Diet Type Recommendation Using Machine Learning

By: Awwab Samad

Contents

Abstract

- **Context:** Importance of tailored dietary recommendations for individuals with specific health conditions.
- **Project Objective:** Utilizing Neural Networks (MLP) with back propagation, for personalized diet type recommendations.
- **Method:** Training an MLP model on a dataset containing various medical parameters, such as BMI, blood glucose levels etc. to predict a sequence of binary values representing the needed diet types for patients with respective parameters.

Introduction

- High prevalence of chronic diseases, diabetes, obesity etc. high mortality rates.
- Need for lifestyle change, diet modification but challenges in personalizing diet plans.
- Traditional dietary plans often fail to adapt to individual health needs.
- Machine Learning Potential, Neural networks can analyze large datasets, learning complex patterns for personalized recommendations.
- Encourage individuals to make better health choices, improve quality of life.

Proposed Scheme

- Project has 5 main phases for developing a diet recommendation system

Data Collection
- Online websites

Data Preprocessing
- Extracting input features
- Separating into training, validation, testing

Model Development
- Setting up and training the neural network

Model Testing
- Validating and testing the trained network

Performance Evaluation
- Analyzing accuracy of model

Dataset Description

- Obtained from Kaggle, [“Metabolic Syndrome - A Comprehensive Dataset on Risk Factors and Health Indicators”]
- Total of 2303 entries
- **9 Input features:** medical information of anonymous patients which that will be used to train our neural network.
- **Target output:** a sequence of 11 binary values representing diet type recommendations for the patient that a registered dietitian has provided for us.
- **Diseases targeted:** Chronic Kidney Disease (CKD), Diabetes, Gout, Hypertriglyceridemia, Obesity.

Input Features

- Sex
- Waist Circumference
- BMI (Body Mass Index)
- UrAlbCr (Urinary Albumin-to-Creatinine Ratio)
- Serum Uric Acid
- Fasting BGL (Fasting Blood Glucose Level)
- HDL (High Density Lipoprotein)
- Triglycerides
- Bias term (added as an additional input to improve network performance)

Target Outputs

- **LoSC:** Low Simple Carbohydrates
- **LoPrt:** Low Protein
- **LoSF:** Low Saturated Fat
- **LoRM:** Low Red Meat
- **LoLgm:** Low Legumes
- **LoNa:** Low Sodium
- **LoP:** Low Phosphorus
- **HiUF:** High Unsaturated Fats
- **HiC:** High Vitamin C
- **HiD:** High Vitamin D
- **HiFib:** High Fiber

Diet Types Recommendations *(1 = Diet Type Recommended, 0 = Diet Type Not Needed)*

Disease	LoSC	LoPrt	LoSF	LoRM	LoLgm	LoNa	LoP	HiUF	HiC	HiD	HiFib
CKD	0	1	0	0	1	1	1	0	0	1	0
Diabetes	1	0	0	0	0	0	0	1	1	0	1
Gout	1	0	0	1	1	0	0	1	1	1	0
Hyper-triglycer-idemia 1	1	0	1	0	0	0	0	0	0	0	1
Hyper-triglycer-idemia 2	1	0	1	0	0	0	0	1	0	0	1
Obesity 1	1	0	1	0	0	0	0	0	0	0	1
Obesity 2	1	0	1	1	0	0	0	0	0	0	1

Software & Hardware

Software:

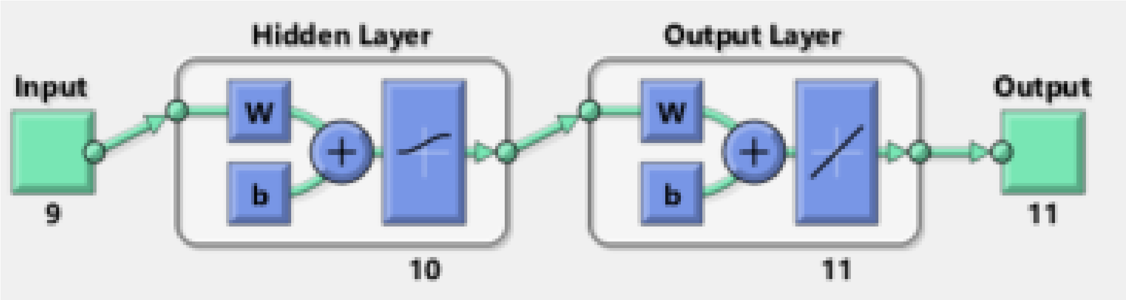
- **Programming Language:** MATLAB R2020a
- **Machine Learning Framework:** Neural Network Toolbox
- **Development Environment:** MATLAB
- **Operating System:** Windows 10
- **Spreadsheet Software:** Microsoft Excel

Hardware:

My PC specifications:

- **Processor:** i5-5200U CPU @2.20GHz
- **RAM:** 16GB RAM DDR3
- **Storage:** 1TB SSD Drive

Neural Network Architecture



Input Layer:

- 9 Input neurons: 8 features + 1 bias neuron.

Hidden Layer:

- 1 hidden layer
- 10 hidden layer neurons: Extracts complex patterns.
- **Activation Function:** Sigmoid function for non-linearity, modelling complex relationships.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

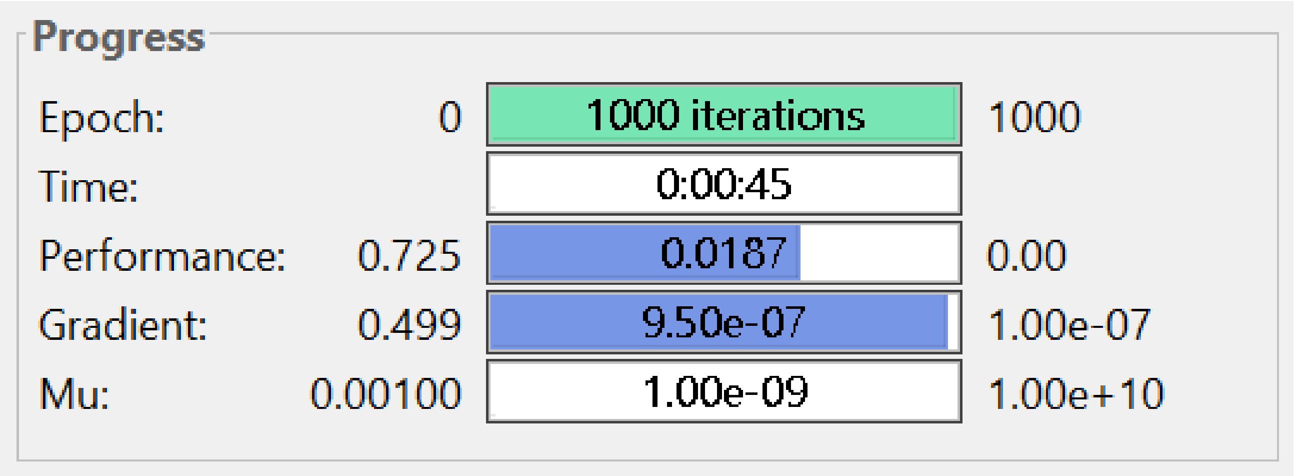
Output Layer:

- 11 Output neurons: Dietary recommendations (e.g., Low Protein, High Fiber represented as a binary 1 when a particular diet type is recommended, and a binary 0 otherwise).

Model Training

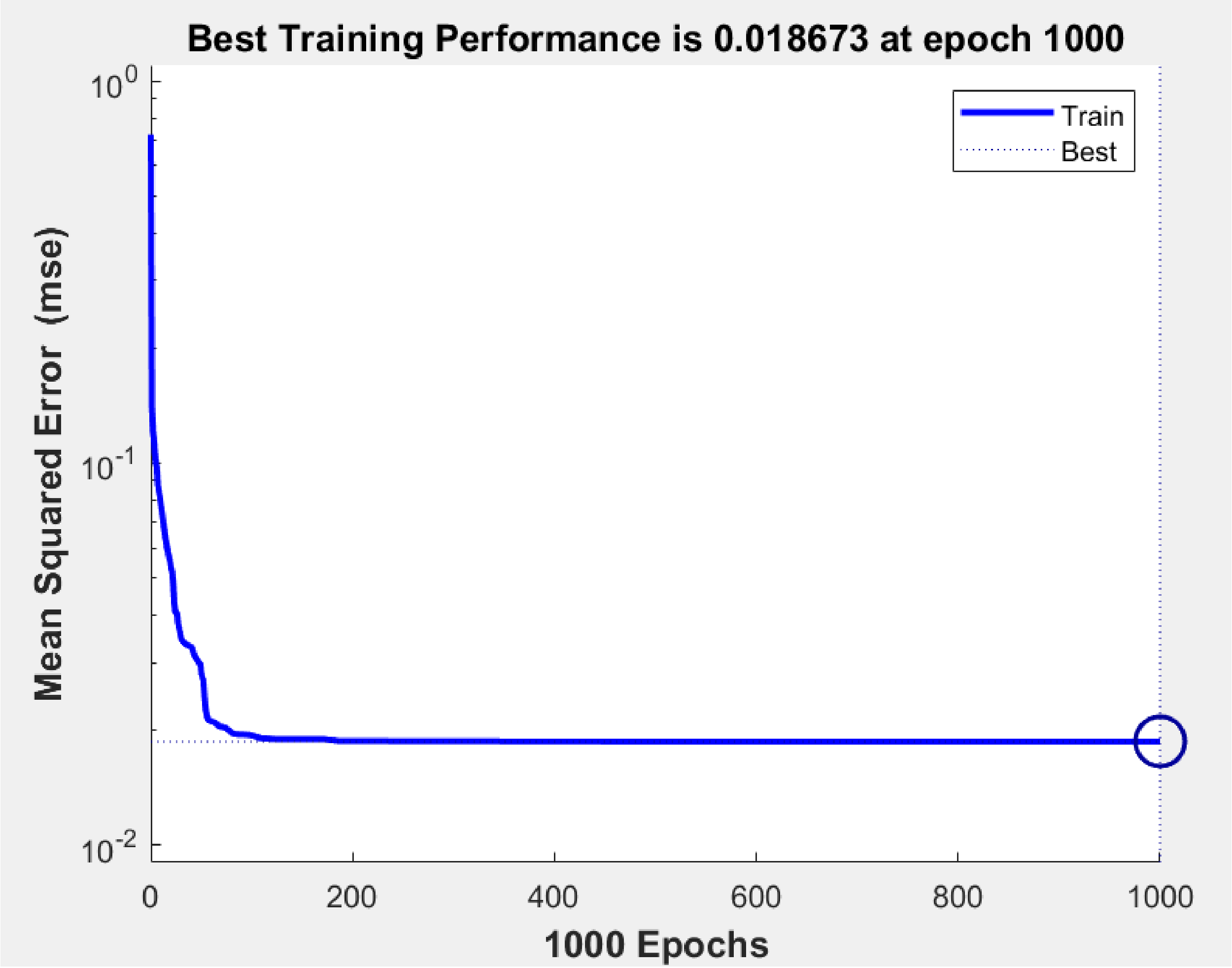
- **Dataset Split:**
  - **Training:** 70% (1612 entries)
  - **Validation:** 15% (345 entries)
  - **Testing:** 15% (346 entries)

- **Training Process:**
  - **Epochs:** 1000
  - **Learning Rate (Mu):** 0.001
  - **Validation Checks:** 100

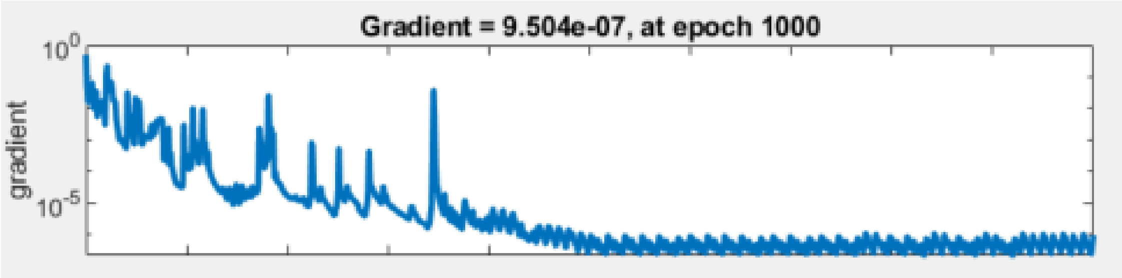


- **Goals:**
  - Improve model's ability to learn complex patterns.
  - Ensure stable convergence using gradient descent method to reach optimal performance.

Mean Squared Error



The lowest mean squared error achieved was 0.018673 after 1000 epochs.



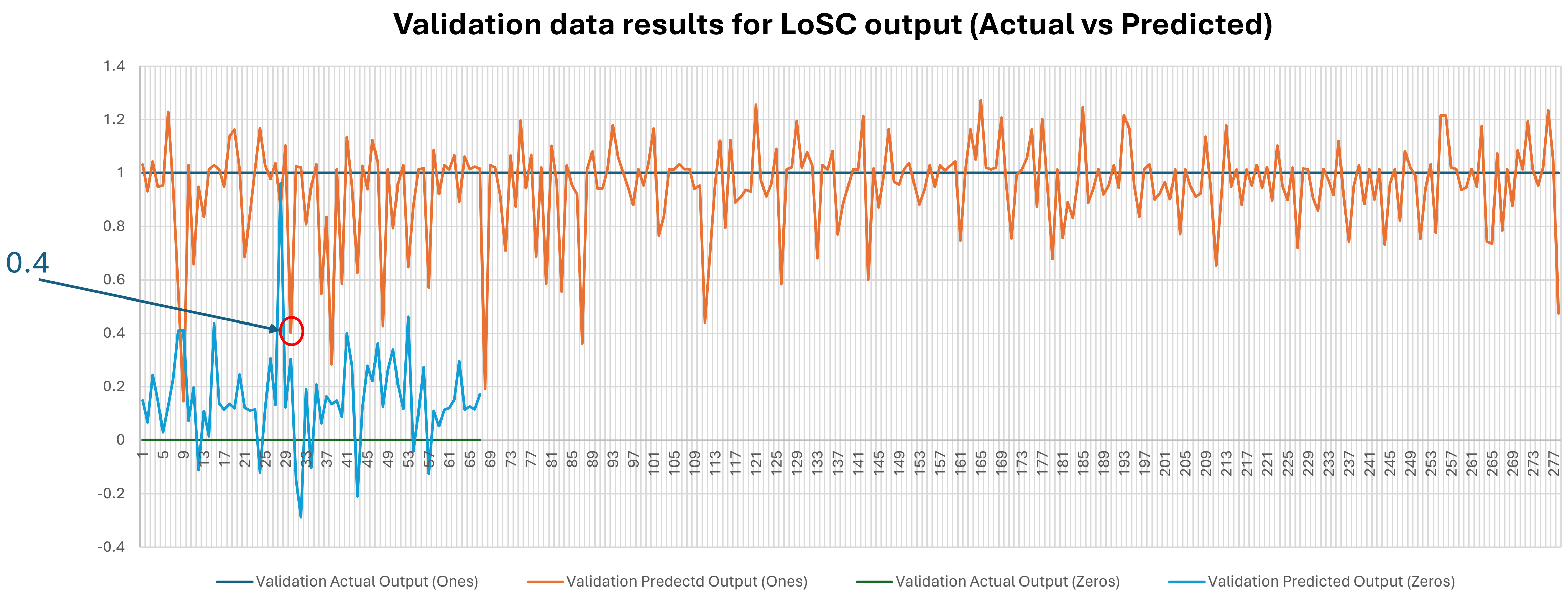
Using gradient descent to adjust weights and minimize error.

Actual vs Predicted Output

- Predicted (estimated) outputs show up as non binary floating point numbers.
- These numbers have a small spread around the values 1 and 0 (bar any anomaly) eg 1.023, 0.998, -0.12, 0.2 etc.
- An appropriate threshold is needed to convert these floating point numbers to binary 1s and 0s for comparing with the actual output.
- A threshold is determined for each output node (diet type) and applied to all predicted values of the output nodes to minimize the performance error.

Actual vs Predicted Output (LoSC)

(Optimum threshold to minimize prediction error = 0.4)



Validation Data Results

Validation Data Overall Results				
Diet Type (DT)	Overall Estimate of DT Recommendation			
	Total	Correct	Incorrect	Success
LoSC	345	333	12	96.52%
LoPrt	345	342	3	99.13%
LoSF	345	321	24	93.04%
LoRM	345	330	15	95.65%
LoLgm	345	340	5	98.55%
LoNa	345	342	3	99.13%
LoP	345	342	3	99.13%
HiUF	345	332	13	96.23%
HiC	345	342	3	99.13%
HiD	345	339	6	98.26%
HiFib	345	329	16	95.36%
Total	**3795**	**3692**	**103**	**97.29%**

- All diet types have been predicted successfully 93% of the time at the minimum.
- Success rate varies from 93% to 99% within various diet type recommendations.
- Overall prediction of diet types is correct 97.29% of the time.

Validation Results Breakdown

Validation Data Results Details				
Diet Type (DT)	When DT Present in Recommendation			
	Total	Correct	Incorrect	Success
LoSC	278	277	1	99.64%
LoPrt	41	41	0	100.00%
LoSF	233	229	4	98.28%
LoRM	207	207	0	100.00%
LoLgm	83	80	3	96.39%
LoNa	41	41	0	100.00%
LoP	41	41	0	100.00%
HiUF	201	196	5	97.51%
HiC	193	192	1	99.48%
HiD	83	80	3	96.39%
HiFib	273	270	3	98.90%
Total	**1674**	**1654**	**20**	**98.81%**

Validation Data Results Details				
Diet Type (DT)	When DT Not Present in Recommendation			
	Total	Correct	Incorrect	Success
LoSC	67	56	11	83.58%
LoPrt	304	301	3	99.01%
LoSF	112	92	20	82.14%
LoRM	138	123	15	89.13%
LoLgm	262	260	2	99.24%
LoNa	304	301	3	99.01%
LoP	304	301	3	99.01%
HiUF	144	136	8	94.44%
HiC	152	150	2	98.68%
HiD	262	259	3	98.85%
HiFib	72	59	13	81.94%
Total	**2121**	**2038**	**83**	**96.09%**

Validation Results Analysis

- The predicted results are accurate 98.81% of the time when a certain diet type is recommended for the patient and 96.09% when not recommended.
- The incorrect predictions happens mostly when a diet type is not recommended to the patient (False Positive).
- LoSC, LoSF, LoRM and HiFib have contributed to most of these false positives
- A possible reason for this could be fewer than needed training data available for better prediction results.

Testing Data Results

Test Data Overall Results				
Diet Type (DT)	Overall Estimate of DT Recommendation			
	Total	Correct	Incorrect	Success
LoSC	346	323	23	93.35%
LoPrt	346	345	1	99.71%
LoSF	346	315	31	91.04%
LoRM	346	323	23	93.35%
LoLgm	346	343	3	99.13%
LoNa	346	345	1	99.71%
LoP	346	345	1	99.71%
HiUF	346	330	16	95.38%
HiC	346	341	5	98.55%
HiD	346	343	3	99.13%
HiFib	346	323	23	93.35%
Total	**3806**	**3676**	**130**	**96.58%**

- The testing data results are very similar to the validation data results.
- All diet types have been predicted successfully 91% of the time at the minimum.
- Success rate varies from 91.04% to 99.71% within various diet type recommendations.
- Overall prediction of diet types is correct 96.58% of the time.

Testing Results Breakdown

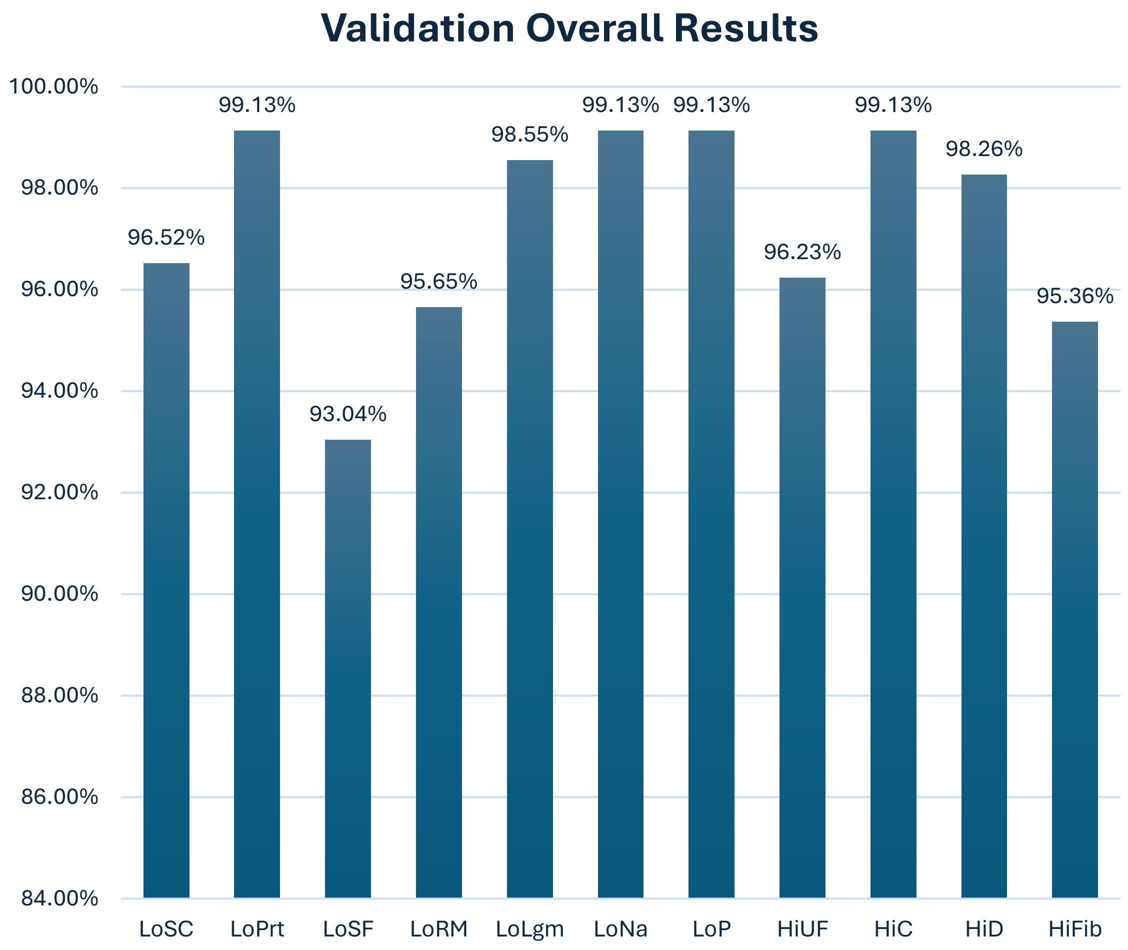
Test Data Results Details				
Diet Type (DT)	When DT Present in Recommendation			
	Total	Correct	Incorrect	Success
LoSC	277	274	3	98.92%
LoPrt	43	43	0	100.00%
LoSF	236	227	9	96.19%
LoRM	214	210	4	98.13%
LoIgm	92	91	1	98.91%
LoNa	43	43	0	100.00%
LoP	43	43	0	100.00%
HiUF	209	202	7	96.65%
HiC	197	195	2	98.98%
HiD	92	91	1	98.91%
HiFib	273	268	5	98.17%
Total	**1719**	**1687**	**32**	**98.14%**

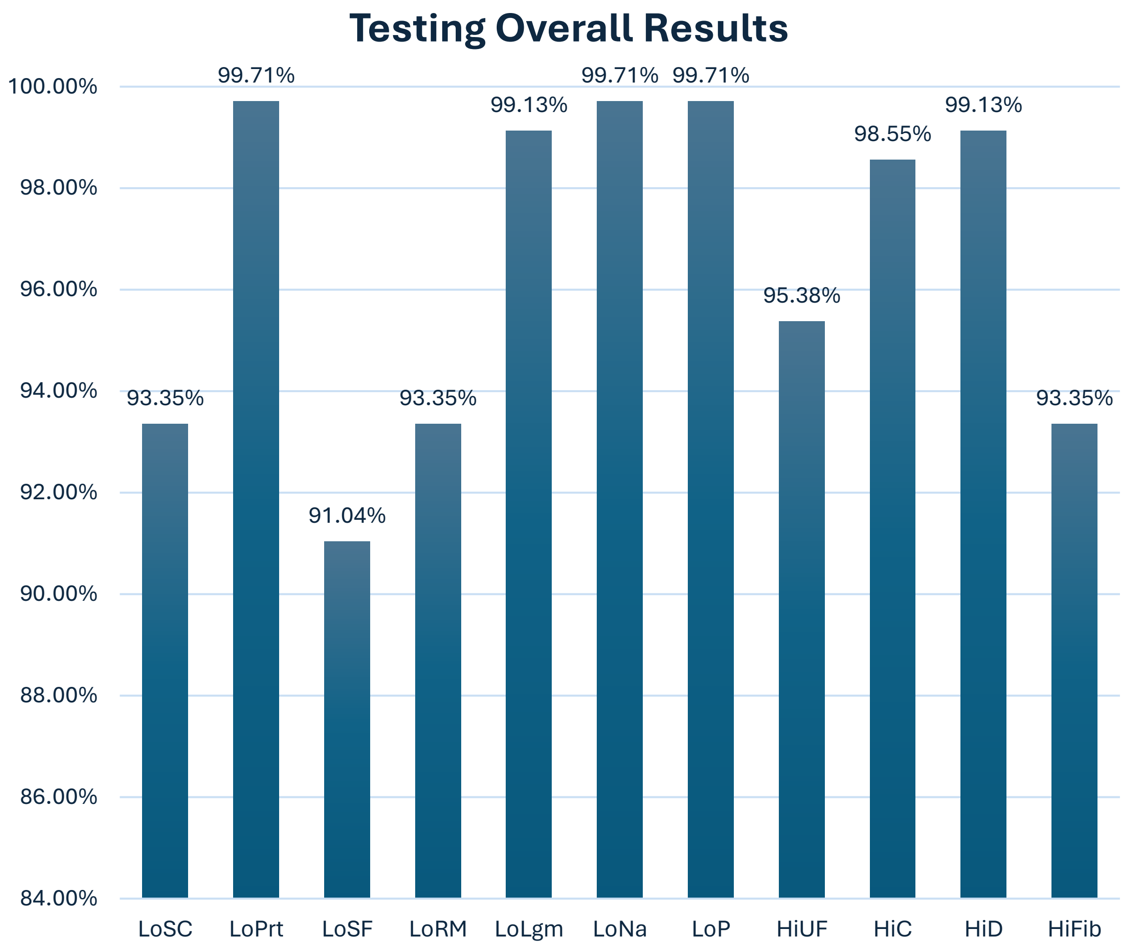
Test Data Results Details				
Diet Type (DT)	When DT Not Present in Recommendation			
	Total	Correct	Incorrect	Success
LoSC	69	49	20	71.01%
LoPrt	303	302	1	99.67%
LoSF	110	88	22	80.00%
LoRM	132	113	19	85.61%
LoIgm	254	252	2	99.21%
LoNa	303	302	1	99.67%
LoP	303	302	1	99.67%
HiUF	137	128	9	93.43%
HiC	149	146	3	97.99%
HiD	254	252	2	99.21%
HiFib	73	55	18	75.34%
Total	**2087**	**1989**	**98**	**95.30%**

Testing Results Analysis

- The predicted results are accurate 98.14% of the time when a certain diet type is recommended for the patient and 95.30% when the diet type is not recommended.
- The incorrect predictions happens mostly when a diet type is not recommended to the patient.
- As before, LoSC, LoSF, LoRM and HiFib have contributed to most of these false positives.
- This prediction is consistent with the analogous validation data.

Validation and Testing Data Outcome Comparison





Conclusion

- Using our model, the overall results when predicting the recommended diet type for a patient with the given input features were predicted quite successfully.
- The overall success rate was well over 90% in both the validation and testing data, model behaved consistently.
- Future research work can proceed in the direction of adding additional hidden layer and/or using a different number of neurons in them for a possibly better trained network model.

References

- Albert Antony, "Metabolic Syndrome, A Comprehensive Dataset on Risk Factors and Health Indicators," [Online]. Dec. 2023.
- Technogineer, "How to use Neural network (NN) toolbox in MATLAB?," [Online Video]. Apr. 26, 2020.
- Nuruzzuman Faruqui, "Diabetes Prediction using Deep Learning," [Online Video]. Sep. 23, 2021.
- Salman Mohagheghi, "Designing Multilayer Perceptron (MLP) Artificial Neural Networks in Matlab," [Online Video]. Apr. 15, 2020.

Acknowledgements

This project received support by the instructor of the Machine Learning course, Professor Dr. Tahira Mahboob, PhD and her Teaching Assistant, Ma'am Malaika Waheed at Information Technology University, Lahore, Pakistan. It also received additional support from a registered dietition, Hodaa Samad.

Thank you