

Klasifikasi Kejadian Hipertensi dengan Metode *Support Vector Machine* (SVM) Menggunakan Data Puskesmas di Kota Bandar Lampung

¹Indah Martika Pasaribu, ²Favorisen Rosyking Lumbanraja, ³Dewi Asiah Shofiana, dan ⁴Aristoteles

^{1,2,3,4} Program Studi Ilmu Komputer, Universitas Lampung
Jalan Soemantri Brojonegoro No. 1 Gedung Meneng, Bandar Lampung, Indonesia
e-mail : ¹indah.martikapasaribu@gmail.com, ²favorisenrosykinglumbanraja@fmipa.unila.ac.id,
³dewiasiahshofiana@fmipa.unila.ac.id, ⁴aristoteles@gmail.com

Abstract — Hypertension is a condition in which a person experiences an increase in blood pressure above the normal value which causes pain and even death. Normal human blood pressure is 120/80 mmHg. Patients with hypertension cannot be cured, but prevention and control can be done. The hypertension cases are always increasing in Indonesia. The Bandar Lampung City Health Service stated that hypertension is a disease that always ranks in the top ten diseases in Bandar Lampung City. Diagnosis of hypertension is currently manually performed by requiring a lot of energy, materials, and time. Based on the condition, there is an idea to apply the field of biomedical data analysis to help diagnosing hypertension using the support vector machine (SVM) method in Bandar Lampung City. This study classifies and measures the accuracy of the support vector machine method in hypertension. The data comes from five health centers in Bandar Lampung City from 2017 to 2019 with 10-fold cross validation data sharing. The kernels used are linear, gaussian, and polynomial kernels. This study successfully classifies hypertension sufferers in Bandar Lampung City. The result of the highest feature correlation analysis is 0.90. The results of the classification using the support vector machine method get the highest accuracy, which is 99.78% on the gaussian kernel.

Keywords: Classification; Hypertension; K-Fold Cross Validation; Kernel; Support Vector Machine.

1. PENDAHULUAN

Hipertensi adalah keadaan seseorang mengalami peningkatan tekanan darah di atas nilai normal yang menyebabkan rasa sakit hingga kematian. Tekanan darah normal manusia adalah 120/80 mmHg [1]. Hipertensi disebut *the silent killer* atau pembunuh gelap, karena beberapa penderita tidak menunjukkan gejala hipertensi [2]. Permasalahan hipertensi selalu mengalami peningkatan di Indonesia. Hasil Riset Kesehatan Dasar (Riskesdas) tahun 2018 [3] menyatakan bahwa *prevalensi* penderita hipertensi adalah 34.1%. *Prevalensi* penyakit hipertensi dari tahun 2007 sampai tahun 2018 menyatakan bahwa penderita hipertensi di Provinsi Lampung selalu mengalami peningkatan. Data dari Dinas Kesehatan Kota Bandar Lampung dari tahun 2015 sampai tahun 2018 mengemukakan bahwa penyakit hipertensi adalah penyakit yang selalu memasuki peringkat sepuluh penyakit terbanyak di Kota Bandar Lampung.

Penelitian Samant et al [4] dilakukan dengan 981 data dan 13 atribut yang mengklasifikasikan empat *class*. *Kernel* yang digunakan adalah *kernel* linear, *quadratic*, *polyorder (order three)*, *multi-layer perceptron* (MLP), dan *radial basis function* (RBF). Penelitian tersebut menunjukkan bahwa kernel linier pada klasifikasi ketiga adalah *kernel* yang terbaik dengan akurasi sebesar 88.4%.

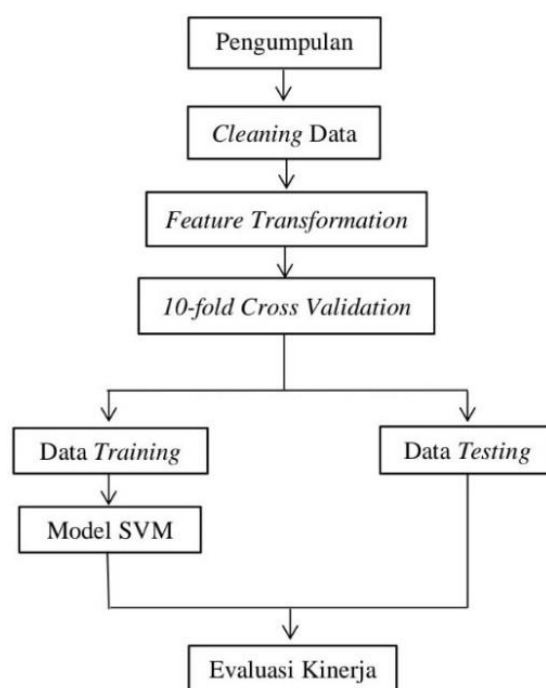
Yudha et al [5] melakukan penelitian klasifikasi pada empat klasifikasi risiko hipertensi. Klasifikasi risiko hipertensi tersebut adalah normal, prahipertensi, hipertensi derajat 1, dan hipertensi derajat 2. *Data set* yang digunakan pada penelitian tersebut berjumlah 125 data sekunder dengan 15 fitur yang berasal dari hasil rekam medis pasien. Penelitian tersebut memiliki tingkat akurasi terbaik sebesar 88% dengan menggunakan nilai $K=10$, $E=4$, dan data latih sebanyak 100 data.

Penelitian yang dilakukan oleh Andriansyah et al [6] menghasilkan akurasi terbaik sebesar 80% yang berasal dari 20 data uji dengan 30 data latih menggunakan nilai FCT dan LDT. Penggunaan atribut pada penelitian tersebut sebanyak 12 atribut. Penelitian tersebut menunjukkan bahwa aturan terbaik berasal dari 30 data latih pada 11 aturan dengan nilai FCT 70% dan LDT 3%.

Diagnosis penyakit hipertensi dilakukan secara manual selama ini. Pasien akan diperiksa oleh tenaga medis secara rutin. Pasien diperiksa dengan menggunakan stetoskop dan tensimeter [2]. Hal ini akan membutuhkan tenaga, materi, dan waktu yang banyak. Berdasarkan hal tersebut, terdapat sebuah gagasan untuk menerapkan bidang ilmu analisis *biomedical* data pada klasifikasi penyakit hipertensi dengan menggunakan metode *support vector machine* (SVM) di Kota Bandar Lampung. Berdasarkan hasil penelitian Osisanwo et al [7] dikemukakan bahwa metode *support vector machine* adalah algoritma presisi dan paling akurat pada klasifikasi. *Support vector machine* adalah metode untuk menemukan fungsi pemisah (*classifier*) yang optimal dari dua kelas yang berbeda. Kelebihan SVM adalah proses komputasi yang cepat karena menggunakan *support vector* [8].

2. METODOLOGI PENELITIAN

Penelitian klasifikasi penyakit hipertensi di Kota Bandar Lampung dengan metode *support vector machine* (svm) dijelaskan pada Gambar 1.



Gambar 1. Alur penelitian klasifikasi hipertensi

Berikut ini adalah tahapan yang dilakukan pada penelitian ini:

2.1. Pengumpulan Data

Penelitian ini menggunakan data penyakit hipertensi di Kota Bandar Lampung tahun 2017 sampai dengan tahun 2019. Data yang digunakan pada penelitian berjumlah 919 data. Data tersebut terdiri dari 609 data negatif hipertensi dan 310 data positif hipertensi. Terdapat 11 atribut pada penelitian ini. Pada Tabel 1 dikemukakan bahwa data penyakit hipertensi meliputi data numerik dan data kategorik.

Tabel 1. *Data set* penelitian klasifikasi hipertensi

No.	Atribut	Tipe Data
1	Usia	Numerik, (18-68; 38.19)
2	Sistolik	Numerik, (90-190; 124.31), mmHg
3	Diastolik	Numerik, (60-130; 78.92), mmHg
4	Tinggi Badan	Numerik, (145-175; 159.13), centimeter
5	Berat Badan	Numerik, (47-87; 62.87), centimeter
6	Lingkar Perut	Numerik, (71-111; 84.63), centimeter
7	Merokok	Kategorik, (Ya, Tidak)
8	Obesitas	Kategorik, (Ya, Tidak)
9	Kurang Makan Sayur Buah	Kategorik, (Ya, Tidak)
10	Konsumsi Alkohol	Kategorik, (Ya, Tidak)
11	Kurang Aktivitas	Kategorik, (Ya, Tidak)

2.2. *Cleaning Data*

Proses *cleaning* data adalah tahap pengurangan atau penghapusan data yang tidak sesuai. Data tersebut adalah data penyakit hipertensi yang kosong, sehingga lebih mudah dalam pengolahan data klasifikasi penyakit hipertensi. Pada penelitian ini dihapus sebanyak 84 data hipertensi yang kosong dari total 1003 data hipertensi yang dikumpulkan dari puskesmas di Kota Bandar Lampung.

2.3. *Feature Transformation*

Pada proses ini dilakukan proses *feature transformation* untuk mengubah data nominal menjadi data numerik. Proses *feature transformation* dilakukan dengan memberi nilai 1 (satu) jika bernilai benar dan nilai 0 (nol) jika bernilai salah [9].

2.4. *K-Fold Cross Validation*

Pada tahap ini dilakukan proses *data training* dan *data testing*. Pada uji validitas dilakukan *10-fold cross validation* dengan membagi sebanyak 10 lipatan. Pada *data training*, dilakukan proses *k-fold cross validation* sebesar 90% dari data. *Data testing* menggunakan sebesar 10% dari data yang tersedia. Pada penelitian ini, 827 data digunakan untuk proses *data training* dan 92 data digunakan untuk proses *data testing*. Proses akan diulangi sebanyak k lipatan (k berjumlah 10) dan rata-rata nilai akan diproses untuk ke tahap selanjutnya.

2.5. *Model SVM*

Tahap yang dilakukan pada proses ini adalah menentukan model dengan memilih tipe *kernel* yang tepat. Nilai yang terpilih akan dilakukan proses klasifikasi dengan menggunakan metode *support vector machine*. Pada penelitian ini, digunakan *kernel* linear, *polynomial*, dan *gaussian*.

2.6. *Testing / Pengujian*

Tahap terakhir pada penelitian ini adalah melakukan evaluasi kinerja. Model *support vector machine* yang tersedia pada tahap sebelumnya akan dilakukan proses evaluasi. Evaluasi kinerja akan mendapatkan nilai tipe *kernel* dan nilai parameter yang baik pada klasifikasi. Pengukuran parameter pada penelitian ini dilakukan dengan menggunakan *confusion matrix* yang terdiri dari parameter *accuracy*, *recall*, dan *precision*. Kemudian, dilakukan perhitungan rata-rata *confusion matrix* untuk mendapatkan hasil kinerja klasifikasi kejadian

hipertensi dengan metode *support vector machine* (SVM) menggunakan data puskesmas di Kota Bandar Lampung.

3. HASIL DAN PEMBAHASAN

3.1. Feature Transformation

Data atribut yang dilakukan proses *feature transformation* adalah merokok, kurang aktivitas fisik, kurang konsumsi sayur dan buah, konsumsi alkohol, dan obesitas. Kelas hipertensi dilakukan proses *feature transformation*.

3.2. Analisis Korelasi Fitur

Analisis korelasi fitur berfungsi untuk mengetahui hubungan antar dua variabel. Proses analisis korelasi menghasilkan nilai yang disebut koefisien korelasi. Nilai koefisien korelasi menggunakan data penelitian ini bernilai positif dan negatif. Nilai positif (terdapat korelasi) memiliki keeratan hubungan antara variabel yang secara teratur dengan arah yang sama pada kenaikan atau penurunan nilai variabel X yang diikuti oleh kenaikan atau penurunan nilai variabel Y. Nilai negatif (tidak ada korelasi) memiliki keeratan hubungan antara variabel secara teratur dengan arah yang berlawanan pada kenaikan atau penurunan nilai variabel X yang diikuti variabel Y akan mengalami sebaliknya. Semakin besar nilai koefisien korelasi mendekati +1 (positif 1) berarti pasangan data variabel X dan Y mempunyai korelasi positif yang sangat kuat (kedua variabel terdapat hubungan) dan jika bernilai 0 (nol), maka kedua variabel tidak terdapat hubungan [10].

Tabel 2. (a) Hasil analisis korelasi fitur

	Usia	Merokok	Kurang.Aktivitas .Fisik	Kurang.Konsumsi. Sayur.dan.Buah	Konsumsi. Alkohol
Usia	1.00	0.12	0.60	0.38	0.18
Merokok	0.12	1.00	0.08	-0.03	0.09
Kurang.Aktivitas.Fisik	0.60	0.08	1.00	0.67	0.25
Kurang.Konsumsi.Sayur.dan. Buah	0.38	-0.03	0.67	1.00	0.21
Konsumsi.Alkohol	0.18	0.09	0.25	0.21	1.00
Obesitas	0.33	-0.05	0.57	0.69	0.24
Sistol	0.63	0.14	0.73	0.54	0.17
Diastol	0.59	0.17	0.71	0.50	0.16
Tinggi.Badan	0.11	0.66	0.07	-0.10	0.07
Berat.Badan	0.29	0.40	0.44	0.41	0.20
Lingkar.Perut	0.21	0.34	0.33	0.29	0.16

Tabel 2. (b) Hasil analisis korelasi fitur

	Obesitas	Sistol	Diastol	Tinggi.Badan	Berat.Badan	Lingkar.Perut
Usia	0.3	0.63	0.59	0.11	0.29	0.21
Merokok	-0.05	0.14	0.17	0.66	0.40	0.34
Kurang.Aktivitas.Fisik	0.57	0.73	0.71	0.07	0.44	0.33

	Obesitas	Sistol	Diastol	Tinggi.Badan	Berat.Badan	Lingkar.Perut
Kurang.Konsumsi.Sayur.dan.Buah	0.69	0.54	0.50	-0.10	0.41	0.29
Usia	0.3	0.63	0.59	0.11	0.29	0.21
Merokok	-0.05	0.14	0.17	0.66	0.40	0.34
Kurang.Aktivitas.Fisik	0.57	0.73	0.71	0.07	0.44	0.33
Kurang.Konsumsi.Sayur.dan.Buah	0.69	0.54	0.50	-0.10	0.41	0.29
Konsumsi.Alkohol	0.24	0.17	0.16	0.07	0.20	0.16
Obesitas	1.00	0.39	0.33	-0.18	0.56	0.38
Sistol	0.39	1.00	0.90	0.13	0.36	0.27
Diastol	0.33	0.90	1.00	0.19	0.39	0.30
Tinggi.Badan	-0.18	0.13	0.19	1.00	0.53	0.38
Berat.Badan	0.56	0.36	0.39	0.53	1.00	0.66
Lingkar.Perut	0.38	0.27	0.30	0.38	0.66	1.00

3.3. Percobaan Model Klasifikasi

Pada penelitian ini dilakukan percobaan dengan model klasifikasi *kernel* linear, *gaussian*, dan *polynomial*. Proses pengelompokan data dilakukan berdasarkan kelas. Terdapat 609 data pada kelas negatif hipertensi dan 310 data pada kelas positif hipertensi. Tahap berikutnya adalah membagi *dataset* menjadi dua bagian, yaitu *data training* dan *data testing*. Proses ini menghasilkan *data training* sebanyak 827 data dan *data testing* sebanyak 92 data. Selanjutnya, melakukan proses *10-fold cross validation* dan membuat model klasifikasi menggunakan tiga *kernel*.

3.3.1 Kernel Linear

Percobaan pertama menggunakan *kernel* linear dengan hasil seperti pada Tabel 3.

Tabel 3. Hasil prediksi model SVM klasifikasi menggunakan *Kernel* Linear

<i>Fold</i>	Pengujian		
	<i>Accuracy</i>	<i>Recall</i>	<i>Precision</i>
1	100	100	100
2	100	100	100
3	100	100	100
4	100	100	100
5	98.91	100	98.39
6	98.91	98.36	100
7	100	100	100
8	98.91	100	98.39
9	100	100	100
10	100	100	100
Rata-rata	99.67	99.83	99.67

Pada percobaan menggunakan menggunakan *kernel* linear, didapatkan hasil rata-rata *accuracy* sebesar 99.67%, hasil rata-rata *recall* sebesar 99.83%, dan hasil rata-rata *precision* sebesar 99.67%

3.3.2 Kernel Gaussian

Percobaan menggunakan *kernel gaussian* dengan membandingkan hasil yang diperoleh dari setiap percobaan. Berikut hasil dengan menggunakan *kernel gaussian* yang dapat dilihat pada Tabel 4.

Tabel 4. Hasil prediksi model SVM klasifikasi menggunakan *Kernel Gaussian*

<i>Fold</i>	Pengujian		
	<i>Accuracy</i>	<i>Recall</i>	<i>Precision</i>
1	100	100	100
2	100	100	100
3	100	100	100
4	100	100	100
5	100	100	100
6	98.91	98.36	100
7	100	100	100
8	98.91	100	98.39
9	100	100	100
10	100	100	100
Rata-rata	99.78	99.83	99.83

Pada Tabel 4, *kernel gaussian* didapatkan hasil rata-rata *accuracy* sebesar 99.83%, hasil rata-rata *recall* sebesar 99.83%, dan hasil rata-rata *precision* sebesar 99.83%.

3.3.3 Kernel Polynomial

Percobaan terakhir menggunakan *kernel polynomial*. Percobaan tersebut mendapatkan hasil *confusion matrix* seperti pada Tabel 5.

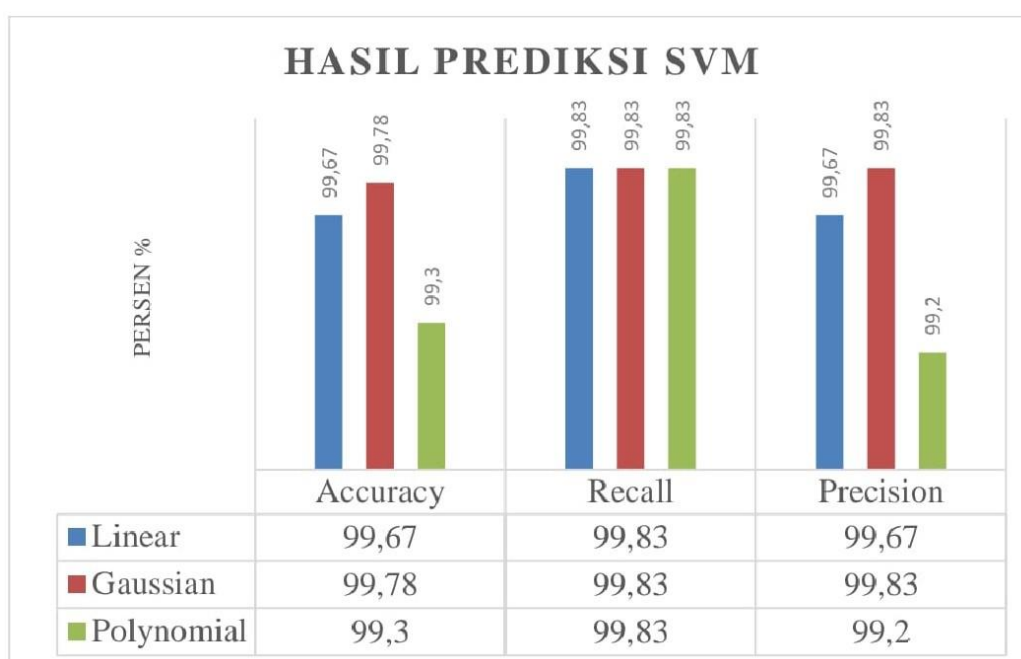
Tabel 5. Hasil prediksi model SVM klasifikasi menggunakan *Kernel Polynomial*

<i>Fold</i>	Pengujian		
	<i>Accuracy</i>	<i>Recall</i>	<i>Precision</i>
1	100	100	100
2	100	100	100
3	100	100	100
4	97.83	100	96.83
5	97.83	100	96.83
6	98.91	98.36	100
7	100	100	100

<i>Fold</i>	Pengujian		
	<i>Accuracy</i>	<i>Recall</i>	<i>Accuracy</i>
8	100	100	100
9	98.91	100	98.39
10	100	100	100
Rata-rata	99.30	99.83	99.20

Percobaan menggunakan *kernel polynomial*, didapatkan hasil rata-rata *accuracy* sebesar 99.30%, hasil rata-rata *recall* sebesar 99.83%, dan hasil rata-rata *precision* sebesar 99.20%.

3.4. Perbandingan *Kernel Linear, Gaussian, dan Polynomial*



Gambar 2. Grafik hasil prediksi menggunakan *kernel linear, gaussian, dan polynomial*

Berdasarkan Gambar 2, hasil rata-rata *accuracy* terbesar pada *kernel gaussian* sebesar 99.78%, dengan hasil rata-rata *recall* sebesar 99.83%, dan hasil rata-rata *precision* sebesar 99.83%. Selanjutnya didapatkan hasil rata-rata *accuracy* terbesar kedua setelah yaitu pada *kernel linear* sebesar 99.67%, dengan hasil rata-rata *recall* sebesar 99.83%, dan hasil rata-rata *precision* sebesar 99.67%. Hasil rata-rata *accuracy* terendah didapatkan oleh *kernel polynomial* sebesar 99.30%, dengan hasil rata-rata *recall* sebesar 99.83%, dan hasil rata-rata *precision* sebesar 99.20%. Nilai *recall* pada *kernel gaussian* memiliki nilai yang besar karena terdapat kesalahan sistem dalam memprediksi bahwa sampel benar hipertensi tetapi pada kenyataannya sampel tidak terkena hipertensi. Nilai *false positive*.

4. KESIMPULAN

Penelitian memperoleh hasil klasifikasi dengan menggunakan metode *support vector machine (SVM)* dengan teknik penentuan parameter terbaik menggunakan *10-fold cross validation*. Proses pembagian data dilakukan dengan aturan 90% *data training* dan 10% *data testing*. Terdapat 919 data pada penelitian ini dengan pembagian *data training* sebanyak 827 data dan *data testing* sebanyak 92 data. Keseluruhan percobaan hasil klasifikasi menggunakan tiga *kernel* didapatkan *accuracy* terbesar pada *kernel gaussian* yaitu sebesar 99.78%,

recall tidak memiliki nilai terbesar pada ketiga *kernel* dikarenakan memiliki nilai yang sama yaitu sebesar 99.83%, dan *precision* terbesar pada *kernel gaussian* yaitu sebesar 99.83%.

DAFTAR PUSTAKA

- [1] Herwati & W. Sartika, "Terkontrolnya Tekanan Darah Penderita Hipertensi Berdasarkan Pola Diet dan Kebiasaan Olah Raga di Padang Tahun 2011," *Jurnal Kesehatan Masyarakat*, vol. 8, no. 1, pp. 8-14, 2014.
- [2] Y. Susilo & A. Wulandari, *Cara Jitu Mengatasi Hipertensi*, Yogyakarta: ANDI, 2011.
- [3] Kementerian Kesehatan Republik Indonesia, "Hasil utama Riset Kesehatan Dasar (Riskesdas)," Badan Litbangkes Kementerian Kesehatan RI, Jakarta Pusat, 2018.
- [4] R. Samant & S. Rao, "A study on Comparative Performance of SVM Classifier Models with Kernel Functions in Prediction of Hypertension," (*IJCSIT*) *International Journal of Computer Science and Information Technologies*, vol. 4, no. 6, pp. 818-821, 2013.
- [5] B. L. Yudha, L. Muflikhah & R. C. Wihandika, "Klasifikasi Risiko Hipertensi Menggunakan Metode Neighbor Weighted K-Nearest Neighbor (NWKNN)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 2, pp. 897-904, 2018.
- [6] M. R. Andriansyah, E. Santoso & Sutrisno, "Klasifikasi Risiko Hipertensi Menggunakan Fuzzy Decision Tree Iterative Dichotomiser 3 (ID3)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 12, pp. 7088-7096, 2018.
- [7] F. Y. Osisanwo, J. E. T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi & J. Akinjobi, "Supervised Machine Learning Algorithms: Classification and Comparison," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 48, no. 3, pp. 128-138, 2017.
- [8] N. H. Farhat, "Photonic Neural Networks and Learning Machines," *IEEE Expert*, vol. 7, no. 5, pp. 63-72, 1992.
- [9] C. Vielhauer & R. Steinmetz, "Handwriting: Feature Correlation Analysis for Biometric Hashes," *EURASIP Journal on Advances in Signal Processing*, vol. 2004, no. 4, pp. 542-558, 2004.
- [10] H. Liu & H. Motoda, "Feature Transformation and Subset Selection," *IEEE Expert*, vol. 13, no. 2, pp. 26-28, 1998.