

IMPLEMENTASI SMOTE DAN *EXTREME LEARNING MACHINES* PADA KLASIFIKASI DATASET *MICROARRAY*

Ivan Sitohang¹, Triando Hamonangan Saragih², Dwi Kartini³, Radityo Adi Nugroho⁴, Mohammad Reza Faisal⁵

Ilmu Komputer, Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat
Correspondent author. Email: ¹1811016110022@mhs.ulm.ac.id, ²triando.saragih@ulm.ac.id,
³dwikartini@ulm.ac.id, ⁴radityo.adi@ulm.ac.id, ⁵reza.faisal@ulm.ac.id

Abstrak

Tumor otak merupakan salah satu penyakit yang menjadi penyebab kematian terbesar secara global. Deteksi terhadap tumor memerlukan suatu algoritma komputasi, salah satunya adalah algoritma klasifikasi. Algoritma yang digunakan untuk masalah klasifikasi adalah *extreme learning machine* (ELM). Penelitian ini menggunakan dataset dari National Center for Biotechnology Information (NCBI) berupa data *microarray* penyakit Tumor pada Otak. Jumlah data pada penelitian ini terdapat 130 data yang dimana 117 adalah data tumor dan 13 data normal, terdapat 21051 fitur yang berada di dataset ini. Hasil penelitian ini adalah untuk mengetahui keakuratan dalam pengklasifikasian tumor otak dengan menggunakan metode *Extreme Learning Machines* dengan dan tanpa menggunakan oversampling SMOTE pada keseluruhan data dengan melakukan percobaan menggunakan 100, 200, 300, 400, dan 500 neuron. Pada penelitian ini didapatkan bahwa penggunaan hidden neuron serta penyeimbangan data pada klasifikasi data *microarray* sangat berpengaruh dalam akurasi yang akan didapatkan dalam penelitian ini. Performa kinerja klasifikasi tertinggi setiap model antara lain model *Extreme Learning Machines* mendapatkan akurasi sebesar 97.43% pada *hidden neuron* = 500. Lalu *Extreme Learning Machines* menggunakan oversampling SMOTE pada keseluruhan data menghasilkan akurasi sebesar 92.30% pada *hidden neuron* = 200. Pada penelitian ini didapatkan bahwa penggunaan *hidden neuron* serta penyeimbangan data pada klasifikasi data *microarray* sangat berpengaruh dalam akurasi yang akan didapatkan dalam penelitian ini.

Kata kunci : tumor otak, *microarray*, *extreme learning machines*, SMOTE.

1. Pendahuluan

Salah satu penyakit paling mematikan dan menjadi salah satu penyebab kematian terbesar secara global adalah tumor. Data epidemiologi berdasarkan sebuah penelitian *systematic review* menunjukkan insidensi tumor otak primer di dunia adalah 10,82 per 100.000 penduduk per tahun (M and Azizah, 2022). Resiko kematian akibat tumor dapat diredam dengan menggunakan strategi deteksi dini terhadap tumor, sehingga dapat dilakukan penanganan lebih awal untuk penyembuhan tumor. Salah satu teknologi yang dapat digunakan untuk deteksi tumor adalah teknologi *microarray*.

Microarray adalah teknologi yang mampu menyimpan ribuan ekspresi gen yang diambil dari beberapa jaringan manusia sekaligus (Rusamsi et al., 2018). Menurut Adiwijaya (2018), *microarray* adalah metode efisien untuk mengatasi kanker (Nurviarelda and Adiwijaya 2018) ekspresi genetik dengan jumlah ribuan dari berbagai gen dapat diamati perubahannya oleh teknologi *DNA microarray* secara bersamaan. Hal ini membuat *DNA microarray* banyak digunakan dalam melakukan deteksi kanker ataupun tumor. Menurut Ranjani Rani & Ramyachitra (2018), data

microarray memiliki permasalahan pada jumlah sampel dan gen yang terkait yang direpresentasikan dalam bentuk matriks. Matriks ini memiliki dimensi data yang tinggi sehingga menyebabkan proses klasifikasi gen tumor yang tidak baik. Adiwijaya (2018), berpendapat Klasifikasi ekspresi gen ataupun klasifikasi kanker adalah proses untuk mengidentifikasi gen informatif yang dapat digunakan untuk memprediksi kelas untuk pengujian sampel baru. Jadi dengan adanya klasifikasi kita mengetahui pasien tersebut terkena tumor atau tidak, dengan cara pengambilan *DNA microarray* kepada si penderita.

Deteksi terhadap tumor memerlukan suatu algoritma komputasi, salah satunya adalah algoritma klasifikasi. Algoritma yang digunakan untuk masalah klasifikasi adalah *extreme learning machine* (ELM). Penelitian yang dilakukan oleh Wahid et al. (2020) mengatakan bahwa *Extreme Learning Machine* memiliki kelebihan yang dimana metode ini mirip dengan metode jaringan syaraf tiruan yang proses pelatihannya tidak memerlukan propagasi balik sehingga proses pelatihan dapat lebih cepat karena tidak perlu memperbaiki nilai bobot dan bias tiap epoch nya. Proses pelatihan dengan menggunakan metode propagasi balik dapat memakan waktu karena sifatnya yang iteratif. Oleh

karenanya diperlukan metode pelatihan yang sifatnya tidak iteratif (*single feed forward pass*), cepat, dan juga menghasilkan akurasi yang tidak kalah baik dengan propagasi balik. Penelitian yang dilakukan oleh Wahid et al. (2020), menggunakan algoritma ELM untuk klasifikasi tumor otak pada citra atau gambar. ELM mampu menghasilkan rata-rata akurasi sebesar 80%. Namun, terdapat permasalahan terhadap jumlah kelas yang tidak seimbang sehingga akurasi yang dihasilkan sangat tergantung pada keseimbangan kelas pada data *microarray*, maka dari itu diperlukan suatu metode untuk menyeimbangkan kelas untuk mendapatkan akurasi yang lebih akurat.

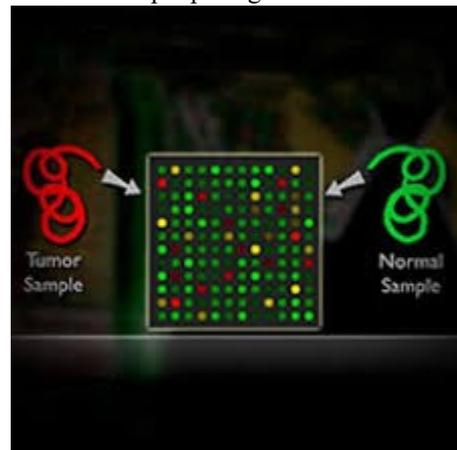
Metode SMOTE merupakan salah satu solusi yang dapat digunakan untuk mengatasi data yang tidak seimbang pada data *microarray*. Gumelar et al. (2021), berpendapat SMOTE merupakan teknik Oversampling menyeimbangkan jumlah distribusi dataset pada kelas minoritas dengan cara mensintesis dataset minoritas hingga jumlah dataset menjadi seimbang dengan jumlah dataset pada kelas mayoritas. Gumelar et al. (2021) melakukan penelitian penyeimbangan data yang tidak seimbang pada data *microarray* dengan menggunakan metode SMOTE untuk menyeimbangkan data tersebut dan membandingkan dengan 4 metode setelah datasetnya sudah di seimbangkan. Penelitian ini berhasil mendapatkan peningkatan *accuracy*, *precision*, *recall* dari 4 metode yang berbeda. Siringoringo (2018) melakukan penelitian klasifikasi membandingkan dataset *Credit Card Fraud* yang tidak menggunakan metode SMOTE dan yang menggunakan metode SMOTE. Dalam penelitian ini berhasil mendapatkan peningkatan *G-Mean* dan *F-Measure* yang signifikan, dengan adanya keseimbangan data. Maka, didapat hasil performa *G-Mean* dari penelitian ini sebesar 81,00% sedangkan yang tidak melakukan penyeimbangan data adalah 53,45% dan untuk performa *F-Measure* didapat 81,83% dengan melakukan penyeimbangan data sedangkan yang tidak melakukan penyeimbangan data adalah 38,71%. Pada penelitian ini akan dilakukan klasifikasi menggunakan metode *Extreme Learning Machine* dan akan dilakukan penyeimbangan data menggunakan SMOTE yang bertujuan untuk mendeteksi tumor dini dengan menggunakan dataset *microarray* yang bertujuan menangani dengan cepat penderita tumor otak dan mengurangi kematian yang diakibatkan oleh tumor otak.

2. Tinjauan Pustaka

Pada bagian ini akan dibahas mengenai teori-teori yang mengenai konsep *microarray*, *extreme learning machine*, dan SMOTE

2.1 Microarray

DNA *microarray* adalah teknologi yang menyimpan urutan asam nukleat di lokasi tertentu yang dapat dilihat dan digunakan untuk menganalisis ribuan sampel sekaligus. Analisis dilakukan menggunakan DNA sampel yang dire-kombinasi menggunakan probe pada chip *microarray* dan telah dilabel menggunakan zat fluorescent (Adiwijaya, 2018). Data *microarray* disimpan dalam bentuk matriks berukuran $m \times n$ dimana m merupakan jumlah sampel data dan n adalah jumlah gen di data seperti contoh sampel pada gambar 1.

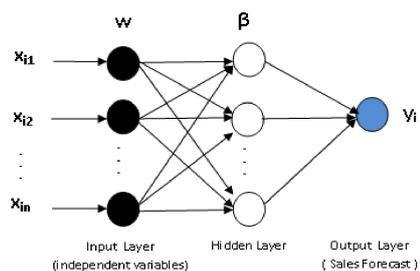


Gambar 1. Sample data *microarray*

Penggunaan *microarray* pada gambar 1 dapat dilihat bahwa pada warna hijau yang berarti sel DNA yang diambil menunjukkan normal sampel dan berwarna merah yang berarti tumor sampel sedangkan berwarna kuning adalah reaksi antara tumor dan normal, ada juga yang tidak berwarna yang artinya tidak bereaksi.

2.2 Extreme Learning Machine

ELM merupakan metode jaringan saraf tiruan yang proses pelatihannya tanpa diperlukannya propagasi balik sehingga untuk proses pelatihan bisa lebih cepat karena tidak perlukannya proses perbaruan nilai bobot dan bias pada tiap epoch nya (Wahid et al., 2020). Ada 3 tahapan dalam melakukan ELM data *training*, jumlah neuron tersembunyi, dan fungsi aktivasi (Fitriani, Ratnawati and Anam, 2019) Arsitek ELM dapat dilihat pada gambar 2.



Gambar 2. Arsitektur *Extreme Learning Machines*

2.3 SMOTE

Synthetic Minority Oversampling Technique (SMOTE) salah satu turunan pada oversampling yang diperkenalkan pertamakali oleh Nithees V. Chawla bekerja dengan mencari k nearest neighbors (yaitu ketetanggaan terdekat data sebanyak k) untuk setiap data di kelas minoritas, selanjutnya dibuat data sintetis sebanyak persentase duplikasi diinginkan pada data minor dan k-nearest neighbors yang dipilih secara *random*. Pendekatan ini bekerja dengan membuat replikasi pada data minoritas. Replikasi tersebut dikenal dengan data sintetis (synthetic data) (Siringoringo, 2018).

Secara umum fungsi algoritma SMOTE ditulis dengan SMOTE (X, N, k), rumus jarak euclidean distance dapat dihitung dengan persamaan

$$dis = \sqrt{(X1 - Y1)^2 + (X2 - Y2)^2 + \dots + (Xn - Yn)^2} \quad (1)$$

Jika atribut dengan 2 data, maka jarak Euclidean distance seperti dalam persamaan :

$$dist = X1 - Y1^2 + X2 - Y2^2 \quad (2)$$

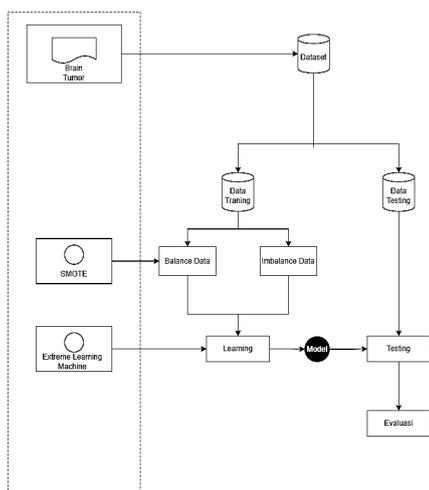
Setelah itu membuat data replikasi dari instance terdekat seperti dalam persamaan :

$$X_{syn} = Xi + X_{knn} - X_{ix} \sigma \quad (3)$$

Dimana X_{syn} adalah data sintesis hasil replikasi, X_i adalah data ke-I pada kelas minoritas, X_{knn} adalah data kelas minoritas yang memiliki jarak terdekat dengan data X_i , sedangkan σ merupakan bilangan random antara 0 dan 1.

3. Metode Penelitian

Alur penelitian dilakukan dengan langkah-langkah yang dirancang, agar penelitian dapat berjalan dengan baik. Alur penelitian ini ditunjukkan pada Gambar 3.



Gambar 3. Prosedur Penelitian

3.1 Pengumpulan Data

Data yang dikumpulkan dan akan digunakan dalam penelitian ini bersumber dari National Center for Biotechnology Information (NCBI) berupa data penyakit Tumor pada Otak. Dataset tersebut adalah dataset microarray yang dimana memiliki ribuan fitur dan memiliki 2 kelas antara kelas tumor dan kelas normal serta memiliki jumlah kelas yang tidak seimbang.

3.2 Pemodelan

Data akan dibagi menjadi dua bagian yaitu data training dan data testing dimana data training 70% dan data testing 30%. Selanjutnya akan dilakukan penyeimbang data menggunakan SMOTE. Lalu data yang sudah seimbang dan yang belum seimbang akan diklasifikasikan menggunakan Extreme Learning Machine (ELM).

3.3 Evaluasi

Evaluasi yang dilakukan pada pengujian ini menggunakan Confusion Matrik yang menghasilkan nilai berupa accuracy, precision, recall dan f1-score.

4. Hasil dan Pembahasan

4.1. Hasil Pengumpulan Data

Bahan penelitian yang digunakan dalam penelitian ini adalah dataset dari National Center for Biotechnology Information (NCBI) berupa data microarray penyakit Tumor pada Otak. Jumlah data pada penelitian ini terdapat 130 data yang dimana 117 adalah data tumor dan 13 data normal, terdapat 21051 fitur yang berada di dataset ini. dimana secara general fitur yang terdapat pada data microarray seperti 1007_s_at, 121_at, dan seterusnya adalah RNA yang sudah di ekstrak menjadi array yang terdapat di sampel tumor pada otak. Berikut ini beberapa dataset yang digunakan yang dapat dilihat pada tabel 1.

Tabel 1. Contoh dataset *Microarray Brain Tumor*

indeks	Kelas	1007_s_at	121_at	1255_g_at	1405_i_at
0	Tumor	12,60860644	4,5671539750	1,9936929430	4,8206939960
1	Tumor	12,70316963	4,6128421050	2,0301462110	4,4662597090
2	Tumor	13,06864086	4,5834609470	1,9893693770	8,8433642310
3	Tumor	12,52516400	4,5151057790	1,9142219020	3,7033137420

4	Tu mor	12,91475 021	4,546872 1440	1,939037 2760	4,046768 7250
5	Tu mor	12,57137 220	4,609247 4260	2,045263 9300	4,567273 6040
6	Tu mor	13,80163 1110	4,623686 6690	2,061709 5280	6,500260 1770
7	Tu mor	12,79935 7310	4,631425 6410	2,011058 8330	4,730917 3370
8	Tu mor	12,61242 7500	4,639853 3450	2,018065 3300	6,181886 2870
9	Tu mor	12,06135 4150	4,603329 3770	2,019011 8060	6,097243 9030
10	Tu mor	13,05818 4230	4,529650 1150	1,908788 3700	5,094843 4710
...
12 9	Tu mor	12,57289 5080	4,518853 520	1,913289 231	5,356086 746

4.2. Klasifikasi Extreme Learning Machines tanpa SMOTE

Pada penelitian ini dilakukan klasifikasi menggunakan algoritma Extreme Learning Machines (ELM) tanpa tambahan teknik apapun dengan menggunakan 100, 200, 300, 400, dan 500 neuron tersembunyi dalam metode Extreme Learning Machines untuk mengklasifikasi tumor otak tanpa menggunakan SMOTE untuk menanganikan kelas yang tidak seimbang pada dataset *microarray* pada tumor otak dengan melihat tabel 2.

Tabel 2. Kelas pada dataset tumor otak tanpa menggunakan SMOTE

Normal	Tumor
10	81

Dapat dilihat dari tabel 2 bahwa kelas pada dataset tumor otak tidak seimbang maka selanjutnya dilakukan pengklasifikasian tanpa menggunakan teknik untuk menyeimbangkan kelas. Hasil yang didapatkan dari hasil klasifikasi tanpa menggunakan teknik untuk menyeimbangkan kelas dengan menggunakan 100 neuron tersembunyi pada Extreme Learning Machine pada dataset tumor otak dapat dilihat pada tabel 3.

Tabel 3. Hasil klasifikasi ELM dengan 100 neuron tersembunyi tanpa penyeimbang kelas.

	precision	recall	f1-score	akurasi
normal	0%	0%	0%	89.74%
tumor	91.74%	97.12%	94.02%	

Didapatkan bahwa hasil dari penelitian yang dilakukan dengan 100 neuron tersembunyi bahwa kelas normal pada klasifikasi dataset *microarray* pada

klasifikasi tumor otak didapatkan presisi, *recall*, dan *f1-score* 0% yang dapat dikatakan bahwa data yang diklasifikasikan ini adalah data yang hasil akurasinya sempurna (*overfitting*) sedangkan untuk kelas tumor selanjutnya mendapat hasil yang cukup baik dimana presisi, *recall* dan *f1-score* memiliki nilai diatas 90%, dan memiliki tingkat akurasi yang bagus 89.74%. Selanjutnya akan dilakukan penelitian dengan menggunakan 200 neuron tersembunyi pada klasifikasi *Extreme Learning Machine* tanpa menggunakan teknik menyeimbangkan kelas maka hasil yang didapatkan dari penelitian ini dapat dilihat pada tabel 4.

Tabel 4. Hasil klasifikasi ELM dengan 200 neuron tersembunyi tanpa penyeimbang kelas

	precision	recall	f1-score	akurasi
normal	29.09%	66.66%	40.47%	84.61%
tumor	97.55%	85.99%	91.42%	

Didapatkan hasil dari klasifikasi yang sudah dilakukan dengan menggunakan 200 neuron tersembunyi presisi dan *f1-score* pada kelas normal buruk yang didapatkan 29.09% dan 40.47%. Dimana hasil dari akurasi yang didapatkan dalam klasifikasi menggunakan 200 neuron tersembunyi adalah 84.61%. Selanjutnya akan dilakukan penelitian dengan menggunakan 300 neuron tersembunyi pada klasifikasi Extreme Learning Machine tanpa menggunakan teknik menyeimbangkan kelas maka hasil yang didapatkan dari penelitian ini dapat dilihat pada tabel 5.

Tabel 5. Hasil klasifikasi ELM dengan 300 neuron tersembunyi tanpa penyeimbang kelas

	precision	recall	f1-score	akurasi
normal	20.00%	66.67%	31.47%	76.92%
tumor	97.55%	96.99%	86.12%	

Dari hasil klasifikasi tumor otak dengan menggunakan 300 neuron tersembunyi didapatkan bahwa hasil dari presisi dan *f1-score* yang lebih buruk dari pada neuron tersembunyi 200 yang dimana didapatkan hasil 20% dan juga 31% menurun dari hasil akurasi dengan 100 dan 200 neuron tersembunyi. Selanjutnya akan dilakukan penelitian dengan menggunakan 400 neuron tersembunyi pada klasifikasi Extreme Learning Machine tanpa menggunakan teknik menyeimbangkan kelas maka hasil yang didapatkan dari penelitian ini dapat dilihat pada tabel 6.

Tabel 6. Hasil klasifikasi ELM dengan 400 neuron tersembunyi tanpa penyeimbang kelas

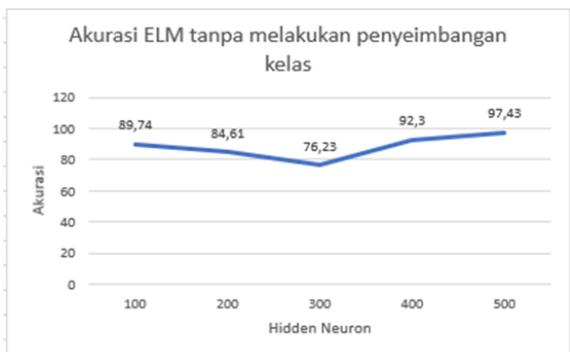
	precision	recall	f1-score	akurasi
normal	49.81%	66.87%	57.69%	92.30%
tumor	97.55%	94.29%	95.95%	

Hasil dari akurasi dengan menggunakan 400 *hidden neuron* didapatkan bahwa hasil presisi, *recall*, dan *f1-score* sudah cukup baik serta meningkat dari antara neuron tersembunyi 100, 200, dan 300. Hasil dari klasifikasi dengan menggunakan 400 *hidden neuron* mendapatkan 92.30% dimana hasil akurasi yang dihasilkan sudah sangat baik dan meningkat dari yang menggunakan 100 *hidden neuron*, 200 *hidden neuron* dan 300 *hidden neuron* dan benar penggunaan neuron tersembunyi yang lebih dari data *training* menyebabkan akurasi yang *overfitting*. Selanjutnya akan dilakukan penelitian dengan menggunakan 500 neuron tersembunyi pada klasifikasi Extreme Learning Machine tanpa menggunakan teknik menyeimbangkan kelas maka hasil yang didapatkan dari penelitian ini dapat dilihat pada tabel 7.

Tabel 7. Hasil klasifikasi ELM dengan 500 neuron tersembunyi tanpa penyeimbang kelas.

	precision	recall	f1-score	akurasi
normal	75.19%	100%	86.41%	97.43%
tumor	100%	96.96%	98.71%	

Hasil dari akurasi pada klasifikasi dengan menggunakan 500 neuron tersembunyi adalah 97.43% yang dimana ini adalah hasil akurasi yang sangat tinggi dari 100 sampai 400 neuron tersembunyi yang dimana presisi, *recall* dan *f1-score* yang dihasilkan sudah sangat baik tetapi ini adalah hasil yang dikatakan *overfitting*. Dari hasil yang sudah didapatkan dalam penelitian 100, 200, 300, 400 dan 500 neuron tersembunyi tanpa melakukan teknik oversampling SMOTE pada penelitian ini maka didapat grafik gambar 3.



Gambar 3. Akurasi ELM tanpa melakukan penyeimbangan kelas

4.3 Klasifikasi ELM dengan Oversampling SMOTE

Pada penelitian ini dilakukan klasifikasi menggunakan algoritma Extreme Learning Machines (ELM) dengan tambahan teknik oversampling SMOTE dengan menggunakan 100, 200, 300, 400, dan 500 neuron tersembunyi dalam metode Extreme Learning Machines untuk mengklasifikasi tumor otak dengan menggunakan SMOTE untuk menangani kelas yang tidak seimbang pada dataset *microarray* pada tumor otak dengan melihat tabel 8.

Tabel 8. Kelas pada dataset tumor otak dengan menggunakan SMOTE

Normal	Tumor
81	81

Dapat dilihat dari tabel 8 bahwa kelas pada dataset tumor otak sudah seimbang dengan menggunakan teknik oversampling maka selanjutnya dilakukan pengklasifikasian menggunakan ELM. Hasil yang didapatkan dari hasil klasifikasi dengan menggunakan teknik untuk menyeimbangkan kelas dengan menggunakan 100 neuron tersembunyi pada Extreme Learning Machine pada dataset tumor otak dapat dilihat pada tabel 9.

Tabel 9. Hasil klasifikasi ELM dengan 100 neuron tersembunyi dengan penyeimbang kelas

	precision	recall	f1-score	akurasi
normal	40.12%	67.41%	50.01%	89.74%
tumor	97.55%	92.29%	94.51%	

Pada hasil klasifikasi tumor otak dengan menggunakan teknik penyeimbangan kelas pada dataset *microarray* didapatkan hasil dari presisi, *recall* dan *f1-score* pada kelas normal yang meningkat dari tanpa menggunakan teknik penyeimbangan kelas SMOTE menjadi presisi 40.12%, *recall* 67.41%, dan 50.01% *f1-score* dan hasil akurasi yang didapatkan sama dari hasil klasifikasi tanpa menggunakan teknik penyeimbangan kelas dapat dilihat dengan penggunaan SMOTE baik digunakan bagi data yang berelatif sedikit untuk menghindari *overfitting*. Selanjutnya akan dilakukan penelitian dengan menggunakan 200 neuron tersembunyi pada klasifikasi Extreme Learning Machine dengan menggunakan teknik menyeimbangkan kelas maka hasil yang didapatkan dari penelitian ini dapat dilihat pada tabel 10.

Tabel 10. Hasil klasifikasi ELM dengan 200 neuron tersembunyi dengan penyeimbang kelas

	precision	recall	f1-score	akurasi
--	-----------	--------	----------	---------

normal	50.41%	67.41%	57.41%	92.30%
tumor	97.55%	94.39%	95.89%	

Pada hasil klasifikasi dengan menggunakan 200 neuron tersembunyi serta dengan dibantu adanya teknik penyeimbangan kelas maka didapat nilai dari akurasi meningkat menjadi 7.69% serta presisi, *recall*, dan *f1-score* yang dihasilkan meningkat dengan adanya bantuan dari *oversampling* SMOTE Selanjutnya akan dilakukan penelitian dengan menggunakan 300 neuron tersembunyi pada klasifikasi Extreme Learning Machine dengan menggunakan teknik menyeimbangkan kelas maka hasil yang didapatkan dari penelitian ini dapat dilihat pada tabel 11.

Tabel 11. Hasil klasifikasi ELM dengan 300 neuron tersembunyi dengan penyeimbang kelas.

	precision	recall	f1-score	akurasi
normal	23,11%	100%	38,37%	74.35%
tumor	100%	71.97%	84.28%	

Pada hasil klasifikasi dengan menggunakan 300 *hidden neuron* pada klasifikasi tumor otak dengan dibantu *oversampling* SMOTE didapatkan nilai akurasi menurun 1.88% dari tanpa adanya *oversampling* SMOTE tetapi presisi, *recall*, dan *f1-score* yang didapatkan meningkat tetapi penggunaan 300 *hidden neuron* tidak lebih baik dari akurasi dengan 100 dan 200 *hidden neuron* Selanjutnya akan dilakukan penelitian dengan menggunakan 400 neuron tersembunyi pada klasifikasi Extreme Learning Machine dengan menggunakan teknik menyeimbangkan kelas maka hasil yang didapatkan dari penelitian ini dapat dilihat pada tabel 12.

Tabel 13. Hasil klasifikasi ELM dengan 400 neuron tersembunyi dengan penyeimbang kelas

	precision	recall	f1-score	akurasi
normal	30.11%	100%	46,17%	82.05%
tumor	100%	80.94%	88.88%	

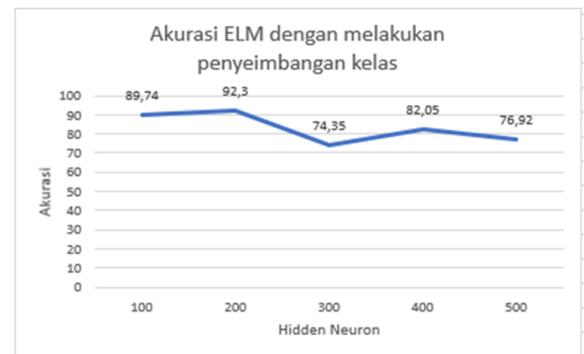
Hasil dari klasifikasi dengan menggunakan 400 *hidden neuron* pada klasifikasi tumor otak dengan menggunakan *oversampling* SMOTE didapatkan bahwa hasil nilai akurasi menurun 10.25% dari tanpa menggunakan *oversampling* SMOTE tetapi *recall* meningkat sangat baik dan akurasi dari penggunaan 400 *hidden neuron* lebih baik daripada 300 *hidden neuron* dan menurun dari penggunaan 100 dan 200 *hidden neuron* yang dimana dapat disimpulkan bahwa penggunaan teknik *oversampling* SMOTE baik untuk menangani data yang *overfitting*. Selanjutnya akan dilakukan penelitian dengan menggunakan 500 neuron tersembunyi pada klasifikasi Extreme Learning

Machine dengan menggunakan teknik menyeimbangkan kelas maka hasil yang didapatkan dari penelitian ini dapat dilihat pada tabel 14.

Tabel 14. Hasil klasifikasi ELM dengan 500 neuron tersembunyi dengan penyeimbang kelas.

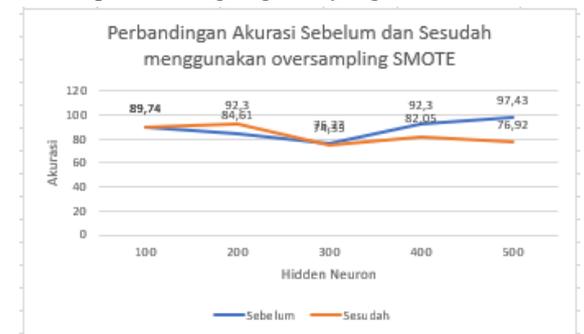
	precision	recall	f1-score	akurasi
normal	30.11%	100%	46,17%	82.05%
tumor	100%	80.94%	88.88%	

Hasil dari klasifikasi yang didapatkan dengan menggunakan 500 *hidden neuron* serta menggunakan teknik *oversampling* SMOTE didapatkan hasil akurasi yang menurun serta presisi, *recall* dan *f1-score* yang menurun drastis dari semua percobaan penelitian jumlah *hidden neuron* yang dimana dapat disimpulkan bahwa penggunaan teknik *oversampling* SMOTE baik untuk menangani data yang *overfitting*. Dari hasil yang sudah didapatkan dalam penelitian 100, 200, 300, 400 dan 500 neuron tersembunyi dengan melakukan teknik *oversampling* SMOTE pada penelitian ini maka didapat grafik gambar 4.



Gambar 4. Akurasi ELM dengan melakukan penyeimbangan kelas

Dari hasil yang sudah di dapatkan maka dapat disimpulkan dengan grafik yang ada dibawah ini



Gambar 5. Perbandingan akurasi ELM dengan melakukan penyeimbangan kelas

Dapat dilihat pada grafik gambar 5 hasil perbandingan yang didapat dari penelitian ini bahwa penggunaan SMOTE berguna untuk memperakurat

hasil akurasi dari klasifikasi *Extreme Learning Machine* dan menjadi salah satu metode untuk mengurangi *overfitting*.

5. Kesimpulan dan Saran

Pada penelitian ini didapatkan bahwa penggunaan hidden neuron serta penyeimbangan data pada klasifikasi data microarray sangat berpengaruh dalam akurasi yang akan didapatkan dalam penelitian ini. Performa kinerja klasifikasi tertinggi setiap model antara lain model Extreme Learning Machines mendapatkan akurasi sebesar 97.43% pada hidden neuron = 500. Lalu Extreme Learning Machines menggunakan oversampling SMOTE pada keseluruhan data menghasilkan akurasi sebesar 92.30% pada hidden neuron = 200. Untuk penelitian kedepannya disarankan untuk mencoba seleksi fitur dikarenakan jumlah fitur pada dataset ini sebanyak 21050 fitur untuk lebih meningkatkan akurasi dari penelitian ini, dan mencoba perbandingan menggunakan teknik oversampling lainnya seperti Adasyn, Outlier-SMOTE, C2 dan teknik oversampling lainnya.

Daftar Pustaka:

- Adiwijaya, A. (2018). Deteksi Kanker Berdasarkan Klasifikasi Microarray Data. *Jurnal Media Informatika Budidarma*, 2(4), 181. <https://doi.org/10.30865/mib.v2i4.1043>
- Asniar, Maulidevi, N. U., & Surendro, K. (2021). SMOTE-LOF for noise identification in imbalanced data classification. *Journal of King Saud University - Computer and Information Sciences*, xxx.
- Balakrishnan, R., & Karthikeyan, T. (2019). Microarray gene expression and multiclass cancer classification using extreme learning machine (ELM) with refined group search optimizer (RGSO). *Int Sci J Sci Eng Technol*, 3, 64–76.
- Fadilla, I., Adikara, P. P., & Setya Perdana, R. (2018). Klasifikasi Penyakit Chronic Kidney Disease (CKD) Dengan Menggunakan Metode Extreme Learning Machine (ELM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(10), 3397–3405.
- Fitriani, I. M., Ratnawati, D. E. and Anam, S. (2019) 'Klasifikasi Senyawa Kimia dengan Notasi Simplified Molecular Input Line Entry System (SMILES) menggunakan Metode Extreme Learning Machine (ELM)', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(5), pp. 4516–4524.
- Gumelar, G., Ain, Q., Marsuciati, R., Bambang, S. A., & Sunyoto, A. (2021). Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance. *Sisfotek*, 250–255.
- Ihsani, D. A., Arifin, A., & Fatoni, M. H. (2020). Klasifikasi DNA Microarray Menggunakan Principal Component Analysis (PCA) dan Artificial Neural Network (ANN). *Jurnal Teknik ITS*, 9(1).
- M, T. A. and Azizah, Q. N. (2022) 'Klasifikasi Tumor Otak Menggunakan Ekstraksi Fitur HOG dan Support Vector Machine', 4(1), pp. 45–50
- Nurviarelda, R., Adiwijaya, A. and ... (2018) 'Klasifikasi Data Microarray Menggunakan Discrete Wavelet Transform Dan Naive Bayes Classification', *eProceedings ...*, 5(1), pp. 1536–1540.
- Ranjani Rani, R., & Ramyachitra, D. (2018). Microarray cancer gene feature selection using spider monkey optimization algorithm and cancer classification using SVM. *Procedia Computer Science*, 143, 108–116.
- Rusamsi, S. R., Adiwijaya, & Rohmawati, A. A. (2018). Deteksi Kanker Berdasarkan Klasifikasi Microarray Data Menggunakan Wavelet Transform dan Backpropagation Termodifikasi dengan Conjugate Gradient Flechter Reeves. *E-Proceeding of Engineering*, 5(1), 1772–1779.
- Siringoringo, R. (2018). Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbor. *Jurnal ISD*, 3(1), 44–49.
- Sulistiyowati, N., & Jajuli, M. (2020). Integrasi Naive Bayes Dengan Teknik Sampling Smote Untuk Menangani Data Tidak Seimbang. *Nuansa Informatika*, 14(1), 34.
- Turlapati, V. P. K., & Prusty, M. R. (2020). Outlier-SMOTE: A refined oversampling technique for improved detection of COVID-19. *Intelligence-Based Medicine*, 3–4(July), 100023.
- Wahid, R. R., Anggraeny, F. T., & Nugroho, B. (2020). Implementasi Metode Extreme Learning Machine untuk Klasifikasi Tumor Otak pada Citra Magnetic Resonance Imaging. *Prosiding Seminar Nasional Informatika Bela Negara (SANTIKA)*, 1, 16–20.

