BRAIN STROKE CLASSIFICATION

deep learning

Abstract

Brain stroke is a life-threatening medical disorder caused by the inadequate blood supply to the brain. After the stroke, the damaged area of the brain will not operate normally. As a result, **early detection is crucial** for more effective therapy.

Abstract

Given a patient's **medical and personal history**, deep learning methods can be used to predict with high accuracy whether the patient will have a stroke

Deep Learning Methods

Logistic regression

supervised learning algorithm that is used for binary classification. It is a linear model that is used to predict the probability of a binary outcome

Multilayer perceptron (MLP)

neural network that consists of multiple layers of artificial neurons (units) connected together, with each layer fully connected to the next one. It is used for supervised learning tasks, such as classification

Deep Learning Methods

Recurrent neural network (RNN)

artificial neural network that is designed to process sequential data, such as time series or natural language. RNNs have a "memory" in the form of hidden states, which allows them to capture dependencies between the input data at different time steps.

Long short-term memory (LSTM)

type of recurrent neural network (RNN) that is specifically designed to capture long-term dependencies in sequential data. LSTM networks are composed of "memory cells" that can store and retrieve information over a long period of time, as well as input, output, and forget gates that control the flow of information into and out of the memory cells

Data Set

origin shape:
4981 rows, 11 columns.
10 features and 1 label - 'stroke'
248 rows '1', 4733 rows '0'

new shape:
9466 rows, 11 columns.
10 features and 1 label - 'stroke'
4733 rows '1', 4733 rows '0'

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
1	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
2	Female	49.0	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
3	Female	79.0	1	0	Yes	Self-employed	Rural	174.12	24.0	never smoked	1
4	Male	81.0	0	0	Yes	Private	Urban	186.21	29.0	formerly smoked	1

Data Set - features

The features selected for the data set are the patient's personal and medical history, and affect the possibility of having a stroke.

Glucose - medical history

Elevated levels of glucose in the blood, also known as hyperglycemia, can increase the risk of stroke. High blood glucose levels can damage blood vessels and increase the risk of complications such as heart disease and stroke

Marriage - personal history

Marriage has been linked to a lower risk of stroke, while not being married has been linked to a higher risk of stroke. married people tend to have better physical and mental health overall, which can reduce their risk of stroke. They also tend to have more social support, which can help them manage stress and other risk factors for stroke.

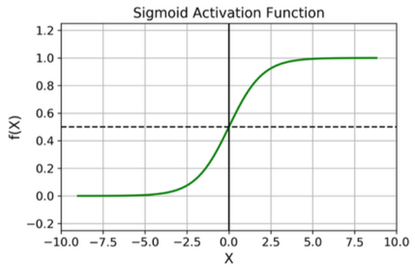
work process

The purposeis is to improve the classification accuracy.
We will describe how each model works on our data, in addition the accuracy and the loss.

machine learning
~logistic regression~

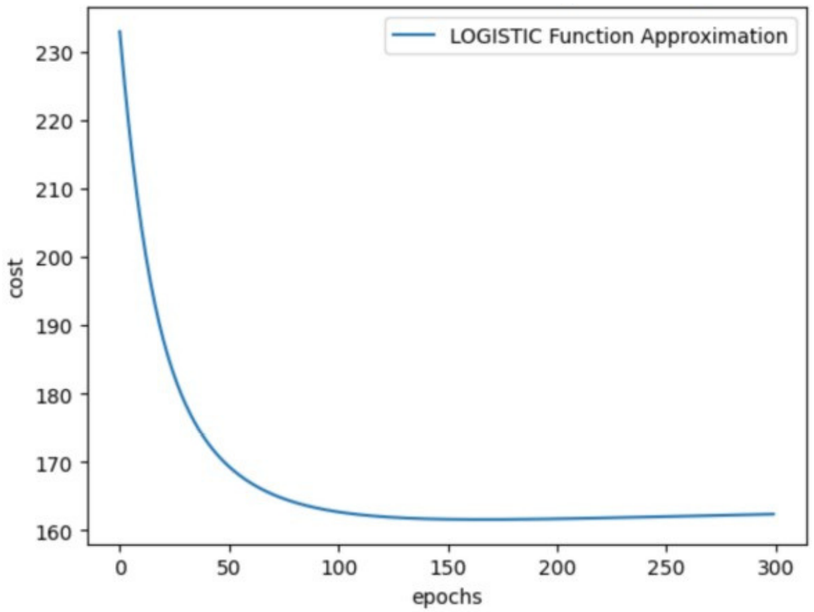
simple neural networks
~MLP~

complex networks
~RNN, LSTM~

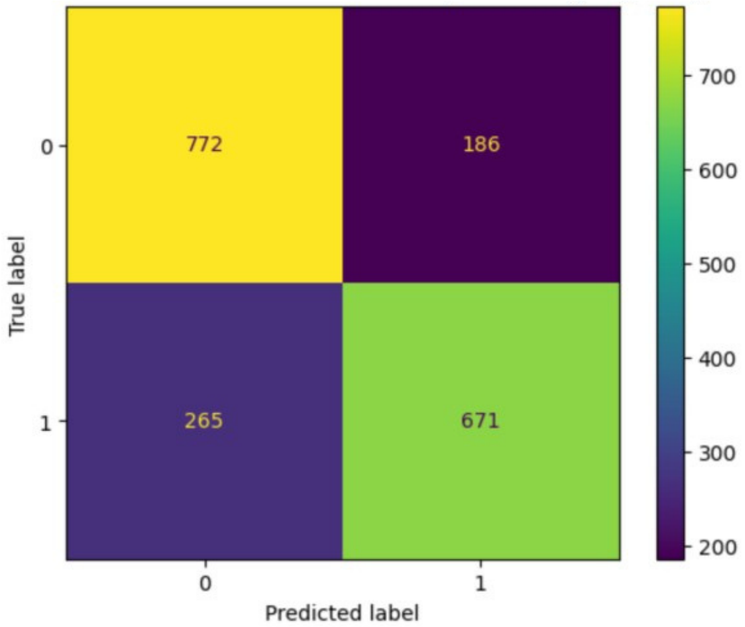


Logistic Regression

The model is trained using stochastic gradient descent with a learning rate of 0.01, using a cross-entropy loss function and the sigmoid_cross_entropy_with_logits function to calculate the loss

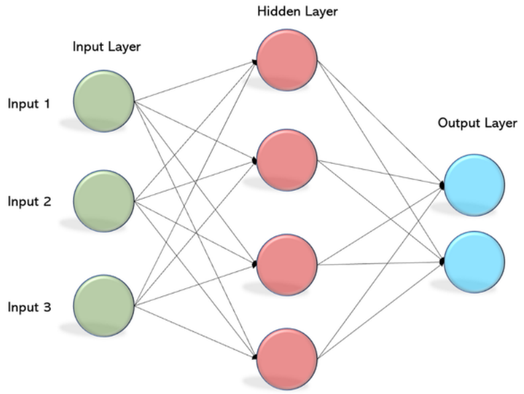


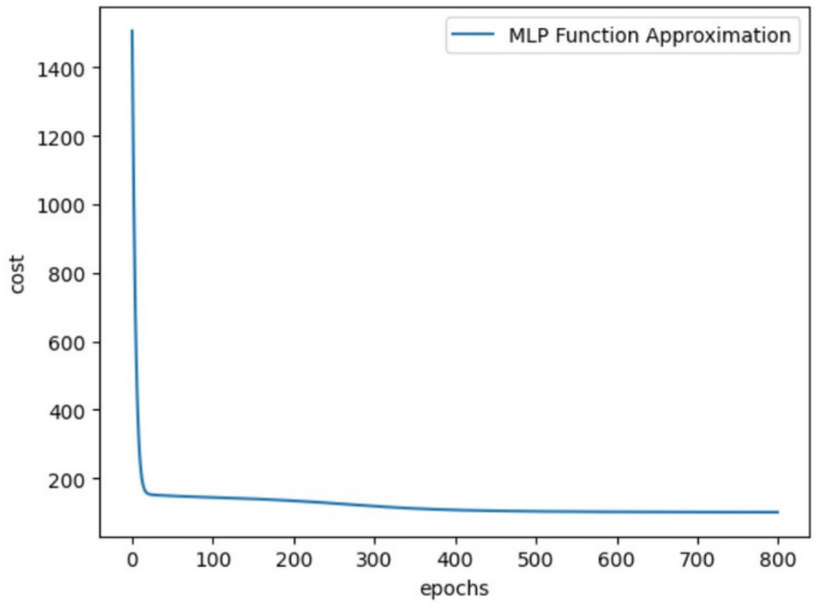
accuracy: 75.5%



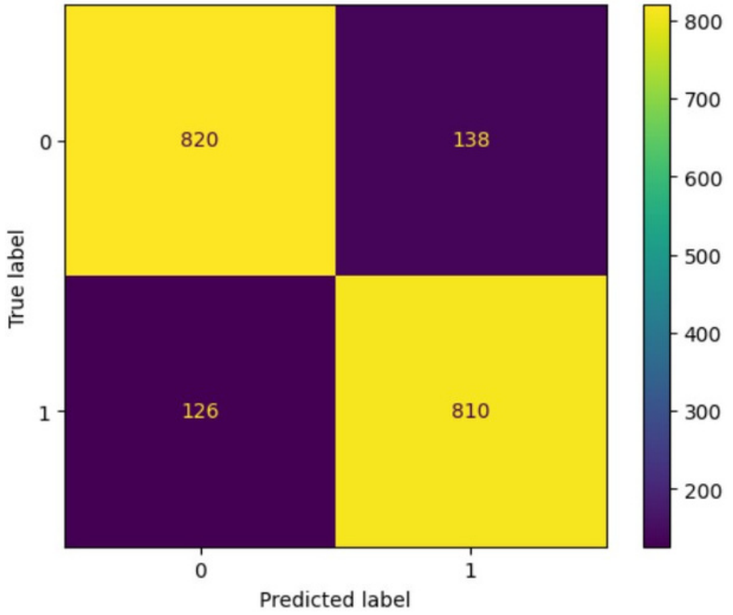
Multilayer Perceptron (MLP)

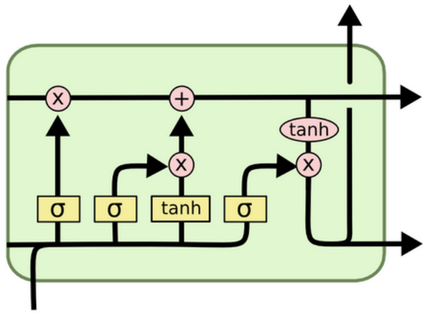
MLP has two layers: a hidden layer with 10 neurons and an output layer with 1 neuron





accuracy: 86%
improve: 8.5%
TP: improve of 20%
FP: improve of 6%





RNN & LSTM

The data is not sequential or time series data, and the order of the data points is not important

↓

LSTM model needs to receive 3D data as input in the following shape: (samples, time-series, features)

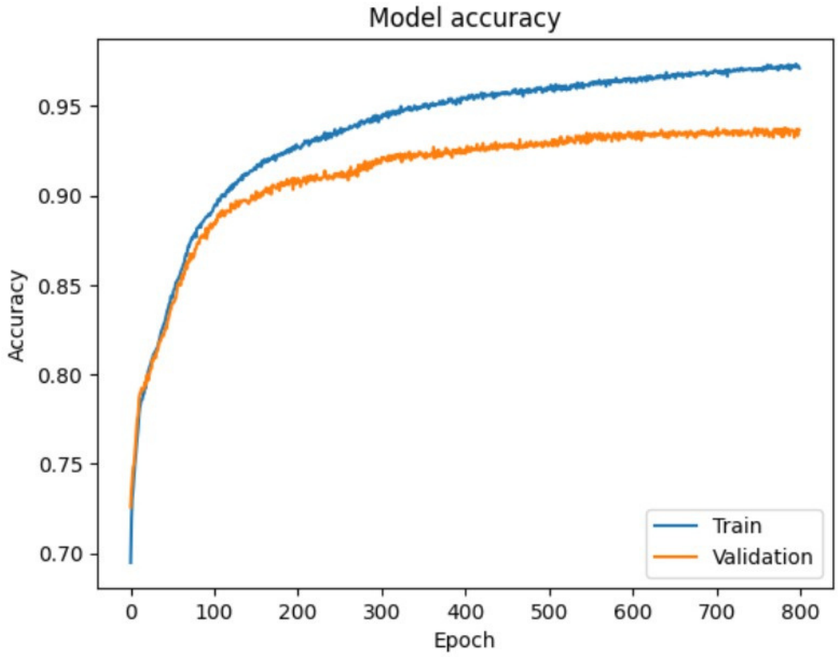
↓

The data needs to be transformed to fit LSTM model,
and therefore we added a time-series dimension and set it to 1.

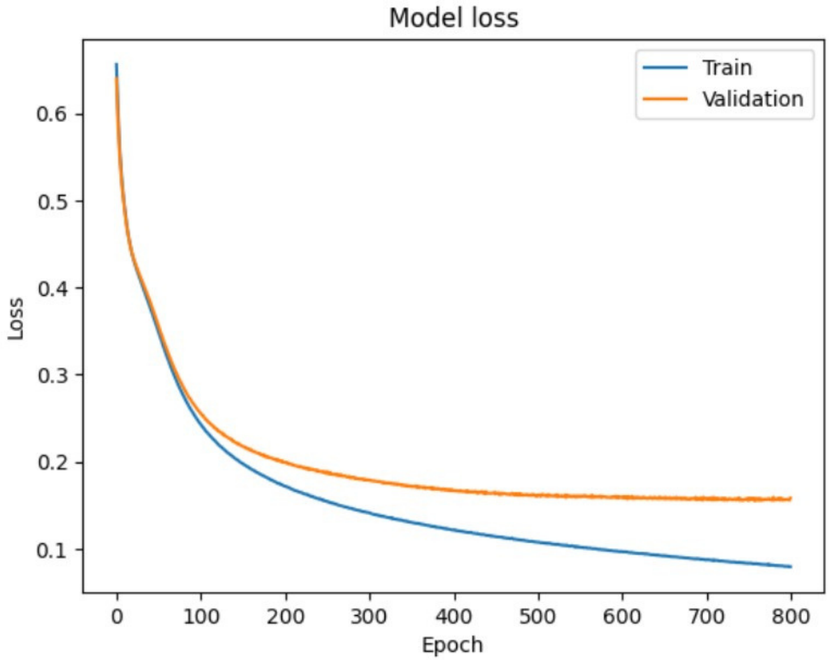
↓

This means that our model will actually treat the input data as a single snapshot of the data,
instead of a sequence of data over time.

RNN & LSTM



accuracy -
train: 97%
validation: 94%
test: 92%
improve: 6%



conclusion

process of improving results using different deep learning methods. First, we used Machine Learning method: Logistic Regression with test accuracy of 75%. Second, we used Multilayer Perceptron (MLP) with an accuracy of 86%. At last, we used a recurrent neural network (RNN) model using long short-term memory (LSTM) layers for the classification of brain stroke, the model achieved good performance, with a test accuracy of 92%.