Predicting Diabetes Through Perceptions: A Novel Approach to Early Detection and Risk Assessment

Santiago Reyes Vargas  
The University of Adelaide

[SantRV/DLF-Diabities-Perceptron: Predict Diabities using perceptron. (github.com)](https://github.com/SantRV/DLF-Diabities-Perceptron)

Image showing research summary

Abstract

This is the abstract of my assessment research hello soy

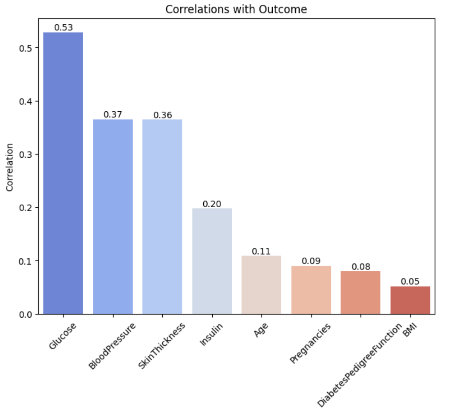
## Introduction

Representing a series of metabolic conditions associated with hyperglycaemia[[1]](#footnote-1) and caused by partial or total insulin insufficiency, diabetes mellitus affects over 422 million people globally which require daily attention to diet, lifestyle, and self-monitoring of blood glucose [1]. While this condition is generally diagnosed by measuring the plasma glucose concentration over a 2 – 3-month period [2], it has been suggested that related conditions include elevated adiposity[[2]](#footnote-2), high insulin, age and blood pressure can be used for the assessment of diabetes existence on a patient [3].

With the aim of developing a software solution that could allow both patients and health practitioners to predict a diabetes diagnosis prior to measuring plasma glucose levels, this research paper focuses on the implementation of a Perceptron (P) and a Multi-layer Perceptron (MLP) model with varied parameters configurations. Hence, this paper builds on the experimentation conducted on the data with deep-learning techniques and data processing methodologies which include data normalisation, layers architecture, and batch optimisation.

The dataset provided by the National Institute of Diabetes and Digestive and Kidney Diseases consists of a single csv file with 768 records of females aged 21 and over of Pima Indian heritage. The dataset contains 8 features namely: pregnancies, glucose, blood pressure, skin thickness, insulin levels, BMI, diabetes pedigree function, and age. One target feature is present as a binary value that indicates the diagnosis of diabetes. Therefore, the performance of the proposed models is limited to the data provided and can only provide a binary forecast, not a diagnostic result for all types of diabetes.

Preliminary data analysis suggests that the existence of strong correlations between glucose, blood pressure and skin thickness and a positive diabetes diagnosis [Figure 1]. Also, it was found that the distribution of the data is skewed towards no-diabetes diagnosis [Figure 2]. Hence, it becomes crucial to normalise the data during training to accelerate convergence and generalise the model given the low number of positive diabetes datapoints [4].



A graph with a blue and yellow bar

Description automatically generated with medium confidence

## Literature Background

Deep neural networks implementations have expanded across the medical field over the last two decades within the areas of medical image analysis, clinical diagnosis, and biological modelling [5]. The predictive diagnosis of diabetes has been addressed by numerous competitors, thus in this section we briefly review the current literature that aligns with this project’s aims.

Machine learning approaches taken include the use of support vector machines (SVM) and K-means clustering algorithms for diagnosis. While research conducted by Alseema et al suggest that information on sex, smoking and family history of diabetes can improve predictions [6], the outcomes of studies by Abnoosian et al suggests that a framework with strong data pre-processing, k-folds cross-validation and grid search can achieve high average accuracy, precision, recall, F1-score, and AUC values of 0.9887, 0.9861, 0.9792, 0.9851, and 0.999, respectively [7].

In addition, the use of scheduled learning rates during model training seem to provide further performance improvement by allowing neural networks to escape local minima and expand the search space exploration over the epoch iterations. This allows the model to find multiple local minima which can be then ensembled to further target a global minimum [8]

## Methodology

The proposed neural networks were developed using Pytorch and Numpy and the main approach taken was to split data into a data loader with batch size of 5, 10, 50 and 100, normalise the data during training using dropout and batch Norm1, have a series of 0 to 3 hidden layers with nodes from 16 to 64, and to implement a linear learning rate scheduler.

### 3.1. Neural Network models

Three neural networks were developed with increasing complexity in the number of hidden layers, all are fully connected and have a one-node output layer that denotes the final binary prediction [Figure 3]. For data pre-processing the models expect a target feature with values of 0 or 1 instead of -1 and 1 as provided in the dataset.

**3.1.1. Perceptron:**

The perceptron is the most basic model with one linear layer with 8 input nodes <s>and one sigmoid activation function.

### Model Evaluation

Equations

Loss function and optimiser

## Perceptron Application

A description of the method. This will typically re-quire explaining some part of the algorithm in detail and providing examples illustrating its effects and deficiencies. If you propose an improvement then you should describe how your method works, in enough detail that a reasonably skilled person would be able to implement it (30 points).

Description includes all key points of the method and would be reproducible.

## Experiments & Analysis

Changing layers and outputs

Describe the tests you have run, and your motivation for having run them. Report the results of the tests and the conclusions that you have drawn. The goal is not to show that your method outperforms all comparators, but rather that you understand what the method aims to achieve, and can

devise, execute, and report upon a set of tests which demonstrate whether it does so. If you have improved upon the base method then you have an opportunity here to show that your improvement is well motivated, and possibly even that it works (30 points)

Clear description of experiments and the aims of your experimental design. Results displayed in a compact and easy to read format. Analysis of results draws accurate conclusions about the method.

## Conclusion

Summarise what you have learned from the process, including ideas about what you could do in

the future to improve the method you are reporting on (10 points)

Clear summary of major project design choices. Ideas for future work based on experimental evidence.

## References

|  |  |
| --- | --- |
| [1] | WHO, “Diabetes,” 5 April 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diabetes. [Accessed 10 September 2023]. |
| [2] | A. M. Egam and S. F. Dinneen, “Diabetes: Basic Facts - What is diabetes?,” *Medicine,* vol. 50, no. 10. https://www.sciencedirect.com/science/article/pii/S1357303922001797, pp. 615-618, 2022. |
| [3] | J. A. Luchsinger, “Type 2 Diabetes and Related Conditions in relation to dementia: An opportunity for prevention?,” *Journal of Alzheimer’s Disease,* vol. 20, no. 11. https://content-iospress-com.eu1.proxy.openathens.net/download/journal-of-alzheimers-disease/jad091687?id=journal-of-alzheimers-disease%2Fjad091687, pp. 723-736, 2010. |
| [4] | L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu and L. Shao, “Normalization techniques in training DNNs: Methodology, analysis and application,” *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,* vol. 45, no. 8. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10056354, p. 10173, 2023. |
| [5] | H. Malmgren , M. Borga and L. Niklasson, Artificial neural networks in medicine and biology, New York: Springer. https://link.springer.com/book/10.1007/978-1-4471-0513-8, 2000. |
| [6] | M. Alssema, D. Vistisen, W. M. Heymans, G. Nijpels, P. Z. Zimmet, P. Z. Shaw and M. Elisson, “Risk scores for predicting type 2 diabetes: using the optimal tool,” *PubMed,* no. https://pubmed.ncbi.nlm.nih.gov/21660635/, 2011. |
| [7] | K. Abnoosian, R. Farnoosh and M. Behzadi, “Prediction of diabetes disease using an ensemble of machine learning multi-classifier models,” *BMC Bioinformatics,* no. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-023-05465-z#:~:text=Prediction%20of%20diabetes%20disease%20using%20an%20ensemble%20of,...%204%20Machine%20learning%20models%20Single%20models%20, pp. 10-94, 2021. |
| [8] | J. Yang and F. Wang, “Auto-ensemble: An adaptive learning rate scheduling based deep learning model ensembling,” *IEEE,* no. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9274468, 2020. |

1. High amounts of glucose in the blood. [↑](#footnote-ref-1)
2. Body fat concentration [↑](#footnote-ref-2)