

INTISARI

KLASIFIKASI TALI PUSAT JANIN BERDASARKAN CITRA ULTRASONOGRAPHY DOPPLER DUA DIMENSI MENGGUNAKAN *MODIFIED SMOTE DAN ENSEMBLE MULTICLASSIFIER*

Oleh

Gede Angga Pradipta

18/435404/SPA/00642

Tali pusat mempunyai fungsi sebagai organ untuk aliran oksigen dan makanan bagi janin yang diperoleh dari ibu. Tali pusat memiliki tiga jenis kategori berdasarkan ciri lilitan atau *coiling* yaitu: *hypercoiling*, *normocoiling*, dan *hypocoiling*. Permasalahan pada penelitian ini adalah adanya kemiripan bentuk antara kelas *normocoiling* dan *hypercoiling* sehingga diperlukan proses ekstraksi fitur yang mampu menghasilkan nilai fitur yang relevan terhadap karakteristik dari ketiga kelas ini. Selain itu dataset memiliki kondisi *imbalanced* yaitu dimana jumlah data pada setiap kelas dalam *dataset* memiliki jumlah dengan rasio yang jauh berbeda dimana kelas *hypercoiling* berjumlah 9 citra, normal *coiling* berjumlah 108 citra, dan *hypocoiling* berjumlah 34. *Imbalanced data sets* umumnya menyebabkan data pada kelas minoritas sulit untuk dikenali dan salah diklasifikasikan oleh *classifier*. Pada penelitian ini mengembangkan sebuah model klasifikasi yang mampu menghadapi kondisi *imbalanced data sets* dengan optimasi baik dari sisi data level dan juga optimasi pada metode klasifikasi yang digunakan. Model memadukan modifikasi *Synthetic Minority Oversampling Technique* (SMOTE) dan Metode *Ensemble Multiclassifier Voting*. Pada tahap pertama, proses segmentasi objek tali pusat janin dilakukan dengan memisahkannya dari *background*. Kemudian dari *region of interest* (ROI) tali pusat dilakukan ekstraksi fitur dengan pendekatan tekstur dan bentuk. Fitur tekstur menggunakan metode *Graylevel Co-Occurrence Matrix* (GLCM) pada dua tipe citra yaitu *grayscale* dan (RGB). Fitur bentuk menggunakan pendekatan dari literatur kedokteran yaitu *Umbilical coiling Index* (UCI). Proses berikutnya adalah melakukan seleksi terhadap fitur dengan metode *Information Gain* yang bertujuan untuk mendapatkan fitur-fitur yang relevan memberikan kontribusi terhadap hasil klasifikasi. Tahapan berikutnya adalah proses *oversampling* dengan metode modifikasi SMOTE untuk pembuatan data sintetik pada kelas minoritas yaitu kelas *hypercoiling*. Pengujian dilakukan pada dataset sebelum dan sesudah seleksi fitur, perbandingan metode *single classifier* dan *ensemble multiclassifier*, dan pengujian modifikasi SMOTE dengan beberapa metode *state-of-the-art* SMOTE lainnya. Berdasarkan pada hasil percobaan, model terbaik adalah dengan kombinasi modifikasi SMOTE (Radius-SMOTE) dan *Ensemble multiclassifier* dengan akurasi 97.66%, presisi 98.33%, *recall* 97.66%, dan *F1 measure* 98.33%. Terdapat peningkatan akurasi 3.1% akurasi, 5% presisi, 3% *recall*, dan 5.3% *F1 measure* saat dibandingkan dengan penggunaan metode SMOTE original.

Kata kunci : *Umbilical cord, Oversampling, Imbalanced data sets, Ensemble multiclassifier, SMOTE*

ABSTRACT

Classification of Umbilical Cord Based On 2-D Ultrasonography Doppler Images Using Modified Smote and Ensemble *Multiclassifier* Voting

by

Gede Angga Pradipta

18/435404/SPA/00642

The fetal umbilical cord is an organ that supplies oxygen and nutrients to the fetus. The umbilical cord has three categories based on coiling characteristics: hypercoiling, normocoiling, and hypocoiling. The problem in this study is the similarity in shape between the normocoiling and hypercoiling classes so that a feature extraction process is needed to produce feature values relevant to the characteristics of these three classes. In addition, the dataset has an imbalanced condition, where the amount of data in each class has a relatively high number ratio where the hypercoiling class is 9 images, 108 normal coiling images, and 34 hypocoiling. Imbalanced data sets generally causes data in minority classes to be challenging to recognize and misclassified by the classifier. This research proposes a classification model that can deal with imbalanced data conditions by optimizing both the data level and the classification method's optimization. The proposed model combines Synthetic Minority Oversampling Technique (SMOTE) and with Ensemble multiclassifier voting method. In the first stage, the fetal umbilical cord object's segmentation process is carried out by separating it from the background object. Then feature extraction is performed on the ROI of the fetal umbilical cord with a texture and shape approach. Texture features use the GLCM method on two types of images, namely grayscale and RGB. The shape features use a knowledge method from the medical literature, namely the Umbilical Coiling Index (UCI). The next process is selecting features using the Information Gain method, which aims to reduce the data dimensions and get the relevant features that contribute to the classifier's performance. The next step is to carry out the dataset's oversampling process in the minority class (Hypercoiling) with the SMOTE modification method for creating new synthetic data.

Evaluation experiments were carried out on datasets before and after feature selection, comparison of single classifier and multiclassifier methods, and the performance of SMOTE modifications with several other state-of-the-art SMOTE methods. Based on experiments and testing for umbilical cord dataset, the best model is proposed modified SMOTE (Radius-SMOTE) with Ensemble multiclassifier voting yield accuracy of 97.66%, precision 98.33%, recall 97.66%, and F1 measure 98.33%. There is a 3.1% increase in accuracy, 5% precision, 3% recall, and 5.3% F1 measure compared to using the original SMOTE method.

Kata kunci : Umbilical cord, Oversampling, Imbalanced data sets, Ensemble multiclassifier, SMOTE