id: editor@ijcrt.org | ESTD: 2013

IJCRT | ISSN: 2320-2882 | IJCRT.ORG



Certificate of Publication

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS | ISSN: 2320 - 2882

An International Open Access, Peer-reviewed, Refereed Journal

The Board of

International Journal of Creative Research Thoughts

Is hereby awarding this certificate to

A. Sindhuja

In recognition of the publication of the paper entitled

Prediction of Amyloid Proteins using Gradient Boosting Model

Published In IJCRT (www.ijcrt.org) & 7.97 Impact Factor by Google Scholar

Volume 12 Issue 3 March 2024 , Date of Publication: 14-March-2024

UGC Approved Journal No: 49023 (18)

PAPER ID : IJCRT2403558

Registration ID : 252980

ISSN 2320-2882 IJCRT

International Journal of Creative Research Thoughts (IJCRT)

EDITOR IN CHIEF

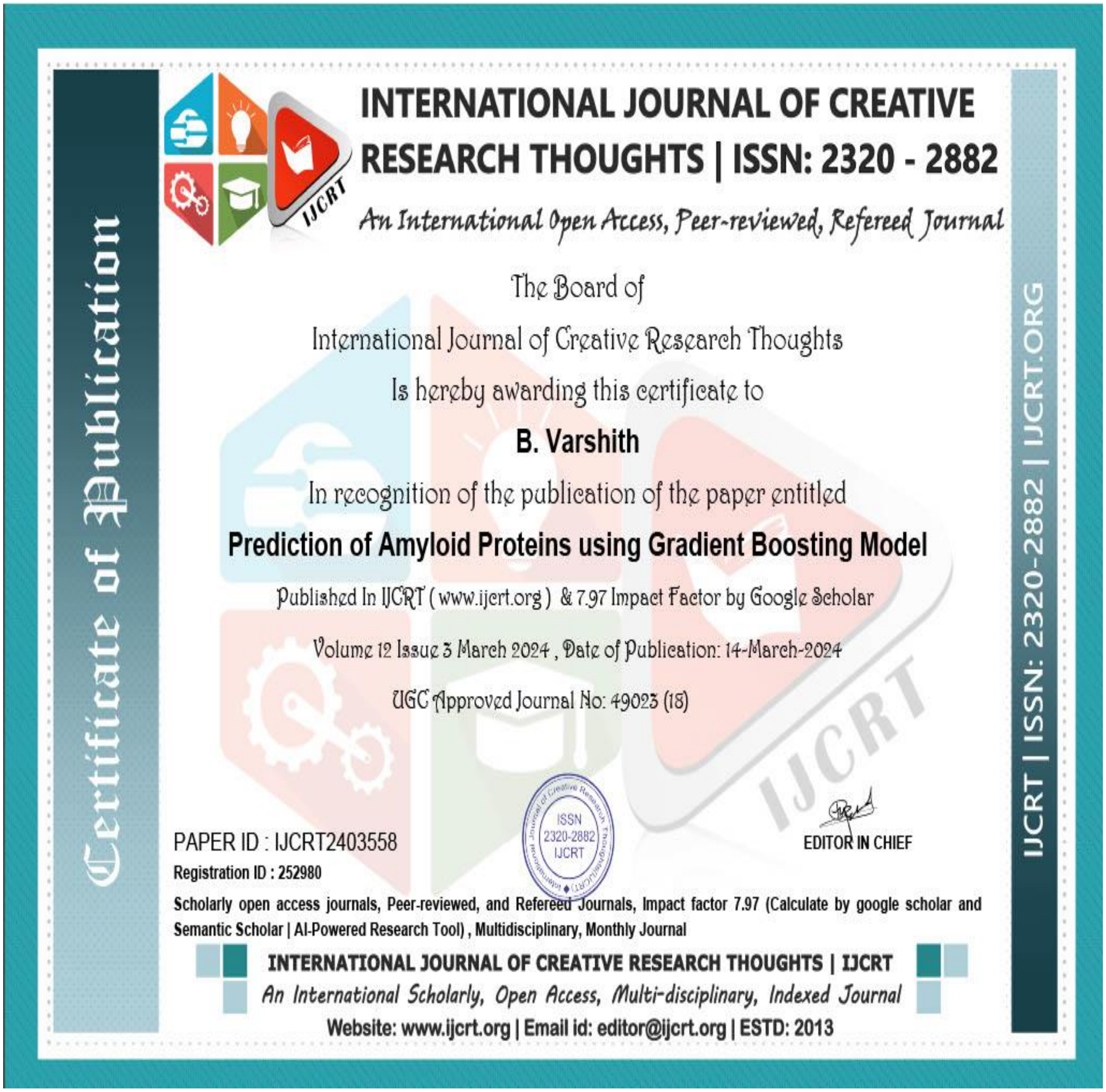
Scholarly open access journals, Peer-reviewed, and Refereed Journals, Impact factor 7.97 (Calculate by google scholar and Semantic Scholar | AI-Powered Research Tool) , Multidisciplinary, Monthly Journal

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS | IJCRT

An International Scholarly, Open Access, Multi-disciplinary, Indexed Journal

Website: www.ijcrt.org | Email id: editor@ijcrt.org | ESTD: 2013

IJCRT | ISSN: 2320-2882 | IJCRT.ORG



Certificate of Publication

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS | ISSN: 2320 - 2882

An International Open Access, Peer-reviewed, Refereed Journal

The Board of

International Journal of Creative Research Thoughts

Is hereby awarding this certificate to

B. Varshith

In recognition of the publication of the paper entitled

Prediction of Amyloid Proteins using Gradient Boosting Model

Published In IJCRT (www.ijcrt.org) & 7.97 Impact Factor by Google Scholar

Volume 12 Issue 3 March 2024 , Date of Publication: 14-March-2024

UGC Approved Journal No: 49023 (18)

PAPER ID : IJCRT2403558

Registration ID : 252980

ISSN 2320-2882 IJCRT

International Journal of Creative Research Thoughts(IJCRT)

EDITOR IN CHIEF

Scholarly open access journals, Peer-reviewed, and Refereed Journals, Impact factor 7.97 (Calculate by google scholar and Semantic Scholar | AI-Powered Research Tool) , Multidisciplinary, Monthly Journal

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS | IJCRT

An International Scholarly, Open Access, Multi-disciplinary, Indexed Journal

Website: www.ijcrt.org | Email id: editor@ijcrt.org | ESTD: 2013

IJCRT | ISSN: 2320-2882 | IJCRT.ORG