



## Optimasi Feature Selection Menggunakan Algoritma Neural Network Untuk Klasifikasi Brain Stroke

Serly Agustin<sup>1</sup>, Rizkia Meinita<sup>2</sup>, Fiqri Khalid Aziz Al-rasyid<sup>3</sup>, Amelia Anjani<sup>4</sup>,  
Rehan Alif Albani<sup>5</sup>, Ricky Firmansyah<sup>6</sup>

<sup>1,2,3,4,5,6</sup> Universitas Adhirajasa Reswara Sanjaya

Korespondensi penulis: [serlyagttn@gmail.com](mailto:serlyagttn@gmail.com)<sup>1</sup>, [rizkiameinita3@gmail.com](mailto:rizkiameinita3@gmail.com)<sup>2</sup>, [amlanzni@gmail.com](mailto:amlanzni@gmail.com)<sup>3</sup>,  
[userofbloods@gmail.com](mailto:userofbloods@gmail.com)<sup>4</sup>, [rehanalif55@gmail.com](mailto:rehanalif55@gmail.com)<sup>5</sup>, [ricky@ars.ac.id](mailto:ricky@ars.ac.id)<sup>6</sup>

**Abstract.** *One of the deadliest strokes is a brain stroke. According to the results of many cases of brain stroke patients, there is a possibility that bad lifestyles such as smoking and drinking alcohol can cause high blood pressure. The goal is to classify triggers for brain structure symptoms by comparing several algorithms. From the results of this comparison, it is possible to obtain triggers with the highest number of triggers so that later brain structures can be diagnosed more quickly. In several algorithms namely nn, feature selection and GA. To group triggers for several brain stroke symptoms, to maximize feature weight and feature selection, data processing using rapidminer was continued with four algorithms: X-Fold validation and split validation with ratios of 0.5, 0.6, 0.7, 0.8 and 0.9. After this test, the most popular AUC values and methods, together with the Neural Net algorithm, the Optimize Selection (Evolutionary) feature, and using a Split Validation ratio of 0.9, produce numbers with very high accuracy. AUC of 0.549 and an accuracy value of 95.88%.*

**Keywords:** *brain stroke, evolutionary, klasifikasi, neural network, rapidminer.*

**Abstrak.** Salah satu struk yang paling mematikan adalah stroke otak. Menurut hasil dari banyak kasus pasien struk otak, ada kemungkinan gaya hidup yang buruk seperti merokok, dan minuman berakohol yang dapat menyebabkan tekanan darah tinggi. Tujuan dari penelitian ini adalah untuk mengklasifikasikan pemicu gejala struk otak dengan cara membandingkan beberapa algoritma. Dari hasil perbandingan tersebut bisa didapatkan pemicu dengan jumlah terbanyak agar nantinya struk otak bisa di diagnosa lebih cepat. Dalam beberapa algoritma yaitu nn, feature selection dan GA. Untuk mengelompokan pemicu dari beberapa gejala stroke otak, dapat memaksimalkan bobot fitur dan seleksi fitur, pengolahan data menggunakan rapidminer dilanjutkan dengan empat algoritma: validasi X-Fold dan validasi split dengan rasio 0.5, 0.6, 0.7, 0.8, dan 0.9. Setelah pengujian ini, nilai dan metode AUC yang paling populer, bersama dengan algoritme Neural Net, fitur Optimalkan Pilihan (Evolusioner), dan menggunakan rasio Validasi Terpisah sebesar 0,9, menghasilkan angka dengan akurasi yang sangat tinggi. AUC sebesar 0.549 dan nilai akurasi sebesar 95.88%.

**Kata kunci:** brain stroke, evolutionary, klasifikasi, neural network, rapidminer.

## **PENDAHULUAN**

Otak adalah jaringan rumit dari subset sel yang beragam yang dapat diprogram ulang dan dibangun kembali secara arsitektural[1]. Menurut definisi AHA/ASA yang diperbarui dari tahun 2013, stroke adalah setiap tanda objektif dari otak, sumsum tulang belakang, atau kematian sel retina yang tidak dapat diubah akibat sumber vascular[1]

Ketika aliran darah otak gagal karena pembuluh darah yang tersumbat atau pecah, mengakibatkan iskemia atau pendarahan, masing-masing, stroke otak berkembang. Stroke saat ini menjadi penyebab kematian nomor dua dan penyumbang signifikan kecacatan jangka panjang di seluruh dunia. Selain itu, peningkatan beban ini diantisipasi karena populasi yang menua di tahun-tahun mendatang, yang akan berdampak besar, serta finansial, pada sistem perawatan Kesehatan.[2]

Setiap tahun, stroke mempengaruhi sekitar 16 juta orang di seluruh dunia dan ada biaya sosial yang tinggi, menurut American Heart Association masalah kesehatan yang serius karena tingkat kematiannya yang tinggi. Juga karena stroke, perawatan di rumah sakit menjadi lebih mahal dan akibatnya kebutuhan meningkat teknologi canggih yang dapat membantu dalam diagnosis klinis, pengobatan, prediksi kejadian klinis, rekomendasi tanggal intervensi pengobatan yang menjanjikan[3].

Deteksi dini stroke merupakan langkah penting dalam pengobatan yang efektif [3]. Wanita memiliki faktor risiko stroke yang lebih tinggi. Data Menyarankan bahwa beberapa faktor risiko stroke lebih umum terlibat pada wanita. Ini dapat diamati terutama ketika: faktor risiko migrain dengan aura, mengonsumsi hormon, dan obesitas. Stroke biasanya terjadi pada wanita yang lebih tua (pascamenopause). Ini ada hubungannya dengan efek perlindungan estrogen. ke dalam pembuluh darah. Frekuensi Stroke pada Orang Tua Ini cenderung memperburuk hasil klinis. Kelompok wanita memiliki insiden stroke tertinggi dibandingkan dengan kelompok umur lainnya berdasarkan jenis kelamin. Fakta-fakta ini menunjukkan bahwa ada perbedaan yang mendalam Frekuensi stroke pada wanita dan pria. Pengidap stroke Wanita yang lebih tua lebih besar dari pria [4]. Hasil klinis berbeda ketika wanita menderita stroke. Meskipun penelitian kausal mengidentifikasi apa faktor penyebab yang memengaruhi efek kesehatan, pembelajaran mesin misalnya yang lain menggunakan faktor-faktor ini untuk menentukan pasien mana yang mengalami hal ini [5]. Proses pemilihan fitur yang sesuai yang akan menghasilkan fitur terbaik disebut Feature Selection. Optimalisasi pemilihan fitur (memilih fitur terbaik) adalah salah satu teknik yang digunakan dalam pembelajaran mesin untuk meningkatkan kinerja model dan mengurangi kerumitannya. Meskipun pengoptimalan pemilihan fitur bukan kategori paling umum dari algoritma pembelajaran mesin, ini dapat

membantu meningkatkan akurasi dan efisiensi model. Opsi masa depan memiliki keunggulan kumpulan data yang lebih kecil dan perhitungan yang lebih sedikit. Feature Selection bertujuan untuk memilih fitur terbaik dari data fitur untuk mencapai nilai akurasi tertinggi.[6]

Dalam Machine Learning, terdapat 5 jenis algoritma yaitu; Classification, Clustering, Regression, Estimation serta Prediction. Fungsi klasifikasi dalam machine learning adalah untuk mengkategorikan data ke dalam kelas atau label yang sudah ditentukan sebelumnya. Algoritma klasifikasi menggunakan pola dan karakteristik data pelatihan untuk membangun model yang dapat digunakan untuk memprediksi kelas dari data baru yang tidak diketahui. Algoritma ini pun bisa dimanfaatkan untuk mengklasifikasikan gejala-gejala awal penyakit agar lebih mudah untuk mendiagnosa.[7]

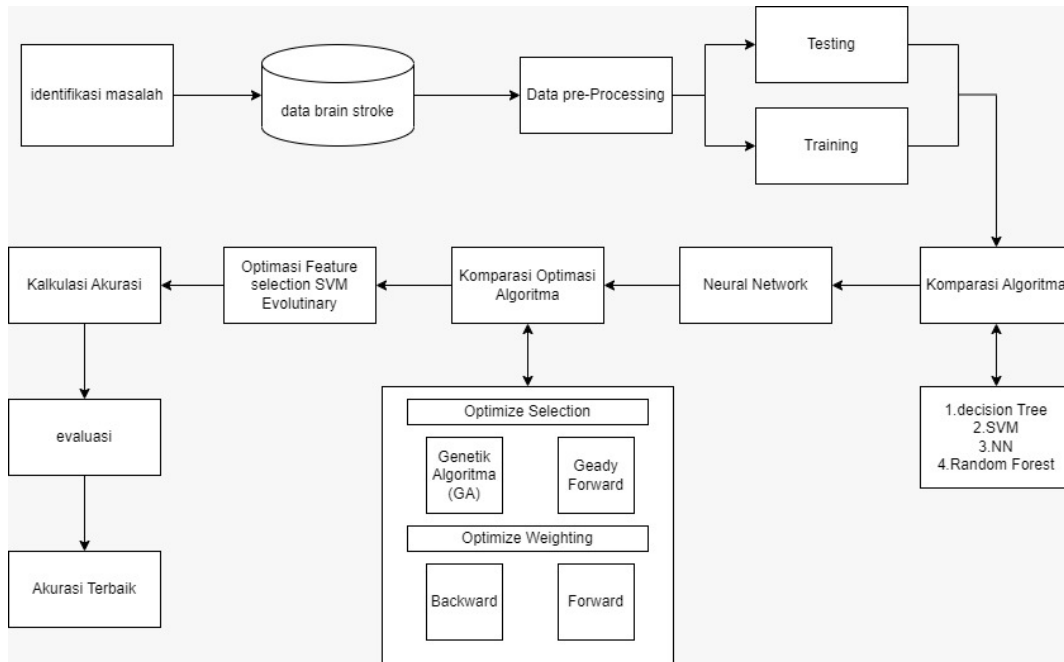
Algoritma genetika (GA) adalah algoritma komputer yang terinspirasi oleh teori evolusi, yang kemudian diterapkan sebagai algoritma komputer untuk menemukan solusi suatu masalah dengan cara yang lebih "alami". Berbeda dengan pendekatan pengoptimalan tradisional, GA menggunakan populasi solusi yang didorong menuju solusi ideal dengan meniru prinsip evolusi alam. Algoritma ini dapat digunakan dalam kasus klasifikasi untuk memilih subset fitur yang optimal untuk meningkatkan kinerja model klasifikasi. Dengan memilih subset fitur yang paling relevan dan informatif, algoritma genetika dapat membantu mengurangi dimensi data, mengatasi masalah overfitting, mempercepat waktu pelatihan, dan meningkatkan akurasi atau performa klasifikasi secara keseluruhan.[7]

Peneliti membandingkan akurasi lima algoritma dalam hal pemilihan bobot berdasarkan penelitian terkait ini. lima Decision Tree, Naïve Bayes, Neural Netwotrck, dan Random Forest dapat digunakan untuk meramalkan masalah bagi pasien penderita struk otak. Algoritma ini memiliki keunggulan dalam situasi tertentu[7].

## **METODE PENELITIAN**

Untuk mengatasi masalah klasifikasi dalam penelitian ini, sejumlah metode digunakan, termasuk pelatihan pemrosesan data dan pengujian data menggunakan dua metodologi (cross validation dan split validation). Perbandingan kategorisasi algoritma juga dilakukan untuk memilih algoritma terbaik. Klasifikasi kemudian akan diperkuat dengan meningkatkan karakteristik dan bobot dataset. Tahapan penelitian ini diakhiri dengan penilaian untuk memilih algoritma klasifikasi dan mengoptimalkan pendekatan yang dapat meningkatkan nilai klasifikasi.

**Gambar 1. Desain Penelitian**



### Identifikasi Masalah

Penyakit struk otak terjadi ketika jaringan otak tidak berfungsi dengan baik dan aliran darah serta oksigen ke dalamnya kurang. Hingga saat ini, diketahui setidaknya ada 9 hal yang dapat menyebabkan dan meningkatkan risiko stroke, yakni : Tekanan darah tinggi, kebiasaan merokok, genetik, mengidap penyakit jantung, kolesterol tinggi yang tak terkontrol, obesitas, penyakit diabetes, usia serta gender. Dari pemicu tersebut akan diklasifikasikan lalu datanya dibandingkan dengan algoritma Neural Network beserta algoritma GA agar kemudian bisa diprediksikan pasien mana yang memiliki penyakit struk otak berdasarkan hasil akurasi yang didapatkan.

### Data Set

Untuk melakukan suatu penelitian tentunya dataset merupakan bahan utama yang akan diolah menggunakan suatu algoritma. Data set yang digunakan adalah brain stroke data set yang memiliki 4981 record dan 11 atribut dengan alamat website <https://www.kaggle.com/datasets/jillanisofttech/brain-stroke-dataset> dan di publish pada tahun 2022 oleh Kaggle.

**Tabel 1. Data Brain Stroke**

	A	B	C	D	E	F	G	H	I	J	K
1	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
2	Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
3	Male	80	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
4	Female	49	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
5	Female	79	1	0	Yes	Self-employed	Rural	174.12	24	never smoked	1
6	Male	81	0	0	Yes	Private	Urban	186.21	29	formerly smoked	1
7	Male	74	1	1	Yes	Private	Rural	70.09	27.4	never smoked	1
8	Female	69	0	0	No	Private	Urban	94.39	22.8	never smoked	1
9	Female	78	0	0	Yes	Private	Urban	58.57	24.2	Unknown	1
10	Female	81	1	0	Yes	Private	Urban	80.43	29.7	never smoked	1
11	Female	61	0	1	Yes	Govt_job	Rural	120.46	36.8	smokes	1
12	Female	54	0	0	Yes	Private	Urban	104.51	27.3	smokes	1
13	Female	79	0	1	Yes	Private	Urban	214.09	28.2	never smoked	1
14	Female	50	1	0	Yes	Self-employed	Rural	167.41	30.9	never smoked	1
15	Male	64	0	1	Yes	Private	Urban	191.61	37.5	smokes	1
16	Male	75	1	0	Yes	Private	Urban	221.29	25.8	smokes	1
17	Female	60	0	0	No	Private	Urban	89.22	37.8	never smoked	1
18	Female	71	0	0	Yes	Govt_job	Rural	193.94	22.4	smokes	1
19	Female	52	1	0	Yes	Self-employed	Urban	233.29	48.9	never smoked	1
20	Female	79	0	0	Yes	Self-employed	Urban	228.7	26.6	never smoked	1
21	Male	82	0	1	Yes	Private	Rural	208.3	32.5	Unknown	1
22	Male	71	0	0	Yes	Private	Urban	102.87	27.2	formerly smoked	1
23	Male	80	0	0	Yes	Self-employed	Rural	104.12	23.5	never smoked	1
24	Female	65	0	0	Yes	Private	Rural	100.98	28.2	formerly smoked	1
25	Male	69	0	1	Yes	Self-employed	Urban	195.23	28.3	smokes	1
26	Male	57	1	0	Yes	Private	Urban	212.08	44.2	smokes	1

### Preprosesing (Split Dan Cross)

Pada tahap ini hasil penelitian akan di pecah menjadi data training dan data testing memanfaatkan split validation cross validation. Split validation digunakan secara acak menjadi dua kumpulan data yang berbeda, data latihan dan data uji, sedangkan cross validation digunakan untuk mengidentifikasi model uji yang akan berkinerja terbaik secara keseluruhan.

### Komparasi Algoritma

Perbandingan algoritma yang digunakan dalam penelitian yaitu untuk menentukan metode yang terbaik, dalam penelitian ini digunakan 4 algoritma diantaranya Decision Tree, Naïve Bayes, Neural Network, dan Random Forest. Dengan perbandingan nilai yang akurasi dihasilkan oleh masing-masing dari 4 Algoritma tersebut, yang salah satunya akan menentukan algoritma yang terbaik dalam mengidentifikasi brain stroke data set.

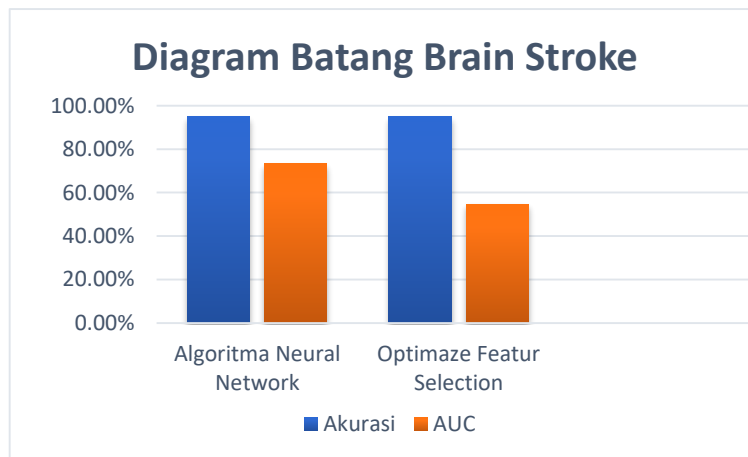
Accuracy	AUC	Algoritma
94.56%	0.638	Decision tree
94.12%	0.553	Svm
93.39%	0.733	Nn
95.02%	0.186	Random forest

## Neural Network

Neural Network digunakan untuk pengelompokan, klasifikasi, prediksi, penambangan pola dan pengenalan pola, jaringan saraf digunakan[8]. Neural Network berfungsi sebagai metafora bagi otak dalam hal pemrosesan informasi. Alih-alih menjadi reproduksi yang tepat tentang cara kerja otak, model ini diilhami secara ilmiah. Sistem yang menggunakan Neural Network telah terbukti sangat menjanjikan.[9]

## Komparasi Optimasi

Penelitian menggunakan algoritma Neural Network split validation dan cross validation yang diperoleh dari Algoritma Neural Network berdasarkan optimasi feature selection Genetic Algoritma.



## GA

Berusaha untuk mengidentifikasi individu yang paling bagus dengan menghasilkan generasi secara iteratif, mengabstraksi ruang masalah sebagai populasi individu. Digunakan sebagai teknik pengoptimalan untuk mengatasi masalah yang menantang di berbagai industry[10]

## Evaluasi

Pada tahapannya akan diketahui akurasi nilai dan AUC terbaik dari klasifikasi data set dengan memprediksi apakah pasien mengalami stroke membandingkan hasil akurasi dan auc yang diperoleh dari algoritma Neural Network split validation dan cross validation yang diperoleh dari Algoritma Neural Network berdasarkan optimasi feature selection Genetic Algoritma.

## HASIL DAN PEMBAHASAN

### Perbandingan Algoritma

Hasil implementasi dari 4 algoritma diantaranya Decision Tree, Nueal Network, Naïve Bayes dan Random forest dengan menggunakan Cross X-Fold Validation pada penentuan klasifikasi dengan mencari hasil nilai accuracy, dalam penelitian ini menggunakan dataset sebanyak 4981 data.

**Tabel 3. Hasil 4 Algoritma**

Algoritma	Accuracy	AUC
Decision Tree	94,56%	0,638
SVM	94,12%	0,553
Neural Net	93,39%	0,733
Random Forest	95,02%	0,186

Sesuai data akurasi dari 4 algoritma yang sudah di dapat pada tabel 3. Kemudian, algoritma yang memiliki tingkat akurasi yang paling tinggi pada algoritma Neural Net dengan Nilai Accuracy 93,39%, maka algoritma tersebut terpilih ke tahapan selanjutnya.

### Hasil dari feature optimize selection dan feature weight optimize

**Tabel 4**

Algoritma	Fitur	Validation	Accuracy	AUC
NN	Optimize selection	x-vols Validation	95.18%	0.547
NN	Optimize selection Evelutianary	x-vols Validation	95.06%	0.452
NN	Optimize Weights Evolutianary	x-vols Validation	95.11%	0.579
NN	Optimize Weights PSO	x-vols Validation	95.10%	0.529

Hasil implementasi dari Optimize Feature Selection dan Optimize Feature Weight yaitu Optimize Selection dan Optimize Selection (Evolutionary) serta Optimize Weight PSO dan Evolutionary dengan menggunakan algoritma Neural Network pada Tools Rapid Miner dengan menggunakan Cross X-Fold Validation pada penentuan klasifikasi dengan mencari hasil nilai accuracy.

**Tabel 5**

Algoritma	Fitur	Validation	Accuracy	AUC
NN	Optimize selection	x-vols Validation	95.18%	0.547

Pada tabel 5 Hasil Optimize yang tertinggi yaitu Optimize Selection dengan accuracy 95.18% dan AUC 0,547 serta terdapat 11 Atribut Weights yaitu. 1) Gender: Select "Male," "Female," or "Other.",Secondly, the patient's age,Hypertension: 0 if there is no hypertension and 1 if there is hypertension,Heart condition: 0 if the patient has no heart conditions, 1 if the patient has a heart condition,Ever married? Indicate "No" or "Yes",Work Type: "Children," "Govtjov," "Never Worked," "Private," or "Self-Employed" 7) "Rural" or "Urban" as the residence type,avgglucoselevel: blood glucose levels on average,Body mass index, or bmi,smoking status: "previously smoked," "never smoked," "smokes," or "Unknown"\* 11) stroke: 1 if the patient had a stroke, 0 otherwise.

**Tabel 6. Hasil Akhir**

Algoritma	Fitur	Validation	Accuracy	AUC
NN	Optimize feature selection	Split Validation 0,5	95,06%	0,529
NN	Optimize feature selection	Split Validation 0,6	95,06%	0,536
NN	Optimize feature selection	Split Validation 0,7	95,11%	0,511
NN	Optimize feature selection	Split Validation 0,8	95,09%	0,544
NN	Optimize feature selection	Split Validation 0,9	95,18	0,351

Pada tabel Hasil Akhir dari keseluruhan yang terkecil yaitu menggunakan Split Validation 0.5 dengan nilai Accuracy 95.06% dan AUC 0,529, sedangkan yang terbaik dan tertinggi yakni dengan menggunakan Split Validation dengan Ratio 0.9 dengan nilai Accuracy 95.18% dengan AUC 0,351.

## KESIMPULAN

Dataset Brain Stroke yang tersedia dari Kaggle,dalam penelitian menggunakan tool Rapidminer 9.0 yang memiliki 4981 record dan 11 atribut, yang diuji 4 algoritma Decision Tree,Naïve Bayes,Neural Network,dan Random Forest, menggunakan Optimize Feature Selection (Optimize Selection dan Optimize Selection dan Optimize Feature Weight dalam metode data mining klasifikasi, yang memiliki nilai akurasi yang tertinggi yaitu dengan menggunakan algoritma Neural Network yang menghasilkan nilai Accuracy = 95.18% dan



AUC = 0.351 yang menunjukkan hasil klasifikasi terbaik, sehingga menghasilkan prediksi pasien yang mengalami stroke otak.

#### DAFTAR PUSTAKA

- J. Dąbrowski et al., “Brain Functional Reserve in the Context of Neuroplasticity after Stroke,” *Neural Plasticity*, vol. 2019. Hindawi Limited, 2019. doi: 10.1155/2019/9708905.
- R. Scapaticci, O. M. Bucci, I. Catapano, and L. Crocco, “Differential microwave imaging for brain stroke followup,” *Int J Antennas Propag*, vol. 2014, 2014, doi: 10.1155/2014/312528.
- M. S. Sirsat, E. Fermé, and J. Câmara, “Machine Learning for Brain Stroke: A Review,” *Journal of Stroke and Cerebrovascular Diseases*, vol. 29, no. 10, p. 105162, Oct. 2020, doi: 10.1016/J.JSTROKECEREBROVASDIS.2020.105162.
- F. Kedokteran, U. / Rs, and B. Yogyakarta, Rizaldy Taslim Pinzon.
- P. Doupe, J. Faghmous, and S. Basu, “Machine Learning for Health Services Researchers,” *Value in Health*, vol. 22, no. 7, pp. 808–815, Jul. 2019, doi: 10.1016/J.JVAL.2019.02.012.
- “View of Optimasi Feature Selection Pada Komentar Media Sosial Terhadap Peralihan Tv Digital Menggunakan Naïve Bayes, Support Vector Machine dan K-Nearest Neighbor.” <https://journal.umkendari.ac.id/index.php/decode/article/view/121/60> (accessed May 13, 2023).
- “Komunitas eLearning IlmuKomputer.Com,” 2003. [Online]. Available: <http://dennyhermawanto.webhop.org>
- M. B. Kursu and W. R. Rudnicki, “Feature Selection with the Boruta Package,” 2010. [Online]. Available: <http://www.jstatsoft.org/>
- “World Academy of Science, Engineering and Technology,” S. Parasuraman, 2008.
- P. Gaur, “Neural Networks in Data Mining.” [Online]. Available: [www.ijecse.org](http://www.ijecse.org)
- L. Haldurai, T. Madhubala, and R. Rajalakshmi, “A Study on Genetic Algorithm and its Applications,” 2016, [Online]. Available: [www.ijcseonline.org](http://www.ijcseonline.org)