

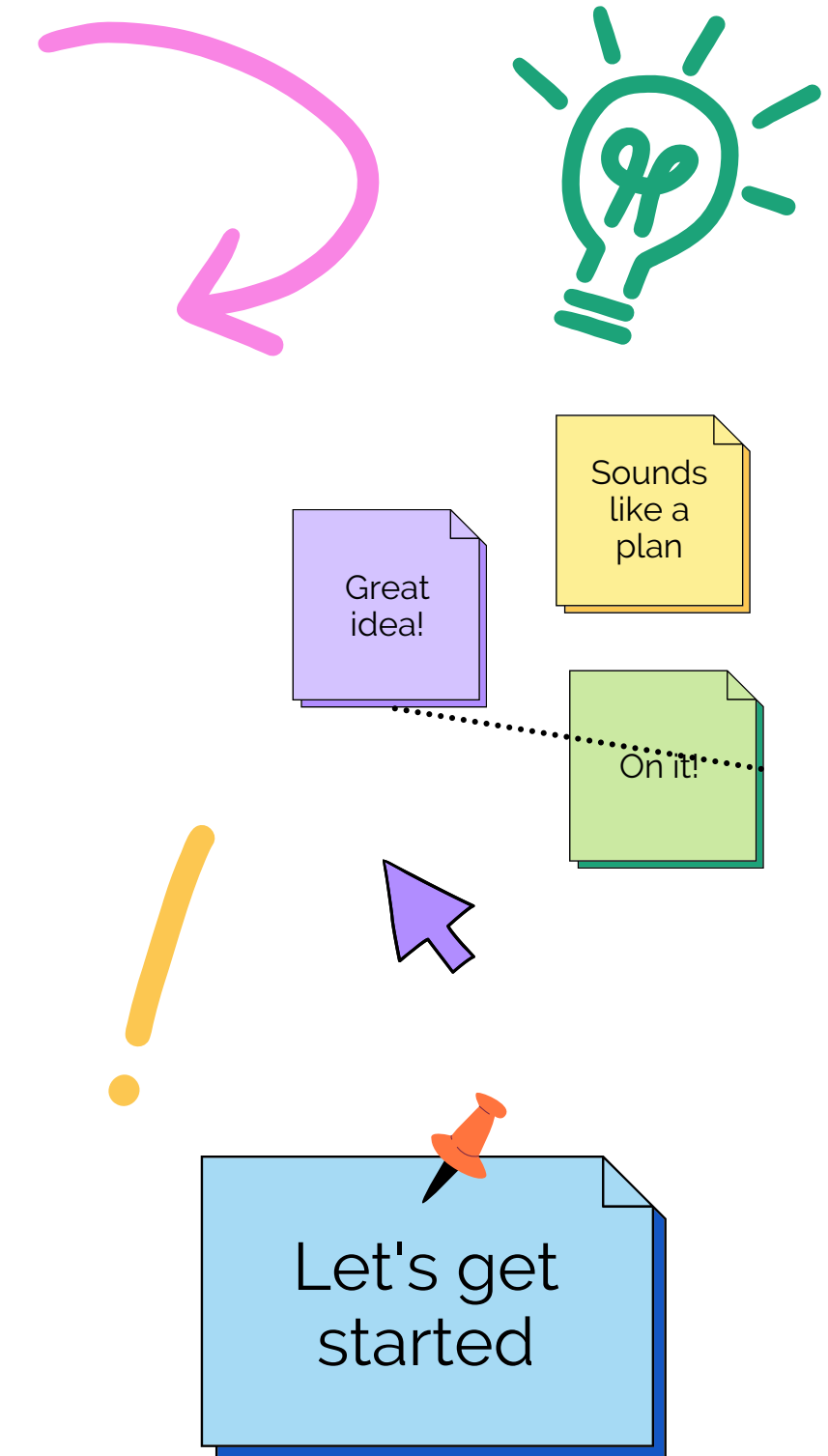
LE01. ✨

Stroke Prediction

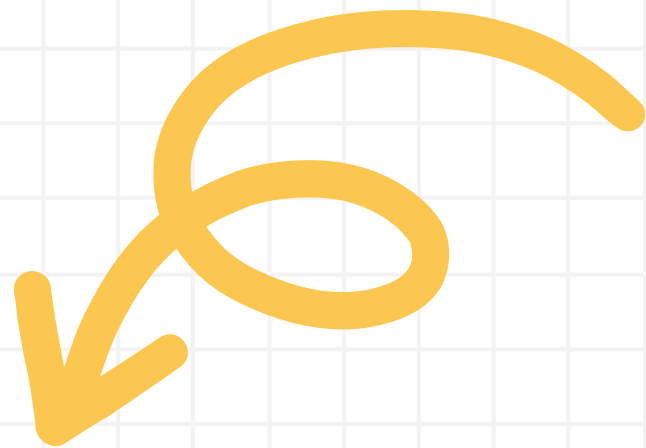
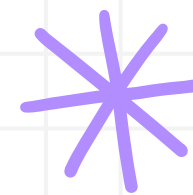
with different models of Machine Learning and Deep Learning

Members :

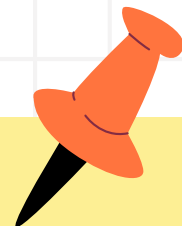
1. 2501996171 - Vincent Colin Tionando
2. 2501992620 - Dharmawan Ruslan
3. 2501993075 - Brandon Ferdinand Koris
4. 2502002791 - Ammar Yaafi Adhinegoro



Agenda



Problem
Background



Purpose
&
Benefit



Literature
Review



Methods

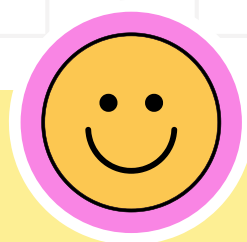


Table of Result
Experiments



Problem Background

Apa itu Stroke ?

Kondisi darurat dimana pasokan darah ke otak terganggu karena penyumbatan (stroke iskemik) atau pecahnya pembuluh darah (stroke hemoragik).

Area tertentu pada otak tidak mendapat suplai oksigen dan nutrisi sehingga terjadi kematian sel-sel otak.

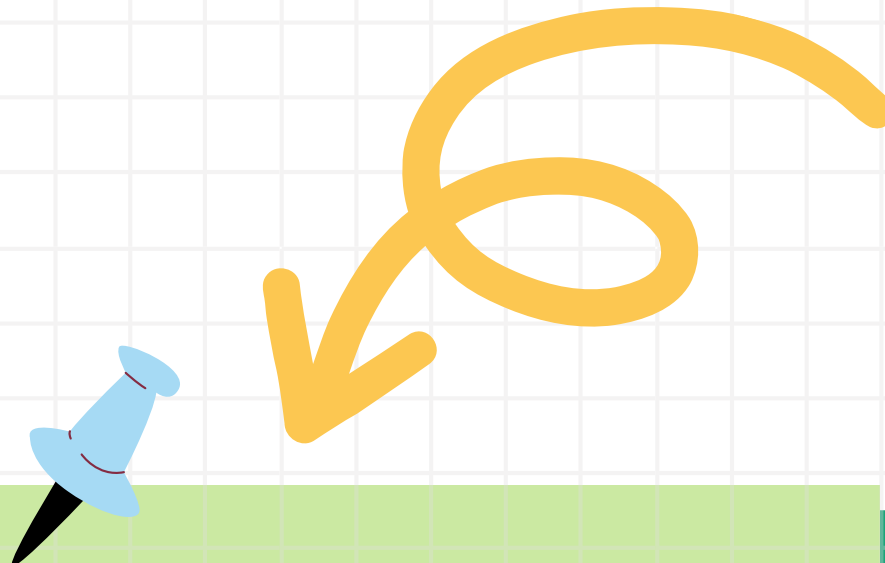
Menurut WHO...

- 1. Stroke adalah penyebab kematian nomor dua di dunia setelah penyakit jantung koroner. 11%**
- 2. Pada tahun 2019, sekitar 11,9 juta orang di seluruh dunia mengalami stroke dan 6,3 juta orang meninggal akibat stroke.**

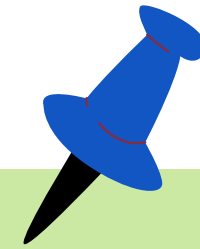
Problem Background

Akibat dari Stroke



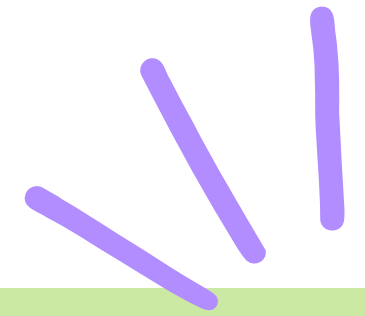
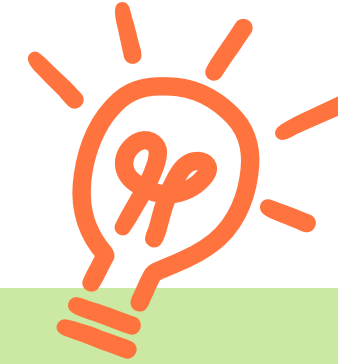
- 
- 1. Kelumpuhan atau melemahnya otot**
 - 2. Gangguan kognitif dan memori**
 - 3. Masalah psikologis**
 - 4. Masalah penglihatan**

Purpose & Benefit



**Memprediksi kemungkinan
orang terkena stroke
berdasarkan faktor-faktor
penyebabnya stroke
menggunakan
model machine learning**

Literature Review



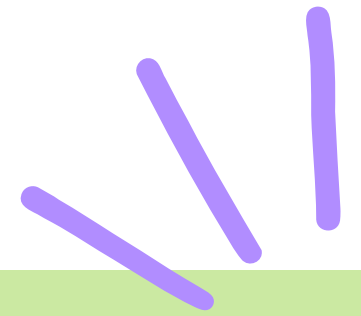
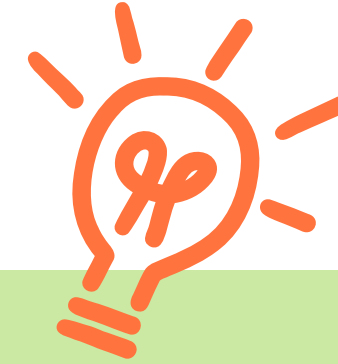
[1] Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database (2017)

Paper ini memprediksi stroke berdasarkan kondisi kesehatan pasien seperti contohnya rekor penyakit jantung, BMI, healthstyle, dll. Penelitian ini menggunakan dataset yang diambil dari 5110 individu. Penelitian ini mencoba berbagai macam model seperti Decision Tree, Random Forest, Naive Bayes, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Logistic Regression, Gradient Boosting, AdaBoost, XGBoost, dan Artificial Neural Network. Ada banyak faktor yang dilihat untuk melihat keunggulan tiap model seperti akurasi, presisi, recall, F1-score, dan area dibawah kurva. Penelitian ini menghasilkan akurasi terbaik yaitu 97% yang dihasilkan oleh model Weighted Voting.

[2] A predictive analytics approach for stroke prediction using machine learning and neural networks (2022)

Paper ini mencoba untuk meningkatkan manajemen penyakit stroke dengan memprediksinya lebih baik. Paper ini melihat faktor-faktor seperti umur, penyakit jantung, dll. Penelitian ini fokus ke dalam relasi antara faktor dan mencari faktor mana yang lebih berpengaruh. Untuk mencari relasi tersebut penelitian ini menggunakan correlation matrix dan mendapatkan bahwa umur merupakan faktor yang paling berdampak terhadap faktor lainnya. Setelah itu data tersebut dimasukkan ke dalam 3 model yaitu Decision tree, NN, dan random forest. Pertama mereka mencoba memasukkan semua faktor lalu memasukkan hanya beberapa faktor yang paling berdampak. Mereka menemukan bahwa penghilangan beberapa faktor yang tidak penting tidak begitu berdampak terhadap akurasi. Bahkan NN mendapatkan peningkatan akurasi sedikit.

Literature Review



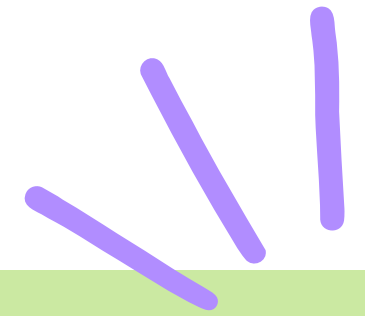
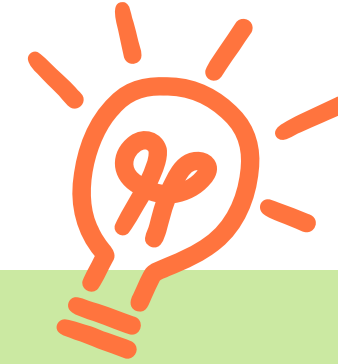
[3] Machine Learning for Brain Stroke: A Review (2020)

Sebuah studi mengulas 39 studi sebelumnya mulai dari 2007 hingga 2019 dari hasil database ilmiah web ScienceDirect tentang kontribusi machine learning dalam memecahkan masalah terkait stroke. Permasalahan terdiri dari pencegahan, identifikasi faktor risiko, pengobatan, dan prognostikasi. Hasil penelitian menunjukkan bahwa ada beberapa tabel untuk setiap masalah terkait stroke dan setiap pendekatan berbasis machine learning memiliki kelebihan dan kekurangan. Untuk diagnosis, ada diagnosis yang didapat mulai dari gambar MRI hingga fitur suara. Untuk pencegahan stroke (early detection) penelitian menunjukkan bahwa Random Forest (RF) dan SVM digunakan untuk masalah klasifikasi 4 kelas untuk menentukan deteksi dini stroke.

[4] Performance Analysis of Machine Learning Approaches in Stroke Prediction (2020)

Sebuah studi bereksperimen dengan pengklasifikasi pembelajaran mesin yang berbeda untuk prediksi stroke menggunakan himpunan data termasuk subjek yang sehat dan tidak sehat dengan faktor risiko stroke. Penelitian ini menggunakan dua teknik prediksi stroke, yaitu algoritma decision tree c4.5 dan KNN. Hasil penelitian menunjukkan bahwa algoritma decision tree c4.5 memiliki akurasi 95% sedangkan KNN memiliki akurasi 94%. Penelitian ini menggunakan 2 model ANN untuk memperkirakan prognosis stroke iskemik dan nilai presisinya adalah 79% dan 95%.

Literature Review



[5] An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure (2019)

Paper ini mengusulkan sebuah expert system berbasis stacked support vector machines (SVM) untuk meningkatkan diagnosis gagal jantung.

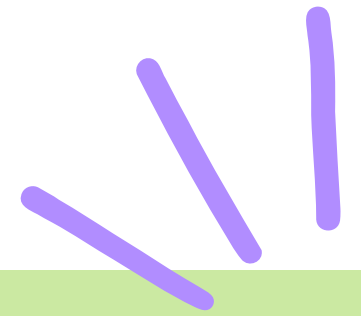
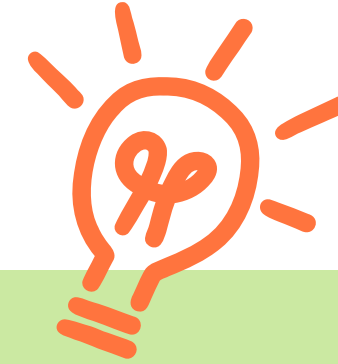
Model SVM pertama digunakan untuk menghilangkan fitur yang tidak relevan, sedangkan model kedua digunakan sebagai model prediksi. Kedua model dioptimalkan menggunakan algoritma pencarian grid hybrid.

Metode yang diusulkan berhasil mengungguli sepuluh metode yang ada dalam literatur dan model pembelajaran mesin state-of-the-art lainnya, meningkatkan kekuatan model SVM konvensional sebesar 3,3%. Metode yang diusulkan juga efisien dalam hal kompleksitas waktu, mengurangi waktu pelatihan model prediksi.

Hasil eksperimen yang dicapai pada dataset gagal jantung menunjukkan bahwa sistem pakar ini dapat meningkatkan proses pengambilan keputusan dokter selama diagnosis.

Studi perbandingan dengan metode yang telah diusulkan sebelumnya yang diterapkan pada dataset penyakit jantung Cleveland juga dilakukan dalam hal akurasi klasifikasi.

Literature Review



[6] Heart Disease prediction using MLP and LSTM models (2020)







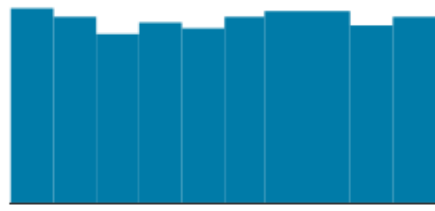
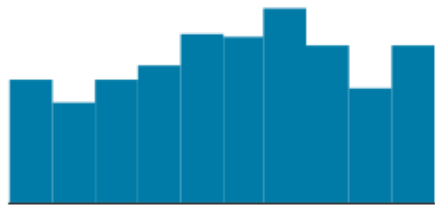



Paper ini menggunakan Heart UCI disease dataset yang memiliki data pasien dengan atribut tertentu dengan tujuan memprediksi apakah pasien terkena kelainan jantung atau tidak.







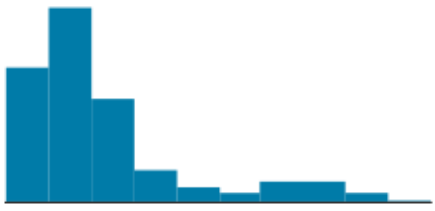

Dalam hal akurasi dan parameter lain, model yang digunakan adalah MLP dan LSTM untuk prediksi penyakit jantung. Namun LSTM secara khusus dipilih karena kemampuannya menangani data timeseries dan binary classification secara efektif.

Dan terbukti model LSTM mengungguli model MLP, dengan mencapai akurasi tinggi 96,5%. Peningkatan efisiensi model LSTM dapat memprediksi penyakit jantung dan menekankan dampak potensialnya pada desain dan implementasi sistem prediksi penyakit jantung di bidang kesehatan.

Kesimpulannya, paper ini berfokus pada analisis komparatif teknik MLP dan LSTM untuk prediksi penyakit jantung dan hasilnya menunjukkan bahwa LSTM menunjukkan kinerja yang menjanjikan dalam memprediksi secara akurat ada tidaknya penyakit jantung.

Dataset

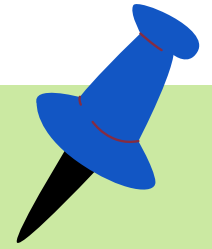
<div><div> id</div><div></div></div> <div>Unique id</div>	<div><div> gender</div><div></div></div> <div>Gender</div>	<div><div> age</div><div></div></div> <div>Age</div>	<div><div> hypertension</div><div></div></div> <div>Hypertension binary feature</div>	<div><div> heart_disease</div><div></div></div> <div>Heart disease binary feature</div>	<div><div> ever_married</div><div></div></div> <div>Has the patient ever been married?</div>
<div><div>6772.9k</div></div>	<div><div>Female59%</div><div>Male41%</div><div>Other (1)0%</div></div>	<div><div>0.0882</div></div>	<div><div>01</div></div>	<div><div>01</div></div>	<div><div>true335366%</div><div>false175734%</div></div>
9046	Male	67	0	1	Yes
51676	Female	61	0	0	Yes
31112	Male	80	0	1	Yes

<div><div> work_type</div><div></div></div> <div>Work type of the patient</div>	<div><div> Residence_type</div><div></div></div> <div>Residence type of the patient</div>	<div><div> avg_glucose_level</div><div></div></div> <div>Average glucose level in blood</div>	<div><div> bmi</div><div></div></div> <div>Body Mass Index</div>	<div><div> smoking_status</div><div></div></div> <div>Smoking status of the patient</div>	<div><div> stroke</div><div></div></div> <div>Stroke event</div>
<div>Private57%</div> <div>Self-employed16%</div> <div>Other (1366)27%</div>	<div>Urban51%</div> <div>Rural49%</div>	<div><div>55.1272</div></div>	<div>N/A4%</div> <div>28.71%</div> <div>Other (4868)95%</div>	<div>never smoked37%</div> <div>Unknown30%</div> <div>Other (1674)33%</div>	<div><div>01</div></div>
Private	Urban	228.69	36.6	formerly smoked	1
Self-employed	Rural	202.21	N/A	never smoked	1
Private	Rural	105.92	32.5	never smoked	1

Methods



Classification using machine learning algorithms



Data Preprocessing :

Menghilangkan kolom "id",
Menghilangkan baris-baris dengan kolom `bmi` bernilai "N/A",
dan kolom `smoking_status` bernilai "Unknown"
Membagi dataset menjadi dua, untuk training dan testing.

KNN :

Menerapkan Label Encoding pada data, dan melakukan training model kNN dengan nilai nearest neighbor yang berbeda-beda.

SVM:

Menerapkan **One Hot Encoding** pada data dan melakukan training model SVM dengan berbagai macam jenis kernel dan parameter.

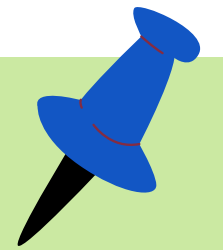
XGBOOST: :

Menerapkan **One Hot Encoding** pada data dan melakukan training model XGBoost dengan parameter `n_estimators=2`, `max_depth= 30`, `learning_rate= 1`, `objective='binary:logistic'`

Methods



Classification using deep learning algorithms



GRU:

Menerapkan Label Encoding pada data, dan melakukan training model LSTM dengan nilai input_dim = 10 dan timesteps = 1 dan dalam proses fitting menggunakan parameter untuk training yakni epochs=20, batch_size=256, validation_split=0.1, callbacks=EarlyStopping(monitor='val_loss', patience=7, min_delta=0.0001))

LSTM:

Menerapkan Label Encoding pada data, dan melakukan training model LSTM dengan nilai input_dim = 10 dan timesteps = 1 dan dalam proses fitting menggunakan parameter untuk training yakni epochs=20, batch_size=256, validation_split=0.1, callbacks=EarlyStopping(monitor='val_loss', patience=7, min_delta=0.0001))

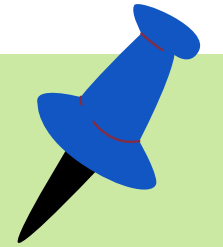
.MLP :

Menerapkan **One Hot Encoding** pada data dan melakukan training model MLP dengan parameter hidden_layer_sizes=(100,) , max_iter=1000, early_stopping=True, validation_fraction=0.15, verbose=True

Methods



Classification using deep learning algorithms



.MLP dengan Random Oversampling

Menerapkan metode random oversampling, dengan mengimplementasikan scaling pada data input, dan menambahkan hidden layer dari MLP.

Result

No	Models	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Support
1	KNN	95	Not Stroke = 95 Stroke = 50 W.Avg = 92	Not Stroke = 100 Stroke = 1 W.Avg = 95	Not Stroke = 97 Stroke = 1 W.Avg = 92	Not Stroke = 3246 Stroke = 180 Total = 3426
2	SVM	95	Not Stroke = 95 Stroke = 47 W.Avg = 90	Not stroke = 100 Stroke = 50 W.Avg = 49	Not stroke = 97 Stroke = 49 W.Avg = 92	Not Stroke = 3245 Stroke = 180 Total = 3425
3	XGBOOST	96	Not Stroke = 96 Stroke = 78 W.Avg = 95	Not Stroke = 100 Stroke = 26 W.Avg = 96	Not Stroke = 98 Stroke = 38 W.Avg = 95	Not Stroke = 3246 Stroke = 180 Total = 3426
4	LSTM	94	Not Stroke = 95 Stroke = 0 W.Avg = 95	Not Stroke = 100 Stroke = 0 W.Avg = 95	Not Stroke = 97 Stroke = 0 W.Avg = 92	Not Stroke = 3246 Stroke = 180 Total = 3426
5	GRU	95	Not Stroke = 95 Stroke = 100 W.Avg = 90	Not Stroke = 100 Stroke = 0 W.Avg = 95	Not Stroke = 97 Stroke = 0 W.Avg = 92	Not Stroke = 3246 Stroke = 180 Total = 3426
6	MLP	95	Not Stroke = 95 Stroke = 100 W.Avg = 95	Not Stroke = 100 Stroke = 0 W.Avg = 95	Not Stroke = 97 Stroke = 0 W.Avg = 92	Not Stroke = 3246 Stroke = 180 Total = 3426
7	MLP + Resampled	96	Not Stroke = 100 Stroke = 92 W.Avg = 96	Not Stroke = 92 Stroke = 100 W.Avg = 96	Not Stroke = 95 Stroke = 96 W.Avg = 96	Not Stroke = 664 Stroke = 635 Total = 1299

References



- [1] C. -Y. Hung, W. -C. Chen, P. -T. Lai, C. -H. Lin and C. -C. Lee, "Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database," 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea (South), 2017, pp. 3110-3113, doi: 10.1109/EMBC.2017.8037515.
- [2] S. Dev, H. Wang, C. S. Nwosu, N. Jain, B. Veeravalli, D. John, A predictive analytics approach for stroke prediction using machine learning and neural networks, Healthcare Analytics, vol.2, 2022, 100032, ISSN 2772-4425, <https://doi.org/10.1016/j.health.2022.100032>.
- [3] M. S. Sirsat, E. Fermé, J. Câmara, Machine Learning for Brain Stroke: A Review, Journal of Stroke and Cerebrovascular Diseases, vol.29, no.10, 2020, 105162, ISSN 1052-3057, <https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105162>.
- [4] M. U. Emon, M. S. Keya, T. I. Meghla, M. M. Rahman, M. S. A. Mamun and M. S. Kaiser, "Performance Analysis of Machine Learning Approaches in Stroke Prediction," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2020, pp. 1464-1469, doi: 10.1109/ICECA49313.2020.9297525.
- [5] L. Ali et al., "An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure," in IEEE Access, vol. 7, pp. 54007-54014, 2019, doi: 10.1109/ACCESS.2019.2909969.
- [6] M. Djerioui, Y. Brik, M. Ladjal and B. Attallah, "Heart Disease prediction using MLP and LSTM models," 2020 International Conference on Electrical Engineering (ICEE), Istanbul, Turkey, 2020, pp. 1-5, doi: 10.1109/ICEE49691.2020.9249935.

You!