Machine-Learning (neural network) driven algorithmic classification of Type 1 or Type 2 diabetes at the time of presentation significantly outperforms experienced clinician classification

CAR Sainsbury, R Muir, J Osmanska, GC Jones

glucose.ai Research Group, Department of Diabetes, Gartnavel General Hospital, Glasgow, Scotland

http://glucose.ai
@csainsbury
c.sainsbury@nhs.net

Background / Aims

Classification of type of diabetes at the point of diagnosis may not always be straightforward. We aimed to develop an algorithmic approach to the problem, using data available within the SCI-Diabetes dataset. Metrics that would be routinely available at the time of first presentation were chosen for inclusion. To asses the potential clinical utility - and to benchmark performance - a subset of individuals were presented to experienced clinicians who classified as either Type 1 or Type 2 diabetes. The accuracy of classifications generated for individuals within the subset by both algorithm and by clinicians were compared, using the established diagnosis within SCI diabetes as the comparator diagnosis. Individuals were included in the analysis only if they had a date of diagnosis at least 12 months from the data extraction date, in order to ensure a high likelihood of a stable and correct diagnosis being achieved.

Data was prepared for analysis using R(1), and analysis code was written in Python(2). An artificial neural network was chosen for the algorithmic approach, implemented in Tensorflow (3) (written using the Keras library (4))

Methods

All individuals within our region with a recorded diagnosis of Type 1 or Type 2 diabetes, diagnosed at least 1 year prior to data-extraction date were identified. Individuals were included within the analysis if they had an HbA1c, Systolic / Diastolic Blood Pressure and BMI measurement available at the time of diagnosis in addition to ethnicity data. Age and sex were recorded. Individuals with missing data were removed from the analysis ie no imputation was attempted.

Categorical data (ethnicity) was encoded using one-hot encoding. Feature scaling was undertaken on continuous data.
The data was divided into training and testing subsets, with the training subset comprising 80% of the total dataset. A simple artificial neural network (ANN) with a single hidden layer of 60 nodes was trained on the training subset. The output layer consisted of a single node with a sigmoid activation function.

A further subset of the test subset was generated for clinician classification. Data from individuals in this subset were presented in batches of 100 to experienced clinicians within our clinic. The proportion of individuals with a recorded diagnosis of Type 1 Diabetes within each batch was varied at random between 0.05 and 0.5 to reduce the possibility of clinicians inferring a diagnosis from the proportion already identified.

The ANN model was applied to the test subset, with the output being a probability for each individual of the correct diagnosis being Type 1 Diabetes. A Receiver Operator Characteristic (ROC) curve was generated using these probabilities. For Sensitivity / Specificity analyses a threshold of 0.2 was applied to the probability value – if the probability was above this threshold a diagnosis of Type 1 Diabetes was deemed to be predicted.

In the case of the subset of patients presented to clinicians for classification, a confusion matrix was generated for both algorithmic and clinician classification outcomes, and Sensitivity, Specificity, Positive Predictive Value and Negative Predictive value were calculated.

Results

There were 49995 individuals in the total dataset with a full set of data available for analysis. There were 9999 individuals were within the test subset, with 791 of these individuals appearing in the in the clinician classification subgroup. A total of 800 individuals were presented to clinicians, but it was later noted that there were 9 subjects duplicated within this subset - these duplicates were therefore removed.

The AUC of ROC curve for test subset was 0.96 (Figure 1). Given that the clinician classification is categorical rather than a continuous probability, a ROC curve cannot be generated.

Subgroup physician classification: Sensitivity 0.72, Specificity 0.92, Positive Predictive Value (PPV) 0.89, Negative Predictive Value (NPV) 0.84.
ANN classification: Sensitivity 0.84 ($p<0.01$ for comparison with physician), Specificity 0.96 ($p = 0.02$), PPV 0.93, NPV 0.85 (Table 1).

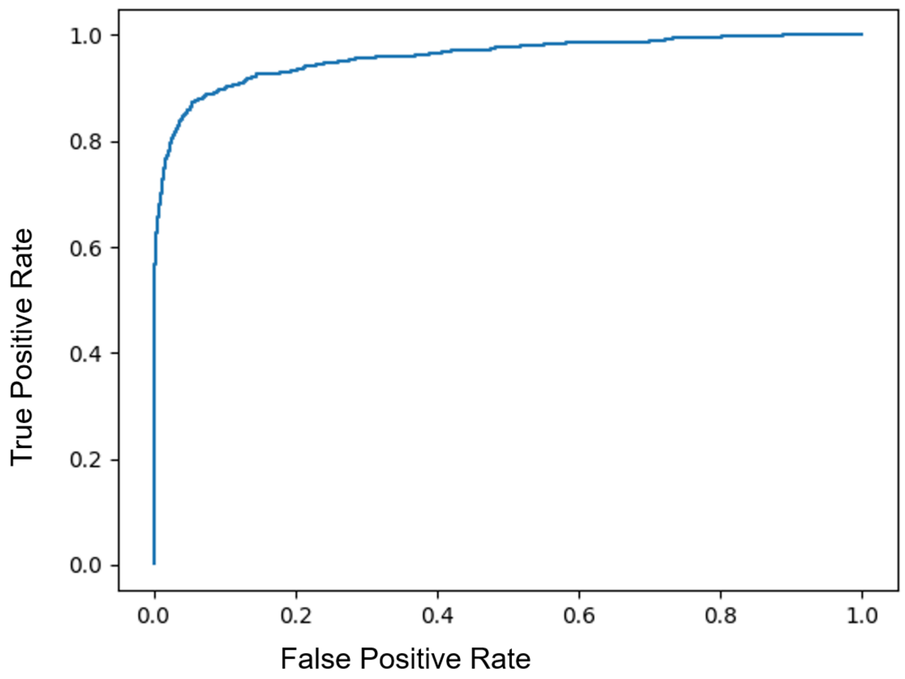

Figure 1. Receiver Operator Characteristic Curve illustrates the diagnostic ability of the neural network to predict a diagnosis of Type 1 Diabetes, as the discrimination threshold is varied. n = 9999

	Clinician Classification	ANN Classification	p value for comparison
Sensitivity	0.72	0.84	<0.01
Specificity	0.92	0.96	0.02
Positive Predictive Value	0.89	0.93	
Negative Predictive Value	0.84	0.85	

Table 1. Sensitivity, Specificity Positive Predictive Value and Negative Predictive Value derived from confusion matricies generated from ANN and clinician classification of individuals
n = 791

Summary and Conclusions

This relatively simple ANN significantly outperforms clinicians at diabetes classification using a limited dataset for decision making. The dataset in this analysis is confined to individuals already classified as having a diagnosis of either Type 1 or Type 2 Diabetes, and relies on this ‘gold standard’ diagnosis being correct - which is unlikely to be the case in reality. Further development of this approach will potentially include the use of unsupervised learning approaches to generate clusters, representations of which might then be used in a multi-class classification algorithm.

References

(1) https://www.r-project.org/
(2) https://www.python.org/
(3) https://keras.io/
(4) https://www.tensorflow.org/