

INTISARI

SELEKSI FITUR MENGGUNAKAN ALGORITMA GENETIKA UNTUK KLASIFIKASI PENYAKIT ANEMIA DENGAN SVM

Oleh

Affifah Mutiara Pertiwi
20/466391/PPA/05957

Hasil Riskesdas Kementerian Kesehatan Republik Indonesia menyampaikan bahwa pada tahun 2018 prevalensi kasus anemia di kalangan remaja Indonesia sebesar 32%. Persentase tersebut sangat besar sehingga diperlukan perhatian khusus dengan diagnosis dini pada penyakit anemia. Data pemeriksaan darah rutin mungkin untuk digunakan sebagai solusi skrining dini anemia karena dapat dilakukan di laboratorium patologi klinik yang mudah ditemui di sekitar masyarakat. Terdapat banyak jumlah parameter hasil pemeriksaan darah rutin. Namun fitur-fitur tersebut kurang signifikan dalam menunjukkan diagnosis anemia tertentu. Oleh karena itu diperlukan suatu seleksi fitur untuk mendapatkan fitur-fitur yang memiliki pengaruh signifikan saat digunakan dalam pengklasifikasian penyakit anemia.

Pada penelitian ini dilakukan seleksi fitur menggunakan algoritma genetika dan *support vector machine* (SVM) sebagai *classifier*. Hasil pemeriksaan darah rutin penyakit anemia diklasifikasikan dalam beberapa kelas, yaitu DB, β -TT, HbE, dan *mix*. Fitur-fitur yang diseleksi adalah indeks-indeks hematologi pada hasil pemeriksaan darah rutin lengkap menggunakan *hematology analyzer*. Data tersebut telah diolah dan diekstraksi oleh Dr. dr. Tri Ratnaningsih, M.Kes., Sp.PK(K) dari pemeriksaan darah di Rumah Sakit Umum Pusat (RSUP) Dr. Sardjito.

Seleksi fitur dilakukan dalam 3 eksperimen dengan nilai P_c berbeda. Proses seleksi fitur diulangi sebanyak 10 percobaan. Hasil eksperimen terbaik *fitness* akurasi adalah eksperimen 3 pada percobaan 9 dengan akurasi 86,2%. Sedangkan *fitness F1-score* hasil terbaik adalah eksperimen 1 pada percobaan 7 dengan skor 81,78%. Implementasi seleksi fitur berhasil meningkatkan akurasi sebesar 4,4% dan sama skornya dengan hasil *F1-score* dibandingkan dengan klasifikasi tanpa seleksi fitur sebesar 81,78%. Terpilih 9 fitur berdasarkan *fitness* akurasi yaitu RBC, MCV, CHCM, CH, Retic %, CHr, RBC % Hypo, RBC % Micro dan RBC Fragment. Berdasarkan *fitness F1-score* terpilih 10 fitur yaitu RBC, HGB, MCV, MCH, MCHC, RDW, Retic %, RBC % Hypo, RBC % Micro dan RBC Fragment.

Kata Kunci: Algoritma Genetika, Anemia, Seleksi Fitur, *Support Vector Machine*

ABSTRACT

FEATURE SELECTION USING GENETIC ALGORITHM FOR CLASSIFICATION OF ANEMIA DISEASES WITH SVM

by

Affifah Mutiara Pertiwi
20/466391/PPA/05957

The results of Riskesdas Ministry of Health of the Republic of Indonesia stated that in 2018 the prevalence of anemia cases among Indonesian teenagers was 32%. This percentage is very large and special attention is needed with early diagnosis of anemia. Routine blood examination data may be used as an early screening solution for anemia, which can be done in clinical pathology laboratories which are easily found in the community. There are many parameters for routine blood test results. However, these features are less significant in indicating specific diagnosis of anemia. Therefore, feature selection is needed to obtain features that have significant influence when used in classifying anemia.

In this research, feature selection was carried out using genetic algorithm and support vector machine (SVM) as classifier. The results of routine blood tests for anemia are classified into several classes DB, β -TT, HbE, and mix. The features selected are hematological indices in the results of complete routine blood examination using hematology analyzer. The data has been processed and extracted by Dr. dr. Tri Ratnaningsih, M.Kes., Sp.PK(K) from blood test at the Central General Hospital (RSUP) Dr. Sardjito.

Feature selection was carried out in 3 experiments with different P_c values. The feature selection process was repeated for 10 repetitions. The best experimental results for fitness accuracy were experiment 3 in repetition 9 with accuracy score 86.2%. Meanwhile, best fitness F1-score result was experiment 1 in repetition 7 with score 81.78%. Implementation of feature selection succeeded in increasing accuracy by 4.4% and the score was same as F1-score compared to classification without feature selection of 81.78%. Nine features were selected based on fitness accuracy RBC, MCV, CHCM, CH, Retic %, CHr, RBC % Hypo, RBC % Micro and RBC Fragment. Based on fitness F1-score, 10 features were selected, RBC, HGB, MCV, MCH, MCHC, RDW, Retic %, RBC % Hypo, RBC % Micro and RBC Fragment.

Key words: Genetic Algorithm, Anemia, Feature Selection, Support Vector Machine