

Prediksi Harapan Hidup Pasien Kanker Paru-Paru Pasca Operasi Bedah Thoraks Menggunakan Boosted Neural Network Dan Smote

Iedam Fardian Anshori¹, Dwiza Riana²

^{1, 2}, Jurusan Magister Ilmu Komputer, STMIK Nusa Mandiri, Jakarta, 10450, Indonesia

E-mail: fardian902@gmail.com

E-mail: dwiza@nusamandiri.ac.id

Abstrak- Kanker paru merupakan penyebab utama kematian utama dalam kelompok kanker. Penyebab paling umum kanker paru adalah paparan dalam jangka waktu yang lama terhadap asap tembakau. Operasi toraks merupakan salah satu cabang ilmu kedokteran yang mempelajari diagnosis dan tindakan operasi pada gangguan kesehatan yang disebabkan penyakit atau cedera pada kerongkongan, paru-paru atau organ tubuh lain yang berada di dada. Namun terdapat banyak resiko dan komplikasi pasca operasi bedah toraks hingga berujung pada kematian. Pada penelitian ini, akan melakukan prediksi harapan hidup pasien kanker paru-paru setelah menjalani kehidupan satu tahun pasca operasi bedah toraks menggunakan *computer aided diagnosis* (CAD). Prediksi ini dilakukan dengan menganalisa kondisi pasien sebelum dan sesudah operasi. Data yang digunakan pada penelitian ini merupakan data sekunder yang berisi 470 data dengan sebaran 400 data pasien yang hidup (*survival*) dan 70 data pasien yang meninggal (*die*). Kondisi jumlah sampel yang tidak seimbang menyebabkan rendahnya prediksi pada kelas *minor*. Sehingga, pada penelitian ini diusulkan metode yang dapat mengatasi ketidakseimbangan kelas. *Adaptive Boost* digunakan sebagai optimasi level algoritma dan *synthetic minority over-sampling technique* (SMOTE) digunakan sebagai optimasi level data yang diterapkan pada algoritma *neural network*. Hasil penelitian menunjukkan bahwa metode yang diusulkan membuat peningkatan yang signifikan dalam prediksi harapan hidup pasien kanker paru pasca operasi bedah toraks. Rata-rata nilai *GMeans* yang dihasilkan oleh metode yang diusulkan sebesar 74.17% dengan nilai prediksi tertinggi sebesar 76.76%. Sementara algoritma *neural network* tanpa optimasi hanya menghasilkan rata-rata nilai *GMeans* sebesar 53.42% dengan hasil tertinggi sebesar 61.26%. Hasil uji beda sebelum dan sesudah menggunakan optimasi menunjukkan bahwa penelitian ini mampu meningkatkan performa algoritma *neural network* dalam mengatasi ketidakseimbangan kelas secara signifikan.

Kata kunci: operasi bedah toraks, *neural network*, *adaptive boost*, SMOTE

Abstract— Lung cancer is the leading cause of death in the cancer group. The most common cause of lung cancer is long-term exposure to tobacco smoke. Thoracic surgery is a branch of medicine that studies the diagnosis and operation of medical problems caused by disease or injury to the esophagus, lungs or other organs in the chest. However, there are many risks and complications after thoracic surgery, leading to death. In this study, will predict the life expectancy of lung cancer patients after one year of life after thoracic surgery using computer aided diagnosis (CAD). This prediction is done by analyzing the patient's condition before and after surgery. The data used in this study is secondary data which contains 470 data with the distribution of 400 patient data (survival) and 70 data of patients who died (die). The condition of the number of samples that is not balanced causes low predictions in the minor class. Thus, in this study, a method that can overcome class imbalances is proposed. Adaptive Boost is used as algorithm level optimization and synthetic minority over-sampling technique (SMOTE) is used as data level optimization applied to the neural network algorithm. The results showed that the proposed method made a significant increase in the prediction of life expectancy of lung cancer patients after thoracic surgery. The average GMeans value generated by the proposed method was 74.17% with the highest predictive value of 76.76%. Meanwhile, the neural network algorithm without optimization only produces an average GMeans value of 53.42% with the highest result of 61.26%. The results of different tests before and after using optimization show that this study can significantly improve the performance of the neural network algorithm in overcoming class imbalances.

Keywords— thoracic surgery, neural network, adaptive boost, SMOTE.

I. PENDAHULUAN

Kanker paru-paru adalah suatu kondisi dimana sel-sel tumbuh secara tidak terkendali di dalam paru-paru oleh sejumlah karsinogen. Kanker paru-paru merupakan salah satu jenis kanker yang paling umum terjadi [1]. Menurut *World Health Organization* (WHO), kanker paru merupakan penyebab utama kematian utama dalam kelompok kanker. Penyebab paling umum kanker paru adalah paparan dalam jangka waktu yang lama terhadap asap tembakau [2].

Gejala dan tanda kanker paru-paru pada umumnya diantaranya batuk, hemoptisis, terdapat abses pada paru, nyeri dada, sesak nafas, suara serak [3]. Penanganan pada kanker paru-paru dilakukan dengan terapi bedah, radioterapi dan kemoterapi.

Penanganan pasien penderita kanker paru-paru salah satunya dapat dilakukan dengan pembedahan toraks. Saluran

pernafasan terletak hampir seluruhnya di dalam toraks. Sehingga toraks mempunyai peranan penting dalam pernapasan [4]. Pasien yang membutuhkan pengobatan atau intervensi dengan operasi untuk penyakit paru-paru harus menjalani operasi toraks [5].

Operasi toraks merupakan salah satu cabang ilmu kedokteran yang mempelajari diagnosis dan tindakan operasi pada gangguan kesehatan yang disebabkan penyakit atau cedera pada kerongkongan, paru-paru atau organ tubuh lain yang berada di dada. Operasi toraks melibatkan banyak dokter ahli dari dokter bedah toraks dan kardiovaskuler serta dokter spesialis jantung. Operasi toraks dapat menangani berbagai penyakit seperti kanker paru, tumor serta pertumbuhan jaringan lunak yang terletak di paru-paru, kanker kerongkongan, sulit bernafas dan sebagainya. Operasi bedah toraks juga dapat dilakukan untuk transplantasi paru-paru dan penyumbatan yang terjadi di jantung atau pembuluh arteri [6].

Masalah utama pada penelitian ini adalah bagaimana melakukan prediksi harapan hidup pasien pasca operasi dengan menggunakan *computer aided diagnosis (CAD) System*. Dataset pasien pasca operasi bedah toraks memiliki dua kelas yaitu, meninggal dalam jangka waktu satu tahun (*die*) dan mampu bertahan hidup (*survival*) dengan jumlah sampel data untuk kelas *die* sebanyak 70 dan untuk kelas *survival* sebanyak 400 sampel [7].

Penelitian sebelumnya pada sampel data pasien operasi toraks untuk prediksi harapan hidup pasca operasi dilakukan oleh beberapa peneliti menggunakan berbagai algoritma, antara lain: *naïve bayes* [8], *rule based classification* [9], *support vector machine* [10], *multi layer perceptron*, *semi naïve bayesian* [11] dan *ensemble svm* [7].

Masalah yang dialami oleh beberapa peneliti sebelumnya adalah kondisi sampel yang tidak berimbang dimana sampel *die* hanya sebanyak 70 sampel dan sampel *survival* mencapai 400 sampel hingga menyebabkan masalah ketidakseimbangan kelas pada dataset ini.

Ketidakeimbangan kelas adalah distribusi yang tidak merata diantara kelas-kelas dimana satu kelas lebih banyak diantara kelas lain [12] [13]. Masalah ini berpengaruh pada prediksi harapan hidup pasca operasi yang rendah [7]. Algoritma klasifikasi yang dipengaruhi oleh masalah ketidakseimbangan kelas untuk satu set data tertentu akan melihat akurasi yang baik secara keseluruhan, tetapi akurasi yang sangat buruk pada kelas minoritas [14].

Algoritma *neural network* dipilih karena memiliki kemampuan untuk mendeteksi dan menganalisa permasalahan yang sifatnya kompleks [15] dan non-linear [16] [17]serta perhitungan dilakukan secara paralel sehingga waktu komputasi lebih cepat [18].

Teknik optimasi yang digunakan pada penelitian ini adalah teknik *boosting*, menggunakan algoritma AdaBoost. AdaBoost digunakan karena sangat mudah diimplementasikan [19], tidak perlu mengatur parameter [7] dan fleksibel [20] sehingga dapat dikombinasikan dengan berbagai algoritma [21].

Optimasi level data dapat digunakan beberapa teknik yang dapat digunakan, diantaranya SMOTE, Random Over Sampling (ROS) dan Random Under Sampling (RUS). SMOTE dan ROS meningkatkan jumlah kelas minoritas sehingga jumlah sampel dapat berimbang. SMOTE memiliki waktu prediksi yang lebih lambat, akan tetapi lebih optimal untuk mengatasi ketidakseimbangan kelas [22]. Teknik SMOTE juga lebih mudah dan cepat diaplikasikan dibandingkan metode lain [23].

Artikel ini disusun dengan pembahasan sebagai berikut: Section II membahas metode yang digunakan pada penelitian ini, Section III membahas hasil eksperimen serta analisa pembahasan dilanjutkan dengan kesimpulan serta rencana penelitian berikutnya pada Section IV.

II. METODE PENELITIAN

A. Desain Penelitian

Menurut [24] ada empat metode penelitian yang paling umum digunakan yaitu: *action research*, *experiment*, *case study*, dan *survey*. Pada penelitian kali ini menggunakan metode *experiment*, yaitu penelitian yang melibatkan penyelidikan kepada beberapa variabel menggunakan tes

tertentu yang dikendalikan sendiri oleh peneliti. Pada penelitian ini dilakukan beberapa tahapan penelitian sebagai berikut:

- 1) *Pengumpulan Data*: pada tahapan ini dijelaskan tentang bagaimana data dalam penelitian ini didapatkan. Pada tahap ini juga ditentukan data yang akan diproses.
- 2) *Pengolahan Data Awal*: pengolahan data awal meliputi pembersihan data, pentransformasian data ke dalam bentuk yang dibutuhkan serta pengelompokan dan penentuan atribut data.
- 3) *Metode yang Diusulkan*: setelah pengolahan data awal, lalu dibuatkan model yang sesuai dengan jenis data. Pembagian data ke dalam data pelatihan dan data pengujian juga diperlukan untuk pembuatan model.
- 4) *Eksperimen dan Pengujian Model*: pada tahapan ini, dilakukan eksperimen dan pengujian model terhadap data yang sebelumnya sudah diolah. Perhitungan dengan setiap algoritma akan diulang untuk mendapatkan parameter terbaik.

TABLE I
ATTRIBUT DATASET

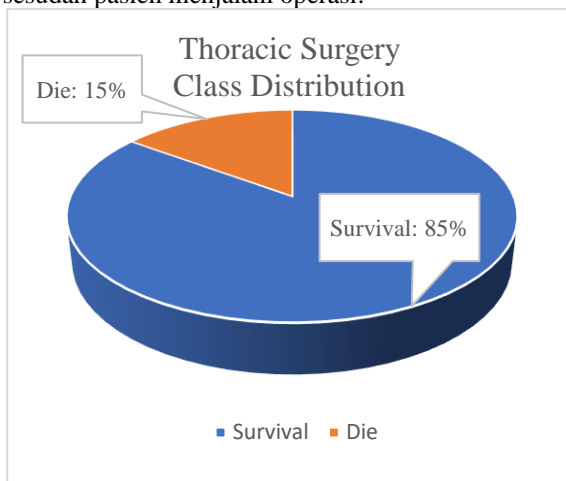
Attributes	Descriptions	Data Type
DGN	Diagnosis - kombinasi spesifik kode ICD-10 untuk tumor primer dan sekunder serta lebih dari satu tumor, jika ada	Nominal
PRE4	Jumlah udara yang bisa dihembuskan secara paksa dari paru-paru setelah mengambil nafas sedalam mungkin (FVC)	Numerik
PRE5	Jumlah udara yang telah dihembuskan pada akhir detik pertama dari FVC (FEV1)	Numerik
PRE6	Ukuran kemampuan umum pasien kanker dalam aktivitas sehari-hari (Zubrod Scale)	Nominal
PRE7	Rasa sakit sebelum operasi	Binary
PRE8	<i>Haemoptysis</i> sebelum operasi	Binary
PRE9	<i>Dyspnoea</i> sebelum operasi	Binary
PRE10	Batuk sebelum operasi	Binary
PRE11	Kondisi lemah sebelum operasi	Binary
PRE14	Ukuran tumor (TNM)	Nominal
PRE17	Diabetes	Binary
PRE19	<i>myocardial infarction</i> (MI) hingga 6 bulan	Binary
PRE25	Penyakit yang menyerang arteri/aliran darah (PAD)	Binary
PRE30	Merokok	Binary
PRE32	Asma	Binary
AGE	Usia saat operasi	Numerik
RISK	Mampu bertahan hidup setelah 1 tahun bernilai T jika meninggal	Binary

5) *Eksperimen dan Pengujian Model*: tahapan ini, eksperimen dan pengujian model dilakukan terhadap data yang telah diolah. Perhitungan dengan setiap algoritma akan diulang untuk mendapatkan parameter terbaik.

6) *Evaluasi dan Validasi Hasil*: tahap evaluasi merupakan tahap yang terakhir dari kegiatan penelitian. Seluruh hasil eksperimen akan di evaluasi untuk menemukan model algoritma yang paling baik.

B. Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan data sekunder. Data sekunder adalah data yang tidak diperoleh langsung dari obyek penelitian, melainkan telah dikumpulkan oleh pihak lain. Data sekunder yang digunakan pada penelitian ini merupakan kumpulan data dari Wroclaw Thoracic Surgery Center, data tersebut merupakan data pasien penderita kanker paru-paru yang menjalani operasi bedah thoraks dari tahun 2009 hingga 2014 [7]. Dataset ini diambil dari UCI Machine Learning Repository yang diunduh melalui <http://archive.ics.uci.edu/ml/datasets>. Dataset ini berisi informasi setiap pasien yang diwakili dalam data yang ditetapkan oleh 16 atribut yang merupakan kondisi sebelum dan sesudah pasien menjalani operasi.



Gambar 1. Distribusi Kelas Dataset *Thoracic Surgery*

16 atribut tersebut merupakan data nominal, numeric dan binary. Dataset pasien pasca operasi bedah toraks memiliki dua kelas yaitu, meninggal dalam jangka waktu satu tahun (*die*) dan mampu bertahan hidup (*survival*) dengan jumlah sampel data untuk kelas *die* sebanyak 70 dan untuk kelas *survival* sebanyak 400 sampel. Distribusi jumlah sampel pada masing-masing kelas dapat dilihat pada Gambar 1. Penjelasan atribut yang ada pada dataset *thoracic surgery* dapat dilihat pada Tabel I. Atribut pada dataset *thoracic surgery* memiliki varian tipe data diantaranya, nominal, binary dan numerik.

C. Pengolahan data Awal

Pengolahan data awal merupakan tindak lanjut dari pengumpulan data. Pada penelitian ini, disebutkan bahwa dataset memiliki distribusi kelas yang tidak seimbang, sehingga pengolahan data awal difokuskan pada penambahan data artifisial pada kelas minor agar distribusi kelas dapat seimbang. Penambahan data menggunakan teknik SMOTE untuk membuat data artifisial dengan acak. Setelah data

artifisial berhasil ditambahkan dan distribusi kelas telah seimbang, maka tahap selanjutnya adalah membagi dataset menjadi dua yaitu, *data training* dan *data testing*.

Data awal yang diperoleh terdiri dari 470 records. Setelah proses SMOTE, dataset terdiri dari 800 records. Untuk menguji model yang dikembangkan, data akan dibagi menjadi dua bagian, yaitu *data training* dan *data testing*. *Data training* digunakan untuk pengembangan model, sedangkan *data testing* digunakan untuk pengujian model. Diketahui bahwa jumlah data adalah 800 dengan pembagian 70% digunakan untuk *data training* dan 30% digunakan untuk *data testing*, dengan masing-masing menghasilkan jumlah 560 dan 240 data. Tahap selanjutnya adalah memilih data yang akan digunakan sebagai *data training* dan *data testing* menggunakan *proportionate stratified random sampling*.

D. Metode Yang di Usulkan

Metode yang diusulkan yaitu menerapkan teknik SMOTE pada dataset *thoracic surgery* untuk menghasilkan data artifisial atau buatan untuk menyeimbangkan distribusi kelas pada dataset *thoracic surgery*. Setelah itu, menerapkan algoritma *neural network* sebagai algoritma *classifier* yang telah dilakukan dioptimasi oleh algoritma *boosting AdaBoost*. Kemudian hasil klasifikasi di evaluasi akurasi menggunakan *confusion matrix*. Gambaran metode yang diusulkan dapat dilihat pada Gambar 2.

E. Eksperimen dan Pengujian Model

Penelitian yang dilakukan dalam eksperimen ini menggunakan komputer untuk melakukan proses perhitungan pada model yang diusulkan. Proses eksperimen pada penelitian ini adalah:

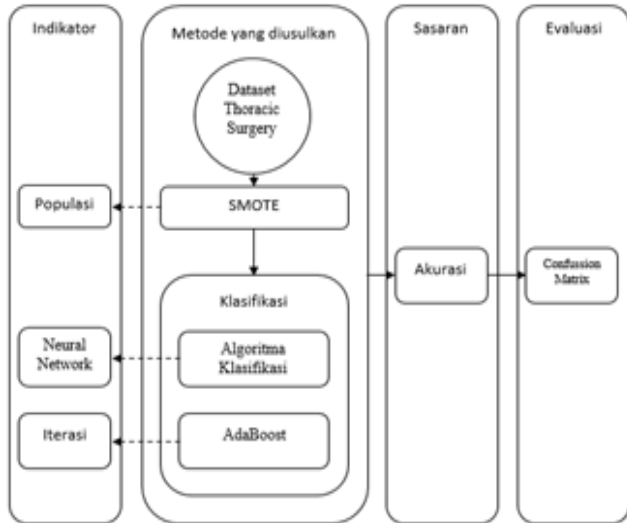
- 1) Menggunakan 2 dataset untuk eksperimen, dataset *training* serta dataset *testing*.
- 2) Mendesain arsitektur *neural network*.
- 3) Melakukan *training* dan *testing* terhadap model *Algoritma neural network* dan mencatat hasil akurasi.
- 4) Melakukan *oversampling* menggunakan SMOTE
- 5) Mendesain arsitektur *neural network* dengan optimasi *AdaBoost*
- 6) Melakukan *training* dan *testing* terhadap dataset yang telah di *oversampling* menggunakan *neural network + AdaBoost* dan mencatat hasil akurasi.
- 7) Melakukan komparasi hasil akurasi pada kedua model dan menguji beda dengan t-Test.
- 8) Mengembangkan aplikasi.

F. Evaluasi dan Validasi Hasil

Pada tahap ini akan dilakukan proses pengujian model yang dihasilkan oleh tool Rapidminer dengan mengevaluasi perbandingan hasil akurasi seluruh eksperimen yaitu eksperimen *neural network* pada dataset yang belum di *oversampling* oleh SMOTE. Eksperimen *neural network* yang telah dioptimasi oleh *AdaBoost* terhadap dataset yang sudah di *oversampling* oleh SMOTE. Sementara itu, evaluasi yang digunakan adalah *confusion matrix*. Kemudian memvalidasi hasil akurasi antara eksperimen-eksperimen tersebut menggunakan uji coba t-Test.

III. HASIL DAN PEMBAHASAN

Hasil dalam penelitian dilakukan dalam dua eksperimen yaitu eksperimen terhadap algoritma *neural network* tanpa optimasi, serta eksperimen terhadap algoritma *neural network* dengan menggunakan SMOTE dan AdaBoost.



Gambar 2. Metode yang diusulkan

TABEL II

HASIL EKSPERIMEN DENGAN *NEURAL NETWORK* + SMOTE + ADABOOST

Rasio	<i>Neural Network</i> + SMOTE + AdaBoost			
	Accuracy	True Positive	True Negative	Gmeans
90-10	76.25	80	73.33	76.59
80-20	72.5	71.95	73.08	72.51
70-30	73.33	71.8	75	73.38
60-40	71.25	69.1	73.94	71.48
50-50	76.75	77.16	76.35	76.75

A. Hasil Experimen dan Pengujian

Eksperimen pada dataset thoracic surgery menggunakan algoritma *neural network* dengan menggunakan optimasi pada level data menggunakan SMOTE dan optimasi pada level algoritma menggunakan AdaBoost. Eksperimen dilakukan sebanyak lima kali dengan variasi menggunakan validasi split validation. Lima eksperimen tersebut dengan berbagai variasi rasio yaitu, 90% - 10%, 80% - 20%, 70% - 30%, 60% - 40% dan 50% - 50%. Indikator untuk mengetahui hasil terbaik ditunjukkan oleh besarnya nilai akurasi untuk masing-masing eksperimen.

Hasil eksperimen pada Tabel II menunjukkan bahwa algoritma *neural network* yang dioptimasi menggunakan AdaBoost dan SMOTE mampu menghasilkan performa akurasi yang lebih baik jika dibandingkan dengan hasil eksperimen lain. Metode yang diusulkan menghasilkan akurasi rata-rata sebesar 74.02% dengan akurasi terbaik didapatkan pada rasio 50-50 dengan akurasi sebesar 76.75%. Kemampuan

mengatasi masalah ketidakseimbangan kelas ditunjukkan dengan nilai *GMeans* rata-rata sebesar 74.17% dengan nilai *GMeans* terbesar diperoleh pada rasio 50-50 dengan *GMeans* sebesar 76.75%. Perbandingan nilai *GMeans* dari hasil eksperimen ini dapat dilihat pada Gambar 3.

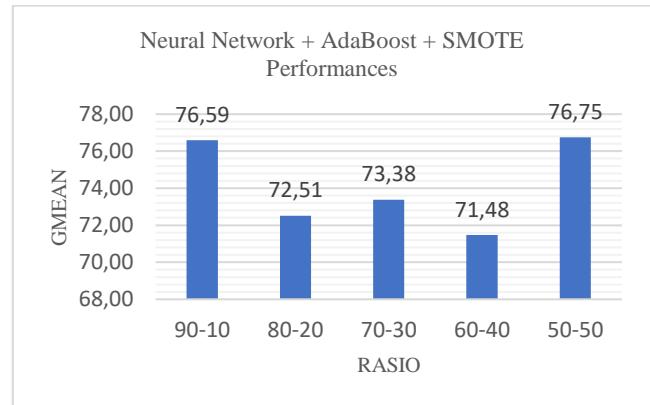
B. Pembahasan

Berdasarkan hasil eksperimen pada beberapa algoritma klasifikasi lain, diperoleh perbedaan nilai akurasi pada pengujian algoritma naïve bayes, decision tree dan support vector machine yang dikombinasi dengan AdaBoost dan SMOTE. Hasil perbandingan eksperimen tersebut dengan hasil eksperimen pada *neural network* dapat dilihat pada Tabel II.

TABEL III

PERBANDINGAN ALGORITMA *NEURAL NETWORK* DENGAN ALGORITMA LAIN

Rasio	Neural Network + AdaBoost + SMOTE	Naïve Bayes + AdaBoost + SMOTE	Decision Tree + AdaBoost + SMOTE	Support Vector Machine + AdaBoost + SMOTE
90-10	76.59	50.00	76.65	72.725
80-20	72.51	62.82	75.315	72.555
70-30	73.38	58.55	75.785	25
60-40	71.48	58.50	70.53	25
50-50	76.75	75.58	70.855	25
AVG	74.17	61.09	73.83	44.06



Gambar 3. Grafik Perbandingan Performa *Neural Network* + SMOTE + AdaBoost

Hasil eksperimen pada Tabel III menunjukkan bahwa rata-rata *GMeans* terbaik diperoleh pada algoritma *neural network* yang telah dioptimasi dengan AdaBoost dan SMOTE yang menghasilkan rata-rata *GMeans* sebesar 74.17%. Algoritma yang paling mendekati hasil *neural network* adalah *decision tree* yang menghasilkan rata-rata *GMeans* sebesar 73.83%. sementara hasil terendah diperoleh algoritma *support vector machine* yang menghasilkan rata-rata *GMeans* sebesar 44.06%. Signifikansi perbedaan nilai pada hasil eksperimen pada algoritma dengan dan tanpa optimasi AdaBoost dan SMOTE dapat dilihat pada Tabel IV.

Berdasarkan Tabel III maka dapat dibuktikan bahwa dari empat algoritma yang dioptimasi, metode yang diusulkan dapat

meningkatkan performa tiga algoritma dengan signifikan. Tiga algoritma tersebut adalah *neural network*, *decision tree* dan *naïve bayes*. Sementara metode yang diusulkan tidak dapat meningkatkan performa algoritma *support vector machine* secara signifikan.

C. Perbandingan Akurasi Metode Yang di Usulkan dengan Penelitian Lain

Berdasarkan rangkaian eksperimen yang telah dilakukan. Metode yang diusulkan lebih unggul jika dibandingkan dengan penelitian [7]. Pada penelitian sebelumnya, *boosted SVM* digunakan untuk melakukan prediksi harapan hidup pasien pasca operasi bedah thoraks, *boosted SVM* menghasilkan rata-rata nilai *GMeans* sebesar 64,81%. Penelitian ini juga mengungguli penelitian yang dilakukan oleh [25] yang menggunakan algoritma *back propagation* untuk melakukan prediksi harapan hidup pasien pasca operasi bedah thoraks dengan hasil rata-rata *GMeans* sebesar 51.1%.

TABEL IV
SIGNIFIKANSI PERBEDAAN NILAI ALGORITMA KLASIFIKASI

Algoritma	P(T<=t) two-tail	Signifikansi
Neural Network+AdaBoost+SMOTE	0.004815395	< 0.05 (Sig.)
Naïve Bayes+AdaBoost+SMOTE	0.002732415	< 0.05 (Sig.)
Decision Tree+AdaBoost+SMOTE	0.01240682	< 0.05 (Sig.)
Support Vector Machine+AdaBoost+SMOTE	0.85868734	> 0.05 (Not Sig.)

TABEL V
PERBANDINGAN *GMEANS* METODE YANG DIUSULKAN

No.	Peneliti	Metode	<i>GMeans</i>
1.	Pandey, Khan, & Swetapadma (2017) [25]	<i>Back Propagation</i>	51.5%
2.	Zieba, Tomczak, Lubicz, & Swiatek (2014) [7]	<i>Boosted SVM</i>	64.81%
3.	Nachev & Reapy (2015) [26]	<i>Naïve Bayes</i>	69.7%
4.	Desuky & Bakrawy (2016) [10]	<i>SVM + Relief-f</i>	71.62%
5.	Penelitian Ini	<i>Boosted Neural Network + SMOTE</i>	74.17%

Sementara itu, penelitian yang diusulkan [10] yang menerapkan *relief-f* yang dikombinasikan dengan *support vector machine* yang menghasilkan rata-rata nilai *GMeans* sebesar 71.62%. Beberapa algoritma lain yang pernah diterapkan untuk prediksi harapan hidup pasien pasca operasi

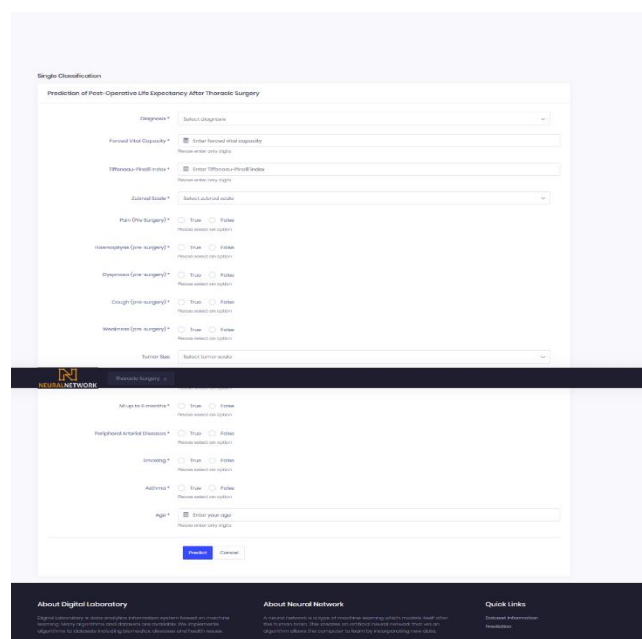
bedah thoraks diantaranya *decision tree*, *naïve bayes* dan *support vector machine* yang dilakukan oleh [26] dengan hasil terbesar memiliki rata-rata nilai *GMeans* sebesar 69.7%.

Perbandingan performa yang dihasilkan oleh metode yang diusulkan dengan penelitian sebelumnya dapat dilihat pada Tabel V.

D. Implementasi

Hasil rangkaian eksperimen dan evaluasi telah dilakukan, sehingga didapatkan model algoritma yang paling baik yakni *Boosted Neural Network* yang dioptimasi oleh SMOTE. Dibuatlah sebuah aplikasi yang digunakan untuk melakukan prediksi harapan hidup pasien kanker paru pasca operasi bedah toraks dengan mengimplementasikan model algoritma tersebut. Aplikasi yang dibangun merupakan aplikasi berbasis web yang dapat diakses pada laman <http://scriptseeker.id/thorax>. Pada aplikasi ini, pengguna dapat menemukan informasi mengenai atribut-atribut yang terdapat pada dataset yang digunakan pada penelitian ini.

Tujuan utama aplikasi ini dibangun adalah untuk melakukan prediksi harapan hidup pasien kanker paru pasca operasi bedah toraks dengan mekanisme, pengguna memasukan kondisi pasien kanker paru sebelum dan sesudah operasi bedah toraks kemudian memproses perhitungan prediksi menggunakan tombol Predict. Setelah proses perhitungan selesai, akan muncul hasil prediksi apakah pasien dapat bertahan hidup selama setahun kedepan atau tidak. Gambaran aplikasi yang dihasilkan dari penelitian ini dapat dilihat pada Gambar 4.



Gambar. 4
Implementasi Metode yang diusulkan

IV. KESIMPULAN

Penelitian ini mengkombinasikan teknik *boosting* AdaBoost sebagai optimasi level algoritma dan SMOTE sebagai optimasi level data pada algoritma *neural network* untuk mengatasi

masalah ketidakseimbangan kelas pada prediksi harapan hidup pasien kanker paru-paru pasca operasi bedah thoraks. Berdasarkan hasil eksperimen pada penelitian ini, maka dapat ditarik kesimpulan sebagai berikut:

- 1) Teknik *boosting* AdaBoost diterapkan untuk mengoptimasi algoritma *neural network* untuk mengatasi ketidakseimbangan kelas. Akan tetapi, hasil eksperimen menunjukkan hasil yang kurang optimal dengan hanya menghasilkan *GMeans* rata-rata hanya sebesar 48.72%, dengan prediksi pada kelas minor masih rendah dengan nilai prediksi terkecil sebesar 0% dan tertinggi sebesar 21.50%
- 2) Teknik SMOTE diterapkan untuk mengoptimasi algoritma *neural network* pada level data untuk mengatasi ketidakseimbangan kelas. Eksperimen menunjukkan hasil yang sangat optimal dengan menghasilkan *GMeans* rata-rata sebesar 73.50% dengan nilai prediksi tertinggi sebesar 76.00% pada rasio 90-10.

Kedua teknik optimasi diterapkan untuk meningkatkan kinerja algoritma *neural network* menggunakan AdaBoost dan SMOTE menghasilkan rata-rata *GMeans* yang sangat baik. Hasil eksperimen menunjukkan hasil rata-rata *GMeans* sebesar 74.17% dengan nilai prediksi tertinggi sebesar 76.76% pada rasio 50-50. Selain itu, penerapan AdaBoost dan SMOTE pada algoritma *neural network* menghasilkan peningkatan kinerja algoritma *neural network* secara signifikan.

Pada penelitian ini, secara umum AdaBoost dan SMOTE diterapkan untuk optimasi pada level algoritma dan level data untuk mengatasi masalah ketidakseimbangan kelas dapat meningkatkan rata-rata nilai *GMeans* pada prediksi harapan hidup pasien kanker paru pasca operasi bedah thoraks, namun beberapa hal dapat diterapkan untuk menyempurnakan penelitian ini, diantaranya:

- 1) Pada penelitian selanjutnya metode usulan pada penelitian ini dapat menggunakan metode metaheuristik lain seperti, particle swarm optimization, ant colony, bee colony dan metode metaheuristik lainnya untuk fitur seleksi.
- 2) Optimasi pada level algoritma juga dapat digunakan menggunakan optimasi parameter pada algoritma neural network untuk mendapatkan parameter yang paling optimal menggunakan algoritma genetika.
- 3) Optimasi pada level data dapat menggunakan teknis sampling lain seperti random under sampling untuk mengurangi jumlah kelas mayor, atau random over sampling untuk meningkatkan jumlah kelas minor sebagai solusi lain untuk mengatasi ketidakseimbangan kelas.

REFERENSI

- [1] L. Jameson, A. S. Fauci dan D. L. Kasper, *Harrison's Principles of Internal Medicine*, Twentieth Edition, New York: McGraw-Hill Education, 2018.
- [2] M. J. Thun, L. M. Hannan dan L. L. Adams-Campbell, "Lung cancer occurrence in never-smokers: an analysis of 13 cohorts and 22 cancer registry studies," *PLOS Medicine*, pp. 1357-1371, 2008.
- [3] K. K. Brown, T. Lee-Chiong, S. Chapman dan G. Robinson, *Oxford American Handbook of Pulmonary Medicine*, Oxford: Oxford University Press, 2009.
- [4] N. G. Y. Asih dan C. Effendy, *Keperawatan Medical Bedah*, Jakarta: Buku Kedokteran EGC, 2014.
- [5] M. K. Ferguson, *Thoracic Surgery Atlas*, Elsevier, 2007.
- [6] J. B. Putnam, "Lung, chest wall, pleura, and mediastinum," dalam *Sabiston Textbook of Surgery. 19th ed*, Philadelphia, Saunders Elsevier, 2012.
- [7] M. Zieba, J. M. Tomczak, M. Lubicz dan J. Swiatek, "Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients," *Applied Soft Computing*, pp. 99-108, 2013.
- [8] P. R. Hachesu, N. Moftian, M. Dehghani dan T. S. Soltani, "Analyzing a Lung Cancer Patient Dataset with the Focus on Predicting Survival Rate One Year after Thoracic Surgery," *Asian Pacific Journal of Cancer Prevention*, pp. 1531-1536, 2017.
- [9] M. Koklu, H. Kahramanli dan N. Allahverdi, "APPLICATIONS OF RULE BASED CLASSIFICATION TECHNIQUES FOR THORACIC SURGERY," dalam *Managing Intellectual Capital and Innovation for Sustainable and Inclusive Society Management, Knowledge and Learning Joint International Conference*, Bari, 2015.
- [10] A. S. Desuky dan L. M. Bakrawy, "Improved Prediction of Post-operative Life Expectancy after Thoracic Surgery," *Advances in Systems Science and Application*, vol. 16, no. 2, pp. 70-80, 2016.
- [11] B. Hui, H. Zhou, Y. Jiang, L. Ji dan J. Chen, "The Research of Postoperative Life Expectancy of Lung Cancer Based on Semi Naive Bayesian," *Computer Science and Artificial Intelligence(CSAI)*, pp. 17-19, 2017.
- [12] T. Deepa dan M. Punithavalli, "An Innovative Optimization Algorithm for Feature Selection - A Comparative Study," *International Journal of Computer Science and Information Technology & Security*, pp. 20-24, 2013.
- [13] R. T. Prasetyo dan P. Pratiwi, "PENERAPAN TEKNIK BAGGING PADA ALGORITMA KLASIFIKASI UNTUK MENGATASI KETIDAKSEIMBANGAN KELAS DATASET MEDIS," *Jurnal Informatika*, vol. II, no. 2, pp. 395-403, 2015.
- [14] D. Tiwari, "Handling Class Imbalance Problem Using Feature Selection," *International Journal of Advanced Research in Computer Science & Technology*, pp. 516-520, 2014.
- [15] R. T. Prasetyo dan E. Ripandi, "Optimasi Klasifikasi Jenis Hutan Menggunakan Deep Learning Berbasis Optimize Selection," *Jurnal Informatika*, vol. VI, no. 1, pp. 100-106, 2019.
- [16] J. Bourquin, H. Schmidli, P. v. Hoogevest dan H. Leuenberger, "Advantages of Artificial Neural Networks (ANNs) as alternative modelling technique for data sets

- showing non-linear relationships using data from a galenical study on a solid dosage form,” *European Journal of Pharmaceutical Sciences*, pp. 5-16, 2018.
- [17] D. Riana, Y. Ramdhani, R. T. Prasetio dan A. N. Hidayanto, “Improving Hierarchical Decision Approach for Single Image Classification of Pap Smear,” *International Journal of Electrical and Computer Engineering*, vol. VIII, no. 6, pp. 5415-5424, 2018.
- [18] A. Sharma dan A. Chopra, “Artificial Neural Networks: Applications In Management,” *IOSR Journal of Business and Management (IOSR-JBM)*, pp. 32-40, 2013.
- [19] A. Sharma dan S. Dey, “A boosted SVM based sentiment analysis approach for online opinionated text,” dalam *Proceedings of the 2013 Research in Adaptive and Convergent Systems*, Montreal, 2013.
- [20] R. T. Prasetio, A. A. Rismayadi dan I. F. Anshori, “Implementasi Algoritma Genetika pada k-nearest neighbours untuk Klasifikasi Kerusakan Tulang Belakang,” *Jurnal Informatika*, vol. V, no. 2, pp. 186-194, 2018.
- [21] Q. Huang, Y. Chen, L. Liu, D. Tao dan X. Li, “On Combining Biclustering Mining and AdaBoost for Breast Tumor Classification,” *IEEE Transactions on Knowledge and Data Engineering*, p. 1, 2019.
- [22] B. J. Park, S. K. Oh dan W. Pedrycz, “The Design of Polynomial Function-Based Neural Network Predictors for Detection of Software Defects,” *Information Sciences*, pp. 40-57, 2013.
- [23] D. Yu, J. Hu, Z. Tang, H. Shen, J. Yang dan J. Yang, “Neurocomputing Improving protein-ATP binding residues prediction by boosting SVMs with random under-sampling,” *Neurocomputing*, vol. 104, pp. 180-190, 2013.
- [24] C. W. Dawson, *Projects in Computing and Information Systems : A Student's Guide 2nd Edition*, London: Pearson Education Limited, 2009.
- [25] A. K. Pandey, M. A. Khan dan A. Swetapadma, “A Back Propagation Neural Network Based Method for Post Life Expectancy Estimation of Thoracic Surgery Patients,” dalam *International Conference on Smart Technology for Smart Nation*, Bengaluru, 2017.
- [26] A. Nachev dan T. Reapy, “Predictive Models for Post-Operative Life Expectancy after Thoracic Surgery,” *Mathematical and Software Engineering*, vol. 1, no. 1, pp. 1-5, 2015.