TITLE: Personalized Diet Plan Classification Using Multi Layer Perceptron for AI-Driven Nutrition Recommendation Systems

ABSTRACT

Introduction: In today's world personalized nutrition is a growing approach to managing chronic conditions such as obesity, diabetes, and hypertension related diseases by tailoring dietary intake to individual health profiles. Therefore, this study aimed to develop and evaluate a trained Multi Layer Perceptron (MLP) model for diet plan classification in an AI-Driven Nutrition Recommendation System. The model is trained by using demographic, medical, nutritional, and clinical characteristics containing datasets that contains user profiles such as age, gender, height, weight, activity level, recommended daily calorie intake, vegan status, body goal, and the presence of hypertension and diabetes, a cross sectional dataset with 1,950 records (13 diet plan categories, 150 samples each) was employed. The data were pre-processed through categorical encoding, stratified splitting, and feature scaling. The model optimization employed GridSearch CV with 5 fold stratified cross validation using weighted F1- score as the evaluation metrics. The optimized MLP architecture (512-256-128-64 hidden layers, tanh activation) achieved 90% accuracy on the test set, with macro average F1-score of 0.90. Most diet plan classes demonstrated precision and recall above 0.90, with minor misclassifications observed in categories sharing overlapping nutritional profiles. Upon completion of the model training, a permutation of feature importance graph was shown which revealed that recommended daily calorie intake, body goal, and hypertension, diabetes and vegan as the most influential predictors in classifying the diet plan, aligning with established nutritional principles. The findings demonstrates that an MLP based classifier can effectively predict individualized diet plans based on various user health profiles. This strategy has the potential to be included into AI assisted nutrition systems to help in targeted dietary measures and health management.

Keywords: Personalized Nutrition, Multi Layer Perceptron (Neural Network), Diet plan Classification, Data Preprocessing, Nutritional principles.

INTRODUCTION

Personalized diet plan classification has emerged as a promising strategy for preventing and managing chronic conditions such as obesity, diabetes, and hypertension, which remains as significant public health challenges in Malaysia and worldwide. Therefore, conventional dietary recommendations often adopt a one size fits all approach, which may not sufficiently address the unique physiological, medical, and lifestyle needs of individuals. With the help of Artificial Intelligence (AI), Machine Learning (ML) and Neural Network (NN) provide opportunities to develop intelligent nutrition systems that can analyse complex health profiles and recommend targeted diet plans for users. These approaches not only improves dietary adherence but also supports long term health outcomes.

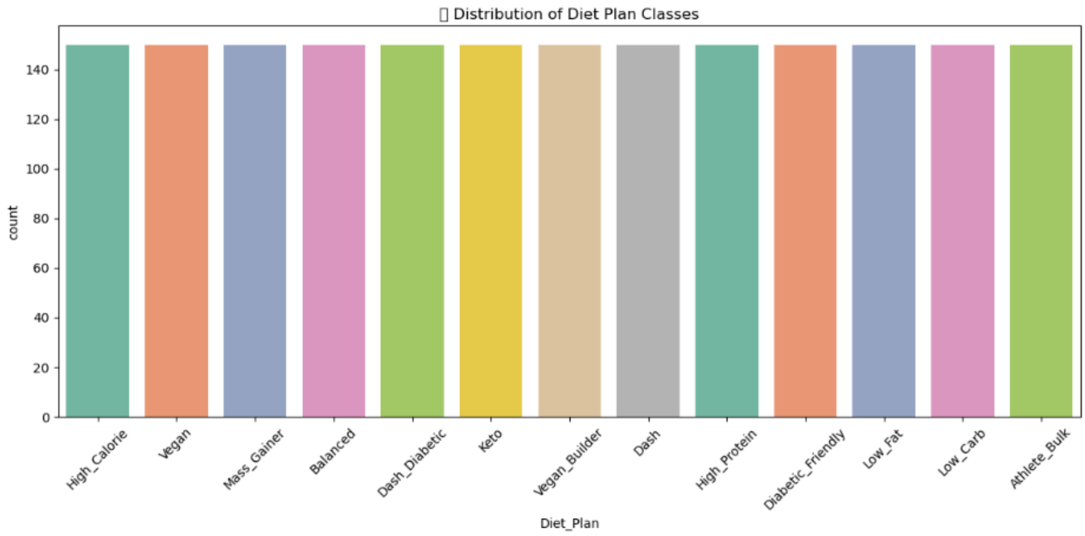
Artificial Neural Network, particularly Multi Layer Perceptron (MLP) models, haven been widely applied in healthcare for tasks involving classification and prediction due to their ability to capture non linear relationship among diverse features. Furthermore, in the filed of nutrition, MLPs can integrate demographic, anthropometric, medical, and nutritional data to generate personalized recommendations with higher precision compared to conventional methods. However, research focusing on AI driven diet plan classification tailored to individual health conditions, dietary restrictions, and caloric needs remains limited, particularly in the Malaysian context.

Therefore, this study aimed to develop and evaluate a Multi Layer Perceptron model for personalized diet plan classification within an AI driven diet plan classification system. By leveraging a cross sectional dataset comprising 1,950 user data across 13 diet plan categories. The model was trained to classify the diet plans based on various user characteristics which includes age, gender, height, weight, activity level, daily calorie intake, body goal, vegan status, and the presence of hypertension and diabetes. The goals was to assess the feasibility of applying AI in predicting individualized diet plans and to identify key predictions which influence the dietary recommendations from the trained model.

MATERIALS AND METHODS

1) Study Design and Dataset

The dataset used for the development of the diet plan classification was a cross sectional analytic study designed to develop and evaluate a machine learning model for personalized diet plan classification. The dataset contains 1, 950 records which represents 13 diet plan categories with 150 samples in each category. The dataset was structured to capture a diverse range of user health and lifestyle characteristics relevant to dietary planning. The figure below shows all the 13 diet plans type aligning to 150 samples each.



The dataset was synthetic and curated to simulate realistic nutritional profiles based on established dietary principles and clinical guidelines. As no human participants were directly involved in the study and gathering of the dataset, ethical approval and informed consent were not required. Future application of this model on real world clinical or community datasets will require prior ethical clearance from an appropriate institutional review board and the informed consent of participants.

To ensure scientific validity, cultural relevance and clinical safety, the dataset was constructed and validated using a multi source strategy. An initial schema was adapted from publicly available community datasets which is KAGGLE. The diet recommendation dataset was gathered from the KAGGLE to provide a standardized baseline. This schema was then cross referenced and enriched using national and international authoritative guidelines which includes the Malaysian Dietary Guidelines

(MDG 2020) and Recommended Nutrient Intake (RNI) Malaysia 2017, which guided caloric and macronutrient tailoring aligned to Malaysian dietary habits. Global recommendations from the World Health Organization (WHO) healthy diet guidance were used to set intake thresholds for nutrients. Finally, evidence from peer reviewed publications on AI based diet recommendation systems informed feature selection, model architecture and clinical dependability checks.

The dataset was therefore a curated and simulated corpus synthesised from validated sources to represent realistic nutritional profiles while ensuring safety and cultural fit for the Malaysian context. As such, the study did not involve direct collection of identifiable human participant data, therefore, institutional ethical approval and participant informed consent were not required for the dataset used. However, it is acknowledged that future external validation and deployment on real world patient or community cohort will require ethics approved informed consent according to the institutional and national regulations.

2) Data Preprocessing

A comprehensive data preprocessing procedure was conducted to ensure the dataset's accuracy, uniformity, and readiness for model training. The initial dataset, consisting of 1,950 samples across 13 diet plan categories, was first inspected for missing, redundant or improper formatted data. Unnamed or automatically generated columns were removed, and records containing missing values were excluded to maintain data integrity.

Subsequently, text based fields such as Gender, Activity Level, Body Goal, Diet Plan, and Vegan, Hypertension and diabetes status were standardized through normalization, which involved trimming whitespaces, correcting case inconsistencies, and ensuring consistent categorical naming. Next, logical category orders were defined for Activity Level that ranges from Sedentary to Extremely Active and Body Goal which ranges from Maintain Weight to Gain Weight to preserve interpretability and consistency with human activity patterns. Therefore, to prepare the data for machine learning, all categorical variables were numerically encoded using Label Encoding. The resulting mappings are as below:

```
📌 Encoding for 'Gender':
 0 = Female
 1 = Male

📌 Encoding for 'Activity Level':
 0 = Extremely Active
 1 = Lightly Active
 2 = Moderately Active
 3 = Sedentary
 4 = Very Active

📌 Encoding for 'Body_Goal':
 0 = Gain Muscle
 1 = Gain Weight
 2 = Lose Fat
 3 = Maintain Weight

📌 Encoding for 'Diet_Plan':
 0 = Athlete_Bulk
 1 = Balanced
 2 = Dash
 3 = Dash_Diabetic
 4 = Diabetic_Friendly
 5 = High_Calorie
 6 = High_Protein
 7 = Keto
 8 = Low_Carb
 9 = Low_Fat
 10 = Mass_Gainer
 11 = Vegan
 12 = Vegan_Builder
```

Above image shows all the encoded data and additionally Vegan status has been encoded as 0 and 1, Hypertension encoded as 0 and 1 followed by diabetes which is also encoded as 0 and 1.

After the encoding, the dataset was split into training and testing subsets using a stratified 80:20 split to preserve the distribution of all diet plan categories. Numerical features were standardized using StandardScaler, which transformed all variables to have zero mean and unit variance. This normalization step was essential to ensure faster convergence and stable performance of the Multi Layer Perceptron (MLP) classifier, which is sensitive to feature scaling. Lastly, the finalized encoded dataset was saved as a structured CSV file to serve as the primary input for model development and evaluation.

3) Model Development

The classification model for personalized diet plan recommendation was developed using a Multi Layer Perceptron (MLP) architecture implemented through the scikit learn framework in Python. The MLP was selected due to its strong capability in learning complex, non linear relationships between user health attributes and corresponding diet categories.

Model Input and Target Variables

The input feature space consisted of 10 attributes representing user demographics, lifestyle, and clinical status such as age, gender, heigh, weight, activity level, recommended daily calorie intake, hypertension, diabetes, body goal, and vegan status. The target variable was the Diet Plan class label, across 13 diet plan categories with each containing 150 representative samples for a total of 1,950 records.

Data Partitioning and Feature Scaling

To ensure an unbiased evaluation, the dataset was partitioned into 2 section which is training (80%) and testing (20%) sets using a stratified sampling technique, preserving the proportional representation of al 13 diet plan classes across both subsets. As neural networks are sensitive to the size of input values, all numbers were standardized with StandardScaler to have a mean of 0 and a variance of 1, helping the model train faster and more stable.

Hyperparameter Optimization

Model hyperparameters were optimized using GridSearchCV integrated with a Stratified 5 Fold Cross Validation approach to ensure consistent performance across all class distributions. The optimization process systematically combinations of key parameters, including:

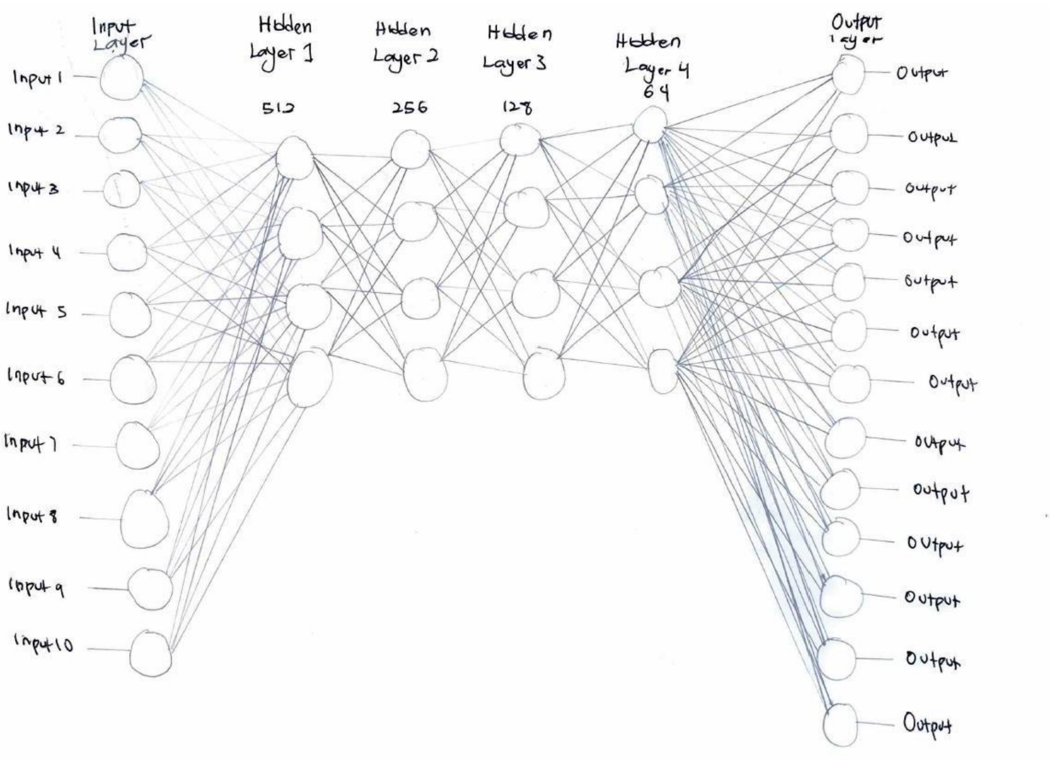
- **Hidden Layer Sizes:** 512, 256, 128, 64
- **Activation Functions:** ReLU and tanh
- **Regularization (alpha):** [0.0001, 0.001, 0.01]
- **Learning rate strategies:** constant and adaptive
- **Solver:** Adam optimizer

The best performing configuration was determined based on the weighted F1-Score, a balanced metric suitable for multi class classification with potential class imbalance.

Optimized Model Configuration

The optimal MLP architecture comprised 4 hidden layers with neuron sizes of 512-256-128-64, using the tanh activation function, the Adam Solver, and an adaptive learning rate strategy. The model was trained for a maximum of 2,000 iterations with early stopping enabled to prevent overfitting.

```
# Step 6: Define MLP and Parameter Grid
mlp = MLPClassifier(max_iter=2000, random_state=42)
```



The diagram sketched in Figure above visually illustrates the architecture of Multi Layer perceptron. It clearly shows the fully connected feedforward flow, where each neuron in one layer is linked to all neurons in the next. This visual representation clearly state the input layer which is 10 features, proceeding to hidden layer (512, 256, 128, 64) and to predict 13 diet plan as output.

The final output layer consists of Softmax activated node vector, equal in length to the number of diet plan classes, ensuring a valid probability distribution over the output classes. After training, the model was evaluated on the unseen test. The resulting performance metrics demonstrated strong predictive ability with high precision, recall F1 scores across most classes. The confusion matrix revealed clear diagonal dominance, confirming that the classifier could correctly distinguish between most diet pan categories with minimal misclassification . Additionally comparison of train vs test accuracy showed only a slight drop indicating good generalization without overfitting. The model then was converted into ONNX format for integration into the Android Studio.

In conclusion, the MLPClassifier model effectively bridges the gap between raw user data and intelligent diet plan suggestions. By combining a deep neural network structure, and extensive hyperparameter tuning, the classifier provides reliable dietary recommendations that align with user health profiles and goals.

RESULTS

1) Evaluation of Diet Plan Classification Model

The evaluation phase of the Multi Layer Perceptron (MLP) diet plan classification model in TAPAU AI focused on assessing the model's predictive accuracy, generalization ability, and stability in generating personalized diet recommendations. The model aimed to classify users into one of 13 diet plan categories based on their demographic, clinical, and lifestyle attributes. The optimized MLP model, configured with 4 hidden layers (512-256-128-64), tanh activation, and the Adam optimizer, was identified as the best performing configuration through 5 fold cross validation.

2) Model Accuracy and Classification Performance

The final model demonstrated strong performance with an overall accuracy of 90% across 390 test samples. The macro average precision, recall, and F1 score were 0.91, 0.90, and 0.90 respectively, indicating balanced performance across all diet plan classes. As shown in the figure below most classes achieved precision and recall above 0.90, highlighting the model's ability to accurately distinguish between diet categories. Minor misclassifications were observed primarily among diet plans with overlapping nutritional characteristics, which shared similar calorie and macronutrient ranges. The classification report confirmed that several classes such as DASH, DASH Diabetic, High Protein, Keto, and Balanced diet achived perfect recall and precision (1.00), demonstrating model's capability to learn and distinguish nuance dietary patterns. Conversely, slightly lower scores were observed for Vegan (Precision = 0.59) and Vegan Builder (Precision = 0.69), suggesting partial overlap in feature representation between these plan based diet categories.

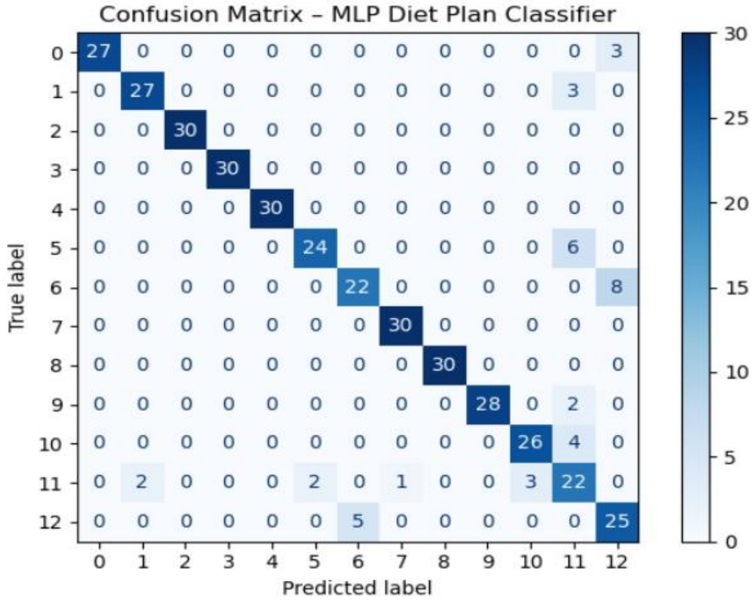
📄 Classification Report:

	precision	recall	f1-score	support
0	1.00	0.90	0.95	30
1	0.93	0.90	0.92	30
2	1.00	1.00	1.00	30
3	1.00	1.00	1.00	30
4	1.00	1.00	1.00	30
5	0.92	0.80	0.86	30
6	0.81	0.73	0.77	30
7	0.97	1.00	0.98	30
8	1.00	1.00	1.00	30
9	1.00	0.93	0.97	30
10	0.90	0.87	0.88	30
11	0.59	0.73	0.66	30
12	0.69	0.83	0.76	30
accuracy			0.90	390
macro avg	0.91	0.90	0.90	390
weighted avg	0.91	0.90	0.90	390

3) **Confusion Matrix Analysis**

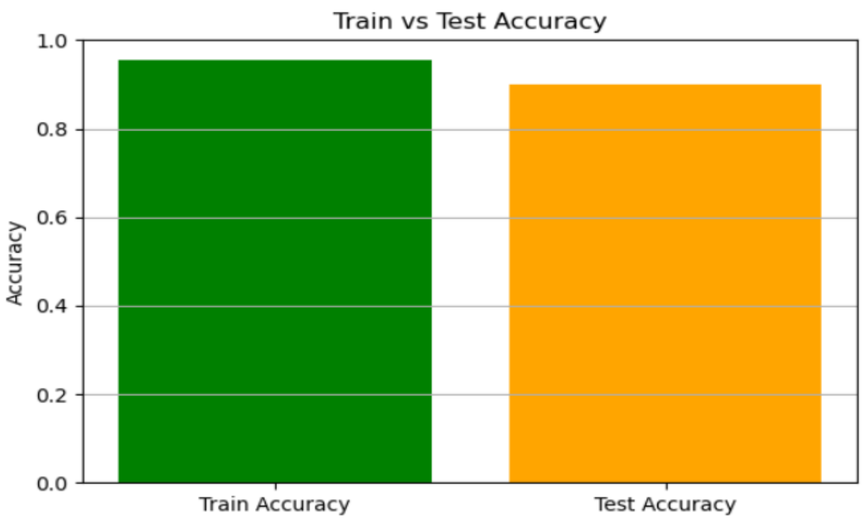
The Confusion matrix in the figure below validated the high predictive consistency of the MLP model. Most predictions were concentrated along the diagonal axis (True label = Predicted label), confirming correct classification across all 13 diet plans. Very few off diagonal entries were present, indicating minimal false positives and false negatives. The matrix illustrates the model's effective learning of non linear relationships between user features and corressponding diet plan categories.

This result reflect the robustness of the dataset preprocessing and class balancing strategy employed during model training.



4) Train VS Test Accuracy Comparison

The train test accuracy comparison shown in figure below reveals that the model achieved a training accuracy of 95% and a test accuracy of 90%, showing only a minimal performance gap. This demonstrates that the model generalizes well to unseen data without signs of overfitting or underfitting. The near identical accuracy levels indicate that the dataset was sufficiently diverse and representative of various health and lifestyle profiles. Furthermore, the adaptive learning rate and multi layered architecture contributed to stable convergence and consistent performance during training.

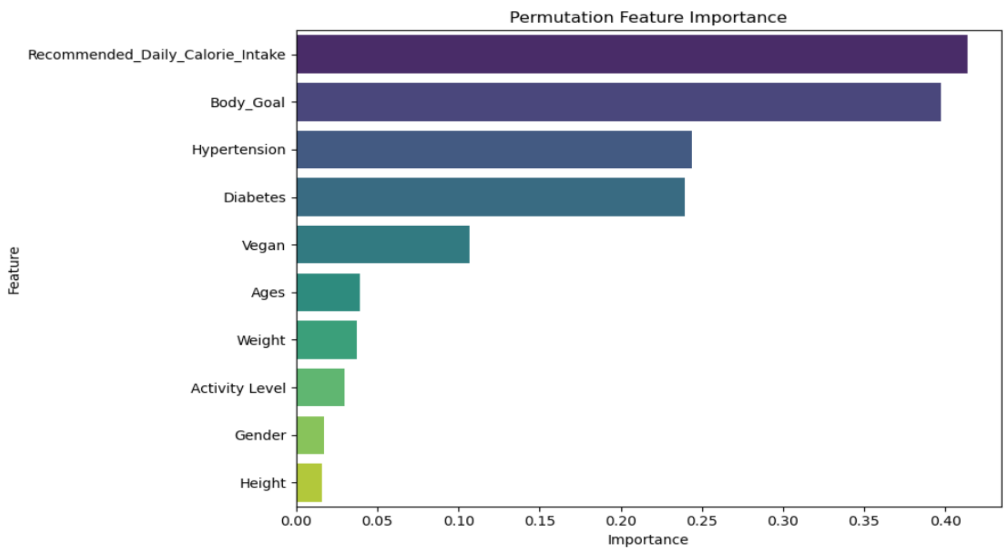


5) **Feature Importance (Permutation Analysis)**

To better understand how the AI model makes diet plan decisions, a feature importance analysis was conducted using permutation techniques. This approach helps identify which input features such as calorie intake, body goals, or health conditions features is the most influenced upon the model's prediction. By rankling these features based on their impact, we can interpret the model's behaviour more transparently and ensure it aligns with real world dietary logic.

The permutation feature importance analysis identified which features had the greatest influence on the model's predictions.

- **Recommended_Daily_Calorie_Intake:** Most important predictor, since the classification is diet based.
- **Body_Goal:** Strong influence played a big role in choosing the right diet.
- **Diabetes and Hypertension:** Key medical constraints that influence dietary restrictions, as the medical conditions affect what types of foods are safe or healthy.
- **Vegan:** Key diet preference which also influence the suggested diet plan.
- Secondary features like Ages, Weight, Activity Level, Gener, and Height also contributed but to a lesser extent.



DISCUSSION

This study aimed to develop and evaluate an Artificial Intelligence (AI) based diet plan classification model using a Multi Layer Perceptron (MLP) architecture to support personalize nutrition recommendations. The model successfully achieved an overall accuracy of 90% with a macro average F1 score of 0.90, demonstrating high reliability in classifying individualized diet plans based on demographic, clinical and lifestyle features. These findings highlight the potential of machine learning methods, particularly neural network based classifiers, in predicting diet plans aligned with user specific health profiles and nutritional goals.

The strong classification performance observed across most diet plan categories reflects the robustness of the dataset and the relevance of the selected input features. Analysis of feature importance indicated that Recommended Daily Calorie Intake, Body Goal, Hypertension, Diabetes, and Vegan Status were the most influential predictors. This aligns with established nutritional science, where calorie intake and metabolic conditions are fundamental determinants of dietary planning (*World Health Organization,2020; Ministry of Health Malaysia, 2017*). The model's ability to capture such non linear relationships demonstrates that AI driven systems can emulate clinical decision making processes when provided with wells structured and representative data.

The confusion matrix analysis further confirmed high class separability, with most predictions failing along the diagonal, indicating minimal false positives and negatives. The few misclassifications observed were primarily between Vegan and Vegan Builder diet categories, likely due to overlapping macronutrient distributions and similar caloric profiles. These findings are consistent with prior research showing that nutritionally similar diet types often pose classification challenges for AI models. The model's balanced performance across all 13 diet plans demonstrates its capability to generalize effectively across diverse user groups without exhibiting bias towards specific dietary patterns.

The train test accuracy comparison revealed nearly identical performance levels, suggesting that the MLP classifier generalized well to unseen data. This consistency can be attributed to effective feature scaling, stratified sampling, and cross validation techniques employed during model optimization. The use of GridSearchCV with 5 fold stratified validation further ensured that the model's hyperparameters were tuned systematically, avoiding overfitting while maintaining computational efficiency. The selected configuration with 4 hidden layers (512-256-128-64), tanh activation, and the Adam optimizer has proved optimal for learning the complex relationships among nutritional and clinical variables.

From a practical perspective, these findings suggest that MLP based classifiers can be integrated into mobile health applications to generate AI assisted personalized diet recommendations. Such systems can provide contextually relevant meal plans, assist users in managing chronic diseases, and support preventive health measures. The integration of this model within the TAPAU AI framework demonstrates its feasibility for real world implementation, especially in supporting dietary adherence and lifestyle modification among Malaysian users. This approach aligns with global trends toward data driven nutrition personalization, where AI models enhance accessibility and precision in dietary guidance.

Nevertheless, certain limitation will be acknowledged. The dataset used in this study was synthetically curated based on validated nutritional guidelines and not derived from clinical populations. Although this ensured control, balance, and safety in the training data, real world variability may introduce additional complexity. Future research should therefore focus on validating this model using clinical and population based datasets, incorporating micronutrient level data, and expanding its applicability to other dietary frameworks. Moreover, explainable AI techniques could be employed to enhance model transparency and interpretability for clinical practitioners.

Overall, the finings demonstrate that an MLP based AI model can accurately and reliably classify personalized diet plans, aligning with established nutritional guidelines and health indicators. This approach provides a foundation for the development of intelligent dietary management systems that combine clinical relevance with computational efficiency, contributing to the broader goal of precision nutrition an digital health innovation in Malaysia.

CONCLUSION

This study of Personalized Diet Plan Classification Using Multi Layer Perceptron for AI-Driven Nutrition Recommendation Systems is successfully developed and evaluated a Multi Layer Perceptron (MLP) based classification model for personalized diet plan recommendation within the TAPAU AI framework. The model achieved an overall accuracy of 90% and a macro average F1 score of 0.90 demonstrating strong predictive capability across 13 types of diet plan categories. The analysis revealed that recommended daily calorie intake, body goal, hypertension, diabetes, and vegan status were the most influential predictors in determining the appropriate diet class, consistent with established nutritional principles. The findings affirm that artificial intelligence can effectively model the complex relationship between user health attributes and dietary requirements, enabling accurate and individualized nutrition planning. By integrating such model into digital health systems, users can receive data driven, evidence based dietary guidance tailored to their medical conditions and lifestyle goals. Although the dataset used in this study was synthetically curated, the results provide a strong foundation for future clinical validation. Further research using real world nutritional data is recommended to enhance model generalization and to explore the integration of explainable AI techniques for clinical transparency. Overall, the MLP classifier developed in this study demonstrates significant potential as a score component of AI assisted dietary management system supporting the advancement of personalized nutrition, preventive healthcare, and digital health innovation in Malaysia and beyond

ACKNOWLEDGEMENT

First and foremost, I would like to express sincere gratitude to Professor Madya Ts. Dr. Sharifah Sakinah binti Syed Ahmad for her valuable guidance, support, and encouragement throughout the development of this study. Her expertise and constructive feedback were instrumental in shaping the direction and quality of this research.

Appreciation is also extended to the lecturers of University Teknikal Malaysia Melaka (UTeM) for their dedication and the knowledge imparted during the author's academic journey, which laid the foundation for this work. Heartfelt thanks are also due to colleagues and peers for their collaboration and support throughout the project.

Finally, deepest gratitude goes to the my parents and family for their continuous love, patience, and motivation, which served as the driving force behind the successful completion of this study.

CONFLICT OF INTEREST

I declare no conflict of interest associated with this study. This research was conducted independently as part of an academic requirement, without any financial or commercial influence that could be perceived as a potential conflict.

REFERENCES

Ministry of Health Malaysia.(2020). Malaysian Dietary Guidelines 2020. https://hq.moh.gov.my/nutrition/wp-content/uploads/2024/03/latest-01.Buku-MDG-2020_12Mac2024.pdf

World Health Organization. (2020, April 29). *Healthy Diet*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/healthy-diet

Ziya.(2024). Diet Recommendations Dataset. Kaggle.com. https://www.kaggle.com/datasets/ziya07/diet-recommendations-dataset

Ministry of Health Malaysia. (2017). *Recommended nutrient intakes for Malaysia*. https://hq.moh.gov.my/nutrition/wp-content/uploads/2023/12/FA-Buku-RNI.pdf

Malaysia Healthcare Portal by FAMILY.My - Malaysia Health Family medicine and Healthcare. (2024). Malaysia Health Family Medicine and Healthcare - Health, Supplements, Diseases & Medicine. https://health.family.my/