

KLASIFIKASI PENYAKIT JANTUNG KORONER MENGUNAKAN SELEKSI FITUR DAN *SUPPORT VECTOR MACHINE*

¹Amiruddin Fadlli, ²Muhammad Imron Rosadi

^{1,2}Program Studi Teknik Informatika
Universitas Yudharta Pasuruan
Pasuruan

¹amiru.fadli@gmail.com, ²imron.rosadi@yudharta.ac.id

Received : September, 2018

Accepted : Nopember, 2018

Published : Desember, 2018

ABSTRACT

Coronary heart disease was blood intake resistance caused by the accumulation of plaque in the coronary arteries that supply oxygen to the heart muscle. Expensive and Impact coronary angiography test to diagnose this disease leads researchers to apply the classification algorithm of Support Vector Machine and engineering ReliefF, Consistency-based and Medical Knowledge-based Feature Selection. RLF and CNS reduction 13 features into 3 and 11, while increasing the accuracy of 81,111% to 83,333 84,444% and 83,333% for CNS reduction. MFS reduces 13 features into 8 while reducing the accuracy of 81,111% to 72,222%, but combining the RLF Along MFS improve accuracy of 72,222% to 83,333 %.

Keyword: Coronary heart disease, Support Vector Machine, Feature Selection

ABSTRAK

Penyakit jantung koroner ialah terhambatnya asupan darah yang disebabkan oleh bertumpuknya plak di dalam arteri koroner yang mensuplai oksigen ke otot jantung. Mahal serta beresikonya uji coronary angiography untuk mendiagnosis penyakit ini menuntun peneliti untuk menerapkan algoritma klasifikasi Support Vector Machine dan teknik ReliefF, Consistency-based dan Medical Knowledge-based Feature Selection. RLF dan CNS mereduksi 13 fitur menjadi 3 dan 11 selagi meningkatkan akurasi dari 81.111% menjadi 84.444% dan 83.333%. MFS mereduksi 13 fitur menjadi 8 selagi mengurangi akurasi dari 81.111% menjadi 72.222%, namun menggabungkan RLF Bersama MFS meningkatkan akurasi dari 72.222% menjadi 83.333%.

Kata kunci: jantung koroner, Support Vector Machine, Seleksi fitur

I. Pendahuluan

Penyakit yang disebabkan oleh bertumpuknya plak di dalam arteri koroner yang mensuplai oksigen ke otot jantung, sehingga mengakibatkan kerusakan yang berat pada jantung disebut penyakit jantung koroner (PJK) (Rajkumar & Reena, 2010). Pendahuluan berisi latar belakang masalah, tujuan penelitian dan manfaat penelitian. Uji *coronary angiography* lebih dipilih oleh pakar jantung dalam mendiagnosis keberadaan PJK terhadap pasien dengan akurasi yang tinggi meskipun *invasive*, beresiko dan juga mahal. Bila melihat kekurangan uji lanjutan ini, maka perlu dikembangkan sebuah metode yang mampu mendiagnosis keberadaan PJK sebelum dilakukan *coronary angiography* (Setiawan, Venkatachalam, & Hani, 2009).

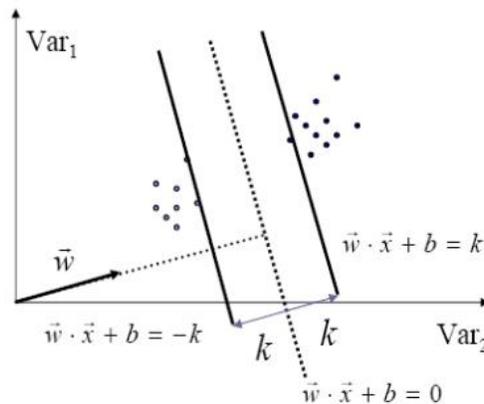
Beberapa penelitian sebelumnya telah menyediakan detail perbandingan dari beberapa classifier untuk deteksi PJK di mana SMO telah menunjukkan potensi sebagai algoritma klasifikasi pada area ini, terutama ketika mempertimbangkan total akurasi sebagai ukuran performa (Nahar, Imam, Tickle, & Chen, 2013). (Purbasari & Nugroho, 2013) telah menunjukkan bahwa teknik seleksi fitur RLF dan CNS memiliki kehandalan dalam mereduksi fitur yang tidak relevan dan redundant, begitu pula dengan MFS yang telah ditunjukkan oleh (Nahar, Imam, Tickle, & Chen, 2013). Maka penulis akan melakukan penelitian untuk prediksi PJK menggunakan algoritma klasifikasi SVM berbasis teknik seleksi fitur MFS yang digabungkan dengan RLF dan CNS dengan harapan dapat memperbaiki akurasi dari penelitian-penelitian sebelumnya.

Hasil dari penelitian ini diharapkan dapat bermanfaat untuk membantu peneliti selanjutnya untuk merancang-bangun sistem prediksi penyakit jantung koroner yang dapat menghasilkan hasil diagnosa yang tidak bias terhadap diagnosis yang dilakukan oleh pakar. Selain itu, juga diharapkan dapat menjadi saran bagi tenaga medis untuk mendiagnosis penyakit jantung koroner sebelum dilakukan metode gold standard, sehingga prosedur diagnosis penyakit jantung koroner yang *invasive*, beresiko, serta tidak ekonomis terhadap pasien dapat dihindari.

II. Tinjauan Pustaka

Klasifikasi adalah proses pembelajaran suatu fungsi target (tujuan) f yang memetakan tiap-tiap himpunan atribut x ke suatu label dari label kelas y yang telah didefinisikan sebelumnya. SVM melakukan klasifikasi dengan membangun *hyperplane* linier yang optimal di dalam dimensi yang lebih tinggi, dengan bantuan *support vector* dan *margin*, yang memisahkan data ke dalam 2 kategori atau kelas (Jiang, Wang, Zhang, & Wang, 2010) seperti pada gambar 1.

Gambar 1. *Hyperplane* yang optimal



Sumber: (Vadicherla & Sonawane, 2013)

Pada Gambar 2.2 SVM menghitung klasifikasi linear dari bentuk,

$$f(x) = w^{\rightarrow} \cdot x^{\rightarrow} + b \quad (1)$$

di mana w adalah *weight vector*, x adalah sampel latih, dan b adalah *bias*. *Hyperplane* yang memisahkan kelas positif dan negatif didefinisikan,

$$w^{\rightarrow} \cdot x^{\rightarrow} + b = 0 \quad (2)$$

Maka data x_i yang termasuk kelas +1 (sampel positif) dapat dirumuskan melalui pertidaksamaan,

$$w^{\rightarrow} \cdot x^{\rightarrow} + b = +1 \quad (3)$$

Begitu juga dengan data x_i yang termasuk kelas -1 (sampel negatif) dapat dirumuskan melalui pertidaksamaan,

$$w^{\rightarrow} \cdot x^{\rightarrow} + b = -1 \quad (4)$$

Di SVM yang dicari adalah *decision boundary* yang memaksimal kan *margin* yang memisahkan poin data training yang positif dari yang negatif. Hal ini dapat dirumuskan sebagai permasalahan *Quadratic Programming* (QP), yakni mencari titik minimal persamaan 2.5 sambil memenuhi 2.6,

$$1/2 \|w^{\rightarrow}\|^2 \quad (5)$$

$$y_i (w^{\rightarrow} \cdot (x_i)^{\rightarrow} + b) \geq 1 \quad (6)$$

y adalah kelas dari x . Permasalahan ini dapat diselesaikan dengan beberapa teknik komputasi, di antaranya adalah *Lagrange Multiplier* berikut,

$$L = \frac{1}{2} \|w\|^2 - \sum_i \alpha_i (y_i (\vec{w} \cdot \vec{x}_i + b) - 1) \quad (7)$$

α_i inilah yang disebut *Lagrange Multipliers*, yang bernilai nol atau positif ($\alpha_i \geq 0$). Nilai optimal dari persamaan 2.7 dapat ditemukan dengan meminimalkan L terhadap w dan b , dan memaksimalkan L terhadap α (alpha). w , b , dan α inilah parameter yang akan menentukan solusi maksimal *margin*. Pada *margin*, poin-poin positif dan negatif yang menyentuh garis batas *margin* disebut dengan *support vectors*. *Support vectors* memiliki nilai α yang positif, sedangkan yang bukan memiliki nilai $\alpha=0$.

Tujuan dari seleksi atribut ialah untuk mengidentifikasi tingkat kepentingan suatu atribut dalam kumpulan data dan membuang seluruh atribut lain seperti informasi yang tidak relevan dan berlebihan. Gambar 2 mendemonstrasikan teknik pemeringkat atribut.

Fitur: {A, B, C, D}	Set Fitur	Skor	Atribut Tunggal Terbaik
Iterasi 0	[]	0.00	
Iterasi 1	[A]	0.20	B
	[B C]	0.40	
	[C D]	0.30	
	[D]	0.15	
Iterasi 2	[A B]	0.38	C
	[B C]	0.65	
	[B D]	0.47	
Iterasi 3	[A B C]	0.60	A
	[B C D]	0.57	
Iterasi 4	[A B C D]	0.62	D

Gambar 2. Proses seleksi fitur
Sumber: (Hall & Holmes, 2003)

Relief merupakan teknik pemeringkat atribut berbasis *instance* yang menyampling acak sebuah *instance* dari data dan mencari tetangga terdekatnya dari kelas yang sama dan yang berlawanan. Nilai atribut tetangga terdekat dibandingkan dengan sampel *instance* dan memperbarui skor relevansi tiap atribut. Ini berulang sesuai jumlah *m instance* yang ditentukan oleh user. Pada awalnya Relief didefinisikan untuk masalah 2 kelas dan kemudian diperluas (ReliefF) untuk menangani noise dan data set multiclass (Kononenko, 1994). ReliefF menghaluskan pengaruh noise data dengan merata-rata kontribusi dari *k-nearest neighbors* dari kelas yang sama dan berlawanan dari tiap sampel *instance* bukannya tetangga terdekat tunggal.

```

set all weights  $W[A] = 0.0$ 
for  $i = 1$  to  $m$  do
  begin
    randomly select an instance  $R$ 
    find  $k$  nearest hits  $H_j$ 
    for each class  $C \neq class(R)$  do
      find  $k$  nearest misses  $M_j(C)$ 
    for  $A = 1$  to #attributes do
       $W[A] = W[A] - \frac{\sum_{j=1}^k \text{diff}(A, R, H_j)}{(m \times k) + \sum_{C \neq class(R)} \left[ \frac{P(C)}{1 - P(class(R))} \sum_{j=1}^k \text{diff}(A, R, M_j(C)) \right]} / (m \times k)$ 
    end
  end

```

Gambar 3. Algoritma ReliefF
Sumber: (Hall & Holmes, 2003)

Fungsi perhitungan dari rumus $\text{diff}(\text{Attribute}, \text{Instance1}, \text{Instance2})$ yang terdapat pada Gambar 3 menghitung perbedaan antara nilai-nilai dari atribut untuk dua *instance*. Untuk atribut diskrit, diff bernilai antara 1 (bila nilai-nilainya berbeda) atau 0 (bila nilai-nilainya sama), namun untuk atribut *continuous* diff bernilai selisih aktual yang dinormalisasikan ke dalam interval 0-1. (Kononenko, 1994) mencatat bahwa semakin besar nilai dari *m* (jumlah sampel *instance*), semakin estimasi ReliefF dapat diandalkan, meskipun menambah nilai *m* juga menambah waktu proses.

Metode pendekatan untuk seleksi subset atribut menggunakan konsistensi kelas diusulkan oleh (Liu & Setiono, 1996) dengan mencari kombinasi dari atribut-atribut yang nilainya membagi data ke dalam himpunan yang mengandung sebuah mayoritas dari kelas tunggal yang kuat. Biasanya, pencarian tersebut membias kepada himpunan fitur yang kecil namun dengan konsistensi kelas yang tinggi. Metrik konsistensinya ialah sebagai berikut:

$$Consistency_s = 1 - \frac{\sum_{i=0}^J |D_i| - |M_i|}{N} \quad (8)$$

di mana s merupakan himpunan atribut, J adalah jumlah kombinasi dari nilai atribut yang berbeda dari s , $|D_i|$ ialah jumlah *occurrences* dari kombinasi nilai atribut ke- i , $|M_i|$ yaitu *cardinality* dari kelas mayoritas terhadap kombinasi nilai atribut ke- i , dan N yakni jumlah total *instance* yang berada di dalam kumpulan data.

Terdapat banyak faktor penyebab penyakit jantung, yang meskipun tidak satupun kurang penting untuk diagnosis, namun beberapa faktor mungkin lebih layak untuk dipertimbangkan saat mendiagnosis penyakit jantung menurut beberapa literatur medis. Melalui ikhtisar dari beberapa literatur medis, (Nahar, Imam, Tickle, & Chen, 2013) mempertimbangkan delapan faktor signifikan medis untuk seleksi fitur sebagai MFS, yakni *age*, *chest pain type*, *resting blood pressure*, *cholesterol*, *fasting blood sugar*, *resting heart rate*, *maximum heart rate*, dan *exercise induced angina*.

III. Metode Penelitian

Menjelaskan metode penelitian yang digunakan serta langkah-langkah penelitian. Jenis penelitian, pendekatan penelitian dan instrumen penelitian.

Terdapat beberapa tahap dalam melakukan pengolahan data eksperimen, pada penelitian ini akan menggunakan model *Cross-Standard Industri for Data Mining* (CRISP-DM). Model CRISP-DM menurut (Sumathi & Sivanandam, 2006) terbagi dalam enam fase sebagai berikut:

3.1 Tahap *Business Understanding*

Dalam penelitian ini akan dilakukan prediksi untuk mendeteksi penyakit jantung koroner menggunakan algoritma *Support Vector Machine* bersama teknik seleksi fitur *ReliefF*, *Consistency-based* dan *Medical Knowledge-based Feature Selection*.

3.2 Tahap *Data Understanding*

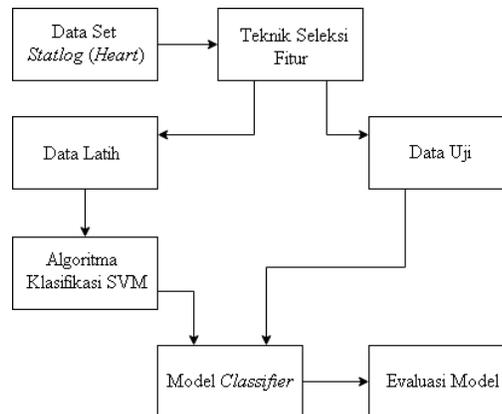
Data yang diolah pada penelitian merupakan data pasien penyakit jantung pada *Statlog (Heart) Data Set* dari repositori UCI yang memiliki 13 atribut dan 1 kelas yakni *age*, *sex*, *chest pain type*, *resting blood pressure*, *serumcholesterol*, *fasting blood sugar*, *resting ecg*, *maximum heart rate*, *exercise induced angina*, *old peak*, *slope for peak peak exercise*, *number of major vessels*, *thal* dan *class*.

3.3 Tahap *Data Preparation*

Dikarnakan data set yang akan digunakan pada penelitian ini tidak terdapat *missing values*, maka pada proses preparasi data set hanya perlu direduksi dengan cara mengurangi fitur atau atribut sebelum memasuki algoritma klasifikasi *Support Vector Machine* menggunakan teknik seleksi fitur *ReliefF* dan CNS yang disediakan oleh Weka dan juga teknik seleksi fitur MFS.

3.4 Tahap *Modelling*

Tahap *modelling* dikerjakan untuk memilih dan menerapkan teknik pemodelan yang tepat, menentukan *tool* yang akan digunakan pada *data mining* serta menentukan parameter dengan nilai yang optimal. Pada tahap ini pula disebut dengan tahap pembelajaran atau *learning*, karna pada tahap inilah data *training* akan dilatih oleh model algoritma yang telah dipilih, dan kemudian data *testing* akan diujikan ke dalam model *classifier* yang sudah tercipta untuk diketahui nilai evaluasinya sebagaimana gambar 4.



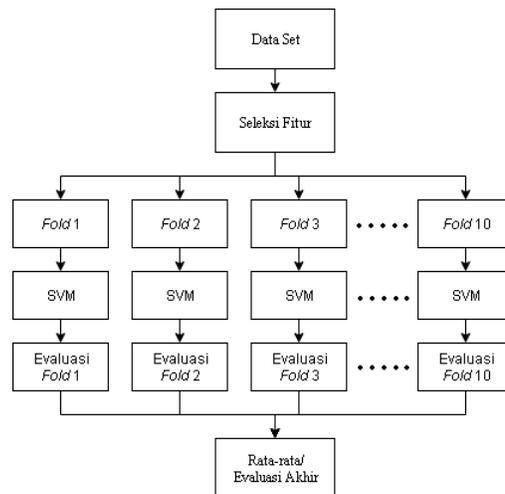
Gambar 4. Pemodelan penelitian

3.5 Tahap *Evaluation*

Pengujian akan dilakukan dengan menggunakan *k-folds cross validation* dan *confusion matrix*. Seluruh data akan dibagi menjadi 10 *subset*, yaitu: *fold 1, fold 2, fold 3, fold 4, fold 5, fold 6, fold 7, fold 8, fold 9, fold 10*. Jadi, pelatihan akan dilakukan berulang-ulang. Pada setiap pengulangan, 9 *fold subset* akan dijadikan data latih dan 1 *fold subset* akan dijadikan data uji. Proses pengulangan akan dilakukan sebanyak 10 kali hingga seluruh *fold* akan pernah berperan sebagai data latih dan data uji.

Gambar 5 berikut menjelaskan proses *10-fold cross validation*, di mana data set yang telah melalui proses seleksi fitur akan dibagi menjadi 10 bagian. Data set yang terdapat pada tiap *fold* kemudian akan dilakukan proses klasifikasi menggunakan algoritma klasifikasi *Support Vector Machine*. Yang kemudian pada tiap *fold* juga akan dilakukan proses evaluasi untuk mencari nilai dari *accuracy, precision, recall* dan *F-measure*. Yang kemudian pada akhirnya hasil evaluasi dari tiap *fold* tersebut akan dirata-rata sebagai hasil evaluasi akhir.

Hasil evaluasi akhir yang akan dijadikan acuan dalam penelitian ini nantinya adalah *accuracy, precision, recall* dan *F-measure*. Semua nilai tersebut akan didapatkan dari *confusion matrix* yang berupa nilai *True Positive, False Positive, True Negative, dan False Negative* sebagaimana ditunjukkan oleh persamaan 9, 10, 11 dan 12.



Gambar 5. Proses *cross validation*

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

$$F - measure = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (12)$$

IV. Hasil dan Pembahasan

Menjelaskan hasil penelitian dan pembahasan. Menguji hipotesis, mengukur akurasi, menjelaskan sebab akibat berdasarkan hasil analisa data dan sebagainya.

Dataset dengan nama *Statlog (Heart) Data Set* yang diambil dari (UCI Machine Learning Repository, 1988) berisi 270 data. Data set tersimpan dalam format '.dat' yang kemudian diimpor ke dalam Microsoft Excel untuk diubah ke dalam format *Comma Separated Value (CSV)*. Dikarnakan data set tersebut tidak memiliki *missing values*, maka proses transformasi data tidak dilakukan.

4.1 Seleksi Fitur *ReliefF*

Penerapan teknik seleksi fitur *ReliefF* menggunakan WEKA dengan seluruh isi data set secara utuh menghasilkan peringkat dan bobot atribut sebagai berikut:

Tabel 1 Peringkat dan bobot atribut *RLF*

Peringkat	No Urut	Nama	Bobot
1	12	nmv	0.16333
2	3	cpt	0.15296
3	13	thal	0.13741
4	2	sex	0.11259
5	11	spe	0.07556
6	7	recg	0.06481
7	9	eia	0.05889
8	10	op	0.02517
9	8	mhr	0.02122
10	4	rbp	0.01788
11	1	age	0.01552
12	6	fbs	0.01148
13	5	sc	0.00668

Kemudian dilakukanlah proses klasifikasi pada WEKA menggunakan algoritma SVM satu-persatu berurutan mulai dari menggunakan semua fitur sampai dengan satu fitur terbaik hasil keluaran dari proses seleksi fitur *ReliefF*. Proses ini tidak menggunakan fitur transformasi data baik *normalization* maupun *standardization* pada WEKA. Hasil *evaluation* dari *cross validation* pada teknik ini tertera pada Tabel 2 berikut:

Tabel 2 Hasil *evaluation* SVM dan RLF

	<i>Acc</i>	<i>Prc</i>	<i>Rcl</i>	<i>F-ms</i>
<i>all</i>	81.11%	0.81	0.86	0.835
<i>best 1</i>	74.07%	0.75	0.80	0.774
<i>best 2</i>	71.11%	0.77	0.69	0.725
<i>best 3</i>	84.44%	0.84	0.89	0.864
<i>best 4</i>	84.44%	0.84	0.89	0.864
<i>best 5</i>	83.33%	0.84	0.86	0.851
<i>best 6</i>	82.22%	0.83	0.85	0.842
<i>best 7</i>	83.70%	0.84	0.88	0.857
<i>best 8</i>	82.96%	0.83	0.87	0.851
<i>best 9</i>	83.33%	0.83	0.87	0.853
<i>best 10</i>	82.59%	0.82	0.88	0.849
<i>best 11</i>	82.59%	0.82	0.89	0.850
<i>best 12</i>	82.59%	0.82	0.89	0.850

Berdasarkan tabel 2 dapat ditarik kesimpulan bahwa 3 dan 4 fitur terbaik yang dipilih oleh *ReliefF* menghasilkan akurasi tertinggi dibandingkan dengan sejumlah fitur terbaik lainnya. 3 fitur terbaik dipilih, karna lebih memenuhi tugas daripada 4 fitur terbaik untuk mereduksi dimensi data set dan mempercepat proses klasifikasi menggunakan algoritma SVM yang kemudian akan dikombinasikan dengan teknik seleksi fitur MFS pada tahap selanjutnya.

4.2 Seleksi Fitur CNS

Penerapan teknik seleksi fitur CNS menggunakan WEKA dengan seluruh isi data set secara utuh menghasilkan sebuah subset atribut dan nilai *Consistency* pada tabel 3.

Tabel 3 Hasil seleksi fitur CNS

Subset	<i>Consistency</i>
{age,sex,cpt,fbs,recg,mhr, eia,op,spe,nmv,thal }	0.985

Kemudian dilakukanlah proses klasifikasi pada WEKA menggunakan algoritma SVM menggunakan semua fitur pada tabel 3. Proses ini tidak menggunakan fitur transformasi data baik *normalization* maupun *standardization* pada WEKA. Hasil yg didapatkan ialah *confusion matrix* pada tabel 4.

Tabel 4 *Confusion matrix* SVM dan CNS

Correct Classification	Classified as	
	+	-
+	133	17
-	28	92

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} = \frac{133 + 92}{133 + 28 + 17 + 92} = \frac{225}{270} = 0.83333 * 100\% = 83.333\%$$

$$Precision = \frac{TP}{TP + FP} = \frac{133}{133 + 28} = \frac{133}{161} = 0.826$$

$$Recall = \frac{TP}{TP + FN} = \frac{133}{133 + 17} = \frac{133}{150} = 0.887$$

$$F - measure = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2}{\frac{1}{0.826} + \frac{1}{0.887}} = 0.855$$

4.3 Seleksi Fitur MFS

Sebagaimana yang telah ditunjukkan oleh (Nahar, Imam, Tickle, & Chen, 2013) dalam kajian literatur medis mereka, atribut-atribut yang terpilih oleh seleksi fitur berbasis pengetahuan medis ini adalah *age, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic, maximum heart rate* dan *exercise induced angina*.

Kemudian dilakukanlah proses klasifikasi pada WEKA menggunakan algoritma SVM menggunakan semua fitur MFS. Proses ini tidak menggunakan fitur transformasi data baik *normalization* maupun *standardization* pada WEKA. Hasil yg didapatkan ialah *confusion matrix* pada tabel 5.

Tabel 5 *Confusion matrix* SVM dan MFS

Correct Classification	Classified as	
	+	-
+	111	39
-	36	84

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} = \frac{111 + 84}{111 + 36 + 39 + 84} = \frac{195}{270} = 0.72222 * 100\% = 72.222\%$$

$$Precision = \frac{TP}{TP + FP} = \frac{111}{111 + 36} = \frac{111}{147} = 0.755$$

$$Recall = \frac{TP}{TP + FN} = \frac{111}{111 + 39} = \frac{111}{150} = 0.740$$

$$F - measure = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2}{\frac{1}{0.755} + \frac{1}{0.740}} = 0.747$$

4.4 Seleksi Fitur MFS + RLF

Pada tahap ini peneliti akan mencoba menggabungkan atribut-atribut hasil seleksi fitur berbasis pengetahuan medis yakni MFS dan atribut-atribut hasil teknik seleksi fitur berbasis komputerisasi yakni teknik seleksi fitur CNS maupun RLF. Namun sayangnya peneliti tidak dapat menggabungkan atribut-atribut hasil seleksi MFS dengan atribut-atribut hasil seleksi CNS, karna atribut-atribut hasil seleksi mereka apabila digabungkan akan menjadi atribut data set asalnya itu sendiri. Maka, peneliti mencoba menggabungkan atribut-atribut hasil seleksi MFS dengan 3 atribut terbaik hasil seleksi *ReliefF* sebagaimana tertera pada Tabel 6.

Tabel 6 Atribut seleksi MFS + RLF

MFS	MFS \cap RLF	RLF
age	<i>chest pain type</i>	<i>number major vessels</i>
<i>resting blood pressure</i>		thal
<i>serum cholesterol</i>		
<i>fasting blood sugar</i>		
<i>resting electro cardio graphic</i>		
<i>maximum heart rate</i>		
<i>exercise induced angina</i>		

Kemudian dilakukan klasifikasi pada WEKA dengan SVM dan gabungan fitur MFS dan RLF. Proses ini tidak menggunakan fitur transformasi data baik *normalization* maupun *standardization* pada WEKA. Hasil yg didapatkan ialah *confusion matrix* pada tabel 7.

Tabel 7 Confusion matrix SVM- MFS+RLF

Correct Classification	Classified as	
	+	-
+	131	19
-	26	94

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} = \frac{131 + 94}{131 + 26 + 19 + 94} = \frac{225}{270} = 0.83333 * 100\% = 83.333\%$$

$$Precision = \frac{TP}{TP + FP} = \frac{131}{131 + 26} = \frac{131}{157} = 0.834$$

$$Recall = \frac{TP}{TP + FN} = \frac{131}{131 + 19} = \frac{131}{150} = 0.873$$

$$F - measure = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2}{\frac{1}{0.834} + \frac{1}{0.873}} = 0.853$$

4.5 Perbandingan Nilai Evaluasi

Tabel 8 berikut ini menampilkan perbandingan nilai evaluasi hasil klasifikasi pada tahap-tahap sebelumnya:

Tabel 8 Perbandingan nilai evaluasi

	Acc	Prc	Rcl	F-ms
All	81.11%	0.81	0.86	0.835
RLF	84.44%	0.84	0.89	0.864
CNS	83.33%	0.83	0.89	0.855
MFS	72.22%	0.76	0.74	0.747
MFS+ RLF	83.33%	0.83	0.87	0.853

V. Kesimpulan

Berdasarkan hasil penelitian yang telah dicapai menggunakan metode yang diusulkan, maka dapat ditarik kesimpulan sebagai berikut:

1. Teknik seleksi fitur ReliefF, CNS dan MFS mampu mengurangi jumlah atribut yang awalnya berjumlah 13 buah menjadi 3, 11 dan 8 buah yang mengakibatkan pada berkurangnya dimensi data set.
2. Teknik seleksi fitur ReliefF dan CNS berhasil meningkatkan akurasi klasifikasi SVM menjadi 84.44% dan 83.33% dari yang pada saat menggunakan semua fitur akurasinya hanya sebesar 81.11%. Di sisi lain, teknik seleksi fitur MFS menurunkan akurasi dari 81.11% menjadi 72.22%. Namun, dengan digabungkannya MFS bersama ReliefF dapat meningkatkan akurasi menjadi 83.33% dari yang pada saat menggunakan MFS saja hanya sebesar 72.22%.

Daftar Pustaka

- Hall, M. A., & Holmes, G. (2003). Benchmarking Attribute Selection Techniques for Discrete Class Data Mining. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, 1437-1447.
- Jiang, Q., Wang, G., Zhang, T., & Wang, Y. (2010). Predicting Human microRNA-disease Associations Based on Support Vector Machine. *IEEE International Conference on Bioinformatics and Biomedicine*, 467-472.
- Kononenko, I. (1994). Estimating attributes: Analysis and extensions of RELIEF. *European Conference on Machine Learning*, 171-182.
- Liu, H., & Setiono, R. (1996). A Probabilistic Approach to Feature Selection: A Filter Solution. *Proc. 13th International Conference on Machine Learning*, 319-327.
- Nahar, J., Imam, T., Tickle, K. S., & Chen, Y.-P. P. (2013). Computational intelligence for heart disease diagnosis: A medical knowledge driven approach. *Expert Systems with Applications* 40, 96-104.
- Purbasari, I. Y., & Nugroho, B. (2013). Benchmarking Algoritma Pemilihan Atribut Pada Klasifikasi Data Mining. *SNASTIA*, 47-54.
- Rajkumar, A., & Reena, G. S. (2010). Diagnosis Of Heart Disease Using Datamining Algorithm. *Global Journal of Computer Science and Technology*, 38-43.
- Setiawan, N. A., Venkatachalam, P., & Hani, A. F. (2009). Diagnosis of Coronary Artery Disease Using Artificial Intelligence Based Decision Support System. *Proceedings of the International Conference on Man-Machine Systems (ICoMMS)*, 1-5.
- Sumathi, S., & Sivanandam, S. N. (2006). *Introduction to Data Mining and its Applications*. Berlin: Springer.
- UCI Machine Learning Repository. (1988, July 1). *UCI Machine Learning Repository: Statlog (Heart) Data Set*. Retrieved April 10, 2018, from UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29>
- UCI Machine Learning Repository. (1998). *UCI Machine Learning Repository: Data Sets*. Retrieved April 10, 2018, from UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/datasets.html>
- Vadicherla, D., & Sonawane, S. (2013). Decision Support System for Heart Disease based on Sequential Minimal Optimization in Support Vector Machine. *International Journal of Engineering Sciences & Emerging Technologies*, 19-26.
