RESEARCH ARTICLE
Published: February 25th, 2025
doi: xxxxxxx

Using Neural Networks to Predict Diabetes Diagnosis through Backpropagation Algorithm

Abwabul Jinan[1], Ryan Rinaldi Hadistio[2], Dede Fika Suryani[3]

[1,2]Informatics, Satya Terra Bhinneka University, [3]Data Science and Artificial Intelligence, University of North Sumatra

Diabetes is a chronic disease caused by the body's inability to produce or effectively use insulin. This condition can lead to various serious complications. The backpropagation algorithm is one of the models found in Artificial Neural Networks (ANN) that employs supervised learning. This algorithm is often used to solve complex problems due to its training through learning methods. This study aims to develop an ANN model using the backpropagation algorithm to predict diabetes based on existing symptoms. By utilizing patients' medical records, this model is expected to provide accurate predictions regarding diabetes diagnosis.

Keywords: backpropagation, diabetes prediction, neural network

OPEN ACCESS
ISSN XXX-XXX (Online)

**Correspondence:*
Abwabul Jinan
abwabul.jinan7@gmail.com

Received: 30th July 2024
Accepted: 22nd January 2025
Published: 25th February 2025

Citation:
Jinan, A., Hadistio, R. R., & Suryani, D. F. (2025) Using Neural Networks to Predict Diabetes Diagnosis through Backpropagation Algorithm. JACIIS (Journal of Advance in Computational Intelligence and Information Systems), 1(1), 24-28.
doi: xxxxxxxxxxxxx

INTRODUCTION

Diabetes mellitus is a metabolic disorder characterized by chronic hyperglycemia due to insulin secretion deficiency, insulin resistance, or both. According to the International Diabetes Federation (IDF) data from 2021, approximately 537 million adults worldwide live with diabetes. This number is projected to increase to 643 million by 2030 and 783 million by 2045 (IDF, 2021). This disease significantly impacts patients' quality of life and imposes a considerable economic burden on global healthcare systems. Early detection and effective management of diabetes are crucial to preventing serious complications such as cardiovascular disease, neuropathy, nephropathy, and retinopathy (American Diabetes Association, 2019). With technological advances in computer science, particularly in developing machine learning and artificial intelligence algorithms, new methods for disease diagnosis and prediction are emerging. One prominent method is using Artificial Neural Networks (ANNs). The Neural Network architecture used is a multilayer perceptron with a single hidden layer. The training process is conducted over several epochs until the error reaches a minimum value.

A Neural Network is a computational model inspired by the human brain's way of recognizing patterns and processing information. ANNs consist of layers of interconnected neurons, where each neuron in one layer receives input from neurons in the previous layer and sends output to neurons in the next layer. Learning in ANNs involves adjusting the weights of the connections between neurons based on the error produced, aiming to minimize this error. The backpropagation algorithm is one of the most commonly used learning techniques in ANNs. This algorithm works by propagating the error from the output layer to the input layer through the hidden layers, allowing the weights to be updated iteratively (Kaur & Kumari, 2020).

A Neural Network model with the backpropagation algorithm can be used to predict diabetes based on existing symptom data. By training with medical record data, this model can provide accurate predictions regarding diabetes diagnosis (Park & Ha, 2019). In the context of diabetes prediction, Neural Networks with the backpropagation algorithm can be trained using symptom data and patient medical records to produce a model capable of predicting the likelihood of someone having diabetes. Some symptoms often associated with diabetes include frequent urination, excessive thirst, unexplained weight loss, fatigue, blurred vision, and slow-healing sores (Chen & Asch, 2017). Using this data, a Neural Network model can be trained to recognize patterns indicative of diabetes. This study aims to develop and test a Neural Network model with the backpropagation algorithm to predict diabetes diagnosis based on patient symptom data. This research is expected to contribute significantly to healthcare, particularly in early detection and management of diabetes, and demonstrate the potential of artificial intelligence technology in improving the accuracy of medical diagnoses.

METHOD

Neural Networks simulate the brain's structure and processes in recognizing complex patterns. A Neural Network consists of an input layer, hidden layers, and an output layer. The backpropagation algorithm is used to train the network by minimizing the error between the expected and generated output.

The backpropagation algorithm consists of several steps (Yildirim & Sensoy, 2019):

1. Initialize weights with small random values.
2. Forward propagate, calculating the output for each unit in the hidden and output layers.
3. Calculate the error and backpropagate to update the weights.
4. Repeat the process until the error reaches a specified minimum value.

Backpropagation Algorithm

Backpropagation is a model within ANN that uses supervised learning. This algorithm is often employed to solve complex problems because it is trained using a learning method. In this network, pattern pairs consisting of input patterns and desired patterns are provided (Hayadi et al., 2021).

The training algorithm for a network with a single hidden layer (using a binary sigmoid activation function) is as follows:

Step 0: Initialize all weights with small random numbers.

Step 1: If the stopping condition is not met, proceed to steps 2-9.

Step 2: For each training data pair, perform steps 3-8.

Phase I: Forward Propagation

Step 3: Each input unit receives the signal and transmits it to the hidden units above it.

Step 4: Calculate all outputs in the hidden units z_j (j = 1,2…,p)

$$z\_net_j = v_{j0} + \sum_{i=1}^{n} x_i v_{ji}$$

$$z_j = f(z\_net_j) = \frac{1}{1+e}$$

Step 5: Use the activation function Y = f(net).

Phase II: Backward Propagation

Step 6: Calculate the δ factor for the output unit based on the error in each output unit y_k (k = 1,2,….,m)

δk = (tk - yk) f '(y_ink)

Step 7: Calculate the δ factor for the hidden unit based on the error in each hidden unit z_j (j = 1,2,….,p)

Phase III: Weight Adjustment

Step 8: Calculate all weight changes: w_{jk} (baru) = w_{kj} (lama) + Δ w_{kj} (k = 1,2,…,m ; j = 0,1,…,p). In this phase, weight adaptation processes are carried out for each weight between the input layer and hidden layer, and between the hidden layer and output layer (Agustin & Prahasto, 2012).

RESULTS AND DISCUSSION

This study uses medical records of diabetic patients to train and test the Neural Network model. The symptom data used includes frequent urination, excessive thirst, weight loss, and others.

Table 1. Symptom Data

No	Symptom Code	Symptom/Characteristic
1	G1	Frequent urination
2	G2	Excessive thirst
3	G3	Unexplained weight loss
4	G4	Extreme fatigue
5	G5	Blurred vision
6	G6	Slow-healing sores

This table outlines the specific symptoms and characteristics used in the study's dataset to train the Neural Network model for predicting diabetes. Each symptom is coded for easy reference during data analysis. The following table shows an example of symptom data used in the study:

Table 2. Symptom Data (Initial Details)

No	Symptom Code	Symptom/Characteristic	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
1	G1	Frequent urination	1	0	1	1	0
2	G2	Excessive thirst	1	1	0	1	0
3	G3	Unexplained weight loss	0	0	1	1	0
4	G4	Extreme fatigue	1	1	1	0	1
5	G5	Blurred vision	0	1	0	1	1
6	G6	Slow-healing sores	0	0	1	0	1

This table represents the symptom data for five patients, where each symptom is coded and recorded as either present (1) or absent (0). This data is used to train the Neural Network model to predict the likelihood of a diabetes diagnosis.

The next step is to transform the symptom data into a format suitable for use by the Neural Network model. This includes data normalization, grouping, and representation in a format suitable for model input and output.

1. **Data Normalization:** If needed, scale the symptom data values to fall within the range [0, 1].
2. **Grouping:** Organize the data in a table format used by the Neural Network model.
3. **Input and Output Representation:** Classify the data into input (symptoms) and output (diagnosis).

The following table shows the transformed symptom data ready for use by the Neural Network model:

Table 3. Transformed Symptom Data for Neural Network Model

No	Symptom Code	Symptom/Characteristic	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
1	G1	Frequent urination	1	0	1	1	0
2	G2	Excessive thirst	1	1	0	1	0
3	G3	Unexplained weight loss	0	0	1	1	0
4	G4	Extreme fatigue	1	1	1	0	1
5	G5	Blurred vision	0	1	0	1	1
6	G6	Slow-healing sores	0	0	1	0	1
Output		Diabetes Diagnosis (0 = No, 1 = Yes)	1	0	1	1	0

In this table, the columns represent different patients, and the rows represent different symptoms coded as G1 to G6. Each symptom is recorded as either present (1) or absent (0). The last row indicates the output, which is the diabetes diagnosis for each patient, represented as 0 (No) or 1 (Yes). This transformed data is now in a suitable format for input into the Neural Network model, allowing it to learn from the symptom patterns and predict diabetes diagnosis accurately.

Diagnosis:

- Diabetes diagnosis is given based on the combination of existing symptoms.
- Patients 1, 2, 3, and 4 are diagnosed with diabetes because they have several key symptoms often associated with diabetes.
- Patient 5 is not diagnosed with diabetes due to having fewer key symptoms.

Data Normalization:

- Symptom data is already in binary form (0 and 1), so no further normalization is required.

Data Grouping:

- Data is organized into a table format used by the Neural Network model.
- Each column represents a symptom, and each row represents a patient.

Explanation and Analysis of Results:

1. **Identification of Key Symptoms:**

- The most common symptoms among patients diagnosed with diabetes are "Frequent urination," "Excessive thirst," "Extreme fatigue," and "Unexplained weight loss."
- Patients 1 and 4 exhibit both "Frequent urination" and "Excessive thirst," which are strongly associated with diabetes.
- The symptom "Extreme fatigue" is present in all patients diagnosed with diabetes.

2. **Diabetes Diagnosis:**

- The diagnosis model indicates that patients presenting with a combination of key symptoms are likely to be diagnosed with diabetes. This aligns with medical literature, which recognizes symptoms like "Frequent urination," "Excessive thirst," and "Unexplained weight loss" as strong indicators of diabetes.
- Patient 5, showing fewer key symptoms associated with diabetes, was not diagnosed with the condition, which is consistent with the pattern observed in the data.

3. **Symptom Combinations:**

- Certain symptom combinations, such as "Frequent urination," "Excessive thirst," and "Unexplained weight loss," significantly increase the probability of a diabetes diagnosis.
- Patients 3 and 4 demonstrate that even if one key symptom is absent, the presence of other symptoms can still lead to a diabetes diagnosis, highlighting the importance of a combination of symptoms rather than a single symptom.

4. **Diagnosis Correction:**

- Symptoms that are less common but still relevant, such as "Blurred vision" and "Slow-healing sores," should also be considered in the diagnostic process.
- Patient 2 was diagnosed with diabetes due to a combination of "Excessive thirst" and "Extreme fatigue," along with additional symptoms, indicating that while individual symptoms are important, the overall combination plays a crucial role in diagnosis.

This analysis underscores the importance of considering multiple symptoms and their combinations when diagnosing diabetes, as well as recognizing the role of less common symptoms in the diagnostic process. The model's ability to identify these patterns reflects its effectiveness in predicting diabetes based on symptom data.

CONCLUSION AND SUGGESTION

The Neural Network model with the backpropagation algorithm, using this symptom data, demonstrates effectiveness in predicting diabetes diagnosis. Data analysis reveals that:

1. Certain Symptom Combinations: Specific combinations of symptoms are very strong indicators of diabetes. This aligns with the model's ability to recognize these patterns and predict the likelihood of diabetes accurately.
2. Model Reliability: The model is reliable, providing accurate predictions with a low error rate. This suggests that the Neural Network effectively learns from the data and generalizes well to new cases.
3. Early Detection and Management: Accurate and early detection of diabetes symptoms can significantly enhance disease management and treatment. Identifying diabetes based on reported symptoms allows for timely intervention and better patient outcomes.

With these results, the developed Neural Network model can serve as an effective diagnostic tool for detecting diabetes based on reported symptoms. Further implementation and validation with larger and more diverse datasets will help to improve the model's accuracy and reliability.

REFERENCES

American Diabetes Association. (2019). Standards of Medical Care in Diabetes. *Diabetes Care, 42*(1), S81-S89. https://doi.org/10.2337/dc19-S008

Chen, J. H., & Asch, S. M. (2017). Machine Learning and Prediction in Medicine – Beyond the Peak of Inflated Expectations. *New England Journal of Medicine, 376*(26), 2507-2509. https://www.nejm.org/doi/abs/10.1056/NEJMp1702071

Hayadi, B. H., Sudipa, I. G. I., & Windarto, A. P. (2021). Model Peramalan Artificial Neural Network pada Peserta KB Aktif Jalur Pemerintahan Menggunakan Artificial Neural Network Back-Propagation. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(1), 11–20. https://doi.org/10.30812/matrik.v21i1.1273

International Diabetes Federation. (2021). *IDF Diabetes Atlas, 10th Edition*. https://www.diabetesatlas.org

Kaur, H., & Kumari, V. (2020). Predictive Modelling and Analytics for Diabetes Using a Machine Learning Approach. *Applied Computing and Informatics, 16*(1/2), 10-17.

Redjeki, S. (2013). Perbandingan Algoritma Backpropagation untuk Identifikasi Penyakit. *Semin. Nas. Apl. Teknol. Inf*, *3*(2), 1-5.