Inside | Jurnal Sistem Informatika Cerdas
Volume 2 Nomor 1 | bulan July tahun 2024 | halaman 8-12
e-ISSN: 3031-4860

Implementation of Deep Learning in Diagnosing Stroke Disease Based on Clinical Data Parameters

Jamilatul Badriyah[1*], Anang Faktchur Rachman[2], Matsaini[3] , Muhammad Yasir Zain[4]
[1,2]Program Studi Teknik Informatika, Universitas Madura
[3]Program Studi Teknik Industri, Universitas Madura
[4]Program Studi Bisnis Digital, Universitas Madura

Corresponding Author:
Jamilatul Badriyah, Universitas Madura, Pamekasan, 69311 – 69383, Indonesia
Email: mila@unira.ac.id, anang@unira.ac.id, matsaini@unira.ac.id, yasir@unira.ac.id

Abstract

Stroke is a medical condition that requires rapid and accurate diagnosis to increase the chances of patient recovery and reduce the risk of long-term complications. This research investigates the application of deep learning techniques to diagnose stroke based on clinical data parameters. It develops and applies deep learning models that use various neural network architectures, such as convolutional neural networks (CNN). The dataset used is an open dataset in analyzing the dataset that includes patient clinical information such as hypertension, cardiac history, married status, employment level, body mass index (BMI), smoking and glucose. The model was trained using a dataset consisting of thousands of medical records of patients with stroke and without stroke. Model evaluation was conducted using performance metrics such as accuracy, precision, recall, and F1-score to assess the effectiveness in classification with an accuracy value of 95.05%. The results showed that the deep learning approach significantly improved the accuracy and speed in detecting stroke compared to conventional diagnosis methods. These findings suggest that the integration of deep learning in clinical diagnostic systems can improve early stroke detection and provide a solid basis for better clinical decisions.

Keywords : Deep learning, Stroke, Classification, Artificial Neural Network, Detection Accuracy

1 INTRODUCTION

Stroke is an emergency medical condition that requires rapid and accurate treatment to minimize neurological damage and improve the patient's chances of recovery. [1]. The occurrence of a stroke often has a far-reaching impact on a patient's quality of life, and requires a precise diagnosis to ensure effective interventions [2]. Despite the existence of conventional diagnostic methods such as physical examination, CT scan, and MRI, stroke diagnosis is often subject to delays or errors, which can significantly affect clinical outcomes. [2][3]. In recent years, technological advancements have enabled the use of artificial intelligence techniques, particularly deep learning, to improve accuracy and efficiency in various medical applications. [4]. Deep learning, as a subfield of machine learning, utilizes complex artificial neural networks to analyze large and complex data, detect patterns, and make predictions [5]. In the context of stroke diagnosis, this technique has the potential to offer faster and more accurate solutions by analyzing diverse medical data.

This research focuses on the application of deep learning for stroke diagnosis based on clinical data parameters, which include various information such as medical history, laboratory test results, and data from the kaggle dataset. [6]. By utilizing deep learning models such as convolutional neural networks (CNN) and recurrent neural networks (RNN) this aims to develop and test a diagnostic system capable of identifying stroke more accurately than traditional methods.

The developed deep learning model will be trained with an open dataset consisting of thousands of patient medical records, both those that have been diagnosed with stroke and those that have not. Model performance evaluation will be conducted using metrics such as accuracy, precision, recall, and F1-score to measure the model's ability to distinguish between patients with stroke and those without [7]. By applying deep learning techniques in stroke diagnosis, this research is expected to overcome the challenges that exist in traditional diagnosis and offer a more advanced and efficient tool for early identification of stroke [8].

Increased accuracy and speed of diagnosis can not only improve clinical outcomes but also contribute to the development of a more responsive and adaptive health system [9], [10].

2 RESEARCH METHOD

2.1 Research methodology

The research method carried out in this study has several stages starting from data collection, data pre-processing, deep learning model building and evaluating and validating to see the accuracy value. The following is a detailed explanation of each stage:

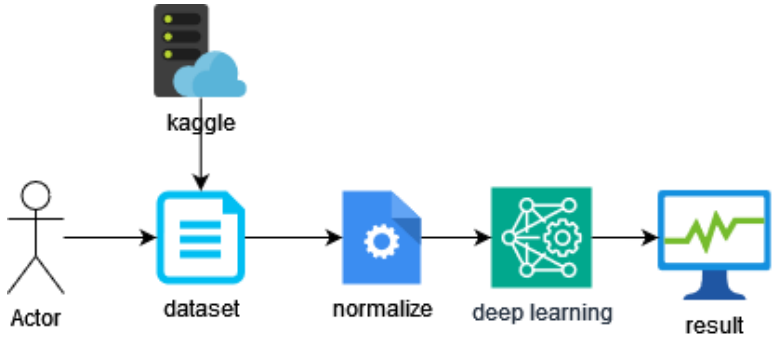

Fig 1. Research Design

2.2 Dataset

The clinical data used in this study utilized an open dataset through the associated kaggle Dataset includes relevant medical parameters such as hypertension, cardiac history, married, occupation, bmi, smoking, glucose as well as other laboratory test results. The following are the variables used in this study. Here are the details of the attributes used:

Table 1. Variabel Dataset

No	Variabel	Value	Description
1	Gender	"Male", "Female" or "Other"	patient gender
2	hypertension	0 if the patient doesn't have hypertension, 1 if the patient has hypertension	patient blood pressure
3	Age	age of the patient	Patient age
4	heart_disease	0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease	heart disease records about patients
5	ever_married	"No" or "Yes"	patient's marital status
6	work_type	"children", "Govt_jov", "Never_worked", "Private" or "Self-employed"	patient job type
7	Residence_type	"Rural" or "Urban"	type of patient residence
8	avg_glucose_level	50-300	average glucose level in blood
9	BMI	10-30	body mass index
10	smoking_status	"formerly smoked", "never smoked", "smokes" or "Unknown"	Smoking status patient
11	stroke	1 if the patient had a stroke or 0 if not	Patient stroke status

The amount of data taken amounted to 5110 data taken which is an open dataset. The following is the raw data that will be taken:

.

	A	B	C	D	E	F	G	H	I	
1	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residenc	avg_glucc	bmi
2	9046	Male	67	0	1	Yes	Private	Urban	228.69	36.6
3	51676	Female	61	0	0	Yes	Self-empl	Rural	202.21	N/A
4	31112	Male	80	0	1	Yes	Private	Rural	105.92	32.5
5	60182	Female	49	0	0	Yes	Private	Urban	171.23	34.4
6	1665	Female	79	1	0	Yes	Self-empl	Rural	174.12	
7	56669	Male	81	0	0	Yes	Private	Urban	186.21	
8	53882	Male	74	1	1	Yes	Private	Rural	70.09	27.4
9	10434	Female	69	0	0	No	Private	Urban	94.39	22.8
10	27419	Female	59	0	0	Yes	Private	Rural	76.15	N/A
11	60491	Female	78	0	0	Yes	Private	Urban	58.57	24.2
12	12109	Female	81	1	0	Yes	Private	Rural	80.43	29.7
13	12095	Female	61	0	1	Yes	Govt_job	Rural	120.46	36.8
14	12175	Female	54	0	0	Yes	Private	Urban	104.51	27.3
15	8213	Male	78	0	1	Yes	Private	Urban	219.84	N/A
16	5317	Female	79	0	1	Yes	Private	Urban	214.09	28.2
17	58202	Female	50	1	0	Yes	Self-empl	Rural	167.41	30.9
18	56112	Male	64	0	1	Yes	Private	Urban	191.61	37.5
19	34120	Male	75	1	0	Yes	Private	Urban	221.29	25.8
20	27458	Female	60	0	0	No	Private	Urban	89.22	37.8
21	25226	Male	57	0	1	No	Govt_job	Urban	217.08	N/A
22	70630	Female	71	0	0	Yes	Govt_job	Rural	193.94	22.4
23	13861	Female	52	1	0	Yes	Self-empl	Urban	233.29	48.9
24	68794	Female	79	0	0	Yes	Self-empl	Urban	228.7	26.6
25	64778	Male	82	0	1	Yes	Private	Rural	208.3	32.5

Fig 2. Dataset

The information taken is still within the shape of raw data so it has to be preprocessed. The reason of the preprocessing prepare is additionally to progress the comes about and precision of information mining for the way better. Some time recently being prepared, the preprocessing prepare is to begin with carried out. The steps of the preprocessing handle begin from information cleaning, information integration and information diminishment. Information cleaning is done by cleaning purge and conflicting values (lost esteem and loud), the moment information integration handle where combining information into one capacity, at that point information diminishment is carried out where the number of qualities utilized is as well much so it should be evacuated or erased concurring to the traits utilized.

2.2 Deep Learing Method

In this inquire about, the information that has been arranged from the preprocessing prepare will be prepared employing a profound learning demonstrate. Deep learning is a subfield of artificial intelligence (AI) that uses artificial neural network algorithms with many layers (in this case called "deep neural networks") to analyze and process data [11]. This technique allows computers to learn from data in a way that resembles human learning, enabling them to recognize complex patterns and make predictions or decisions based on the given data [12]. Through an iterative process of training and parameter adjustment, deep learning models can learn to recognize patterns and make accurate decisions from given data.

2.3 Evaluasi Model

Model evaluation using evaluation metrics is a measure or criterion used to assess model performance in the machine learning process, including in the context of disease classification such as stroke [13]. This metric helps determine how well the model is able to make predictions and provides insight into the quality of the results obtained.

3 RESULT AND ANALYSIS

This research uses rapid miner tools to create data mining models. In the initial process the dataset will be imported as input data to be analyzed. Furthermore, the use of normalization and set of reference parameters for stroke sufferers for classification. The data that has been set will then be processed in a deep learning classification model with a difference in the presentation of 60% traning data and 40% testing data. The following is a table of accuracy presentation on the use of deep learning algorithms. The following is an image of the results of accuracy using the deep learning method model

accuracy: 95.05% +/- 0.19% (micro average: 95.05%)

	true 1	true 0	class precision
pred. 1	1	5	16.67%
pred. 0	248	4856	95.14%
class recall	0.40%	99.90%	

Fig 3. Accucary Result

The results of the deep learning accuracy above show that the use of deep learning models produces a large or optimal accuracy value. Accuracy of 95.05% shows high accuracy so that it can be a reference in early detection of stroke disease. By utilizing the ability of deep learning to process large and complex datasets, it is found that the developed model can provide more accurate and faster diagnosis results. In addition, analyzing data from various other sources and applying deep learning techniques can improve the accuracy of diagnosis, enable early detection, and optimize patient care management.

4 CONCLUSION

This research has shown that the application of deep learning techniques in stroke diagnosis based on clinical data parameters can produce a very high accuracy of 95.05%. These results indicate that deep learning methods, especially using advanced neural network architectures, are able to analyze medical data very effectively to detect and classify stroke cases. The use of deep learning allows the model to learn from complex patterns in a dataset that includes various clinical parameters, including laboratory results and medical history. With an accuracy of 95.05%, the model not only demonstrated excellent performance in detecting stroke but also reduced the likelihood of misdiagnosis that can occur with traditional methods. The success of this model reflects the great potential of deep learning in improving the quality and speed of stroke diagnosis. The implementation of this technology in clinical practice can aid in the early detection and more effective treatment of stroke, as well as provide additional tools that can improve clinical decisions and patient care outcomes.

REFERENCES

[1] A. Faisal and A. Subekti, "Deep Neural Network untuk Prediksi Stroke," *JEPIN J. Edukasi Dan Penelit. Inform.*, vol. 7, no. 3, pp. 443–449, Dec. 2021, doi: 10.26418/jp.v7i3.50094.

[2] I. O. Herlistiono and S. Violina, "Model Prediksi Risiko Stroke Menggunakan Machine Learning," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 7, no. 4, pp. 1230–1238, Jul. 2024, doi: 10.31539/intecoms.v7i4.10942.

[3] P. Patmawati, "STROKE PREDICTION USING A SUPPORT VECTOR MACHINE (SVM)," *Bull. Netw. Eng. Inform.*, vol. 1, no. 1, p. 12, Apr. 2023, doi: 10.59688/bufnets.v1i1.5.

[4] A. Muliawan, A. Rizal, and S. Hadiyoso, "Heart Disease Prediction based on Physiological Parameters Using Ensemble Classifier and Parameter Optimization," *J. Appl. Eng. Technol. Sci. JAETS*, vol. 5, no. 1, Art. no. 1, Dec. 2023, doi: 10.37385/jaets.v5i1.2169.

[5] A. Muliawan, D. A. Fauziah, and F. Wiranto, "Experiment Time Series Forcasting Using Machine Learning (Case studi : Stock Value Prediction)," *PROCEEDING Int. Conf. Econ. Bus. Inf. Technol. ICEBIT*, vol. 4, pp. 834–839, Jul. 2023.

[6] "Stroke Prediction Dataset." Accessed: Aug. 08, 2024. [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset

[7] A. Alhamad, A. I. S. Azis, B. Santoso, and S. Taliki, "Prediksi Penyakit Jantung Menggunakan Metode-Metode Machine Learning Berbasis Ensemble – Weighted Vote," *JEPIN J. Edukasi Dan Penelit. Inform.*, vol. 5, no. 3, Art. no. 3, Dec. 2019, doi: 10.26418/jp.v5i3.37188.

[8] M. Gupta, P. Meghana, and K. Reddy, "Deep learning-based approach for prediction of brain stroke from MR images for IoT in healthcare," *J. Auton. Intell.*, vol. 7, Jan. 2024, doi: 10.32629/jai.v7i3.1101.

.

[9] D. A. Fauziah, A. Muliawan, and M. A. Rohim, “Stock Price Prediction of PT. Jasa Marga (Persero) Tbk Using Linear Regression Algorithm,” *PROCEEDING Int. Conf. Econ. Bus. Inf. Technol. ICEBIT*, vol. 4, pp. 810–815, Jul. 2023.
[10] N. Sakinah, T. Badriyah, and I. Syarif, “Analisis Kinerja Algoritma Mesin Pembelajaran untuk Klarifikasi Penyakit Stroke Menggunakan Citra CT Scan,” *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 7, no. 4, Art. no. 4, Aug. 2020, doi: 10.25126/jtiik.2020743482.
[11] M. Parmar, “Heart Diseases Prediction using Deep Learning Neural Network Model,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, pp. 2244–2248, Jan. 2020, doi: 10.35940/ijitee.C9009.019320.
[12] S. Subramani *et al.*, “Cardiovascular diseases prediction by machine learning incorporation with deep learning,” *Front. Med.*, vol. 10, Apr. 2023, doi: 10.3389/fmed.2023.1150933.
[13] D. A. Fauziah, A. Maududie, and I. Nuritha, “Klasifikasi Berita Politik Menggunakan Algoritma K-nearst Neighbor,” *Berk. SAINSTEK*, vol. 6, no. 2, Art. no. 2, Dec. 2018, doi: 10.19184/bst.v6i2.9256.