

INTISARI

Penanganan Data Tidak Seimbang Berbasis Pembelajaran *Cost-Sensitive* Pada Klasifikasi Nodul Tiroid

Oleh

Akhmad Faizal

23/529787/PPA/06701

Tiroid adalah kelenjar endokrin krusial yang mengatur metabolisme dan perkembangan tubuh. Seringkali, kelenjar ini mengembangkan nodul tiroid, yaitu pertumbuhan abnormal yang sangat umum terjadi dan memerlukan perhatian klinis untuk membedakan sifat jinak atau ganasnya. Diagnosis nodul tiroid melalui ultrasonografi (US) masih menghadapi tantangan berupa rendahnya spesifisitas dan bias akibat distribusi data yang tidak seimbang, di mana kasus ganas hanya mewakili sebagian kecil dari populasi. Penelitian ini bertujuan mengembangkan sistem berbasis pembelajaran mesin yang mampu mengatasi permasalahan tersebut melalui segmentasi dengan U-Net dan klasifikasi menggunakan SVM serta XGBoost dengan pendekatan *cost-sensitive learning* (CSL). Model U-Net yang digunakan untuk tahap segmentasi mencapai nilai *Intersection over Union* (IoU) sebesar **79,82%**, menunjukkan kemampuan yang baik dalam mengekstraksi area nodul secara akurat. Hasil evaluasi menunjukkan bahwa SVM tanpa CSL gagal melakukan klasifikasi secara efektif dengan sensitivitas **100%** namun spesifisitas dan NPV sebesar **0%**. Penerapan CSL memperbaiki kemampuan diskriminasi model. SVM + CSL Default menghasilkan spesifisitas **79,1%** dan sensitivitas **33%**, sementara SVM + CSL Bobot Random menghasilkan spesifisitas **65,7%** dan sensitivitas 46,1%. Di sisi lain, XGBoost menunjukkan performa sangat unggul dengan sensitivitas 98,2%, PPV 97,8%, dan AUC **85,3%**. Meskipun model Cine-CNNTrans memiliki AUC tertinggi (**88%**), XGBoost memberikan keseimbangan metrik yang lebih optimal untuk deteksi kasus ganas. Temuan ini menegaskan bahwa kombinasi U-Net dan XGBoost dengan CSL lebih relevan secara klinis dalam menangani ketidakseimbangan data pada klasifikasi risiko nodul tiroid.

Kata Kunci: Ultrasound, SVM, XGBoost, Cost-sensitive Learning, Klasifikasi Tiroid

ABSTRACT

Unbalanced Data Handling Based On Cost-Sensitive Learning In Thyroid Nodule Classification

Oleh

Akhmad Faizal

23/529787/PPA/06701

The thyroid is a crucial endocrine gland that regulates metabolism and body development. Often, this gland develops thyroid nodules, which are very common abnormal growths that require clinical attention to distinguish whether they are benign or malignant. Thyroid nodule diagnosis using ultrasonography (US) still faces challenges due to its low specificity and bias arising from imbalanced data distribution, where malignant cases represent only a small fraction of the population. This study aims to develop a machine learning-based system to address these challenges by employing U-Net for segmentation and SVM as well as XGBoost with a cost-sensitive learning (CSL) approach for classification. The U-Net segmentation model achieved an Intersection over Union (IoU) of **79.82%**, demonstrating good capability in accurately extracting nodule regions. Evaluation results show that SVM without CSL failed to classify effectively, with **100%** sensitivity but **0%** specificity and NPV. The application of CSL improved the model's discriminative ability; SVM + CSL Default yielded **79.1%** specificity and **33%** sensitivity, while SVM + CSL Random Weights produced **65.7%** specificity and **46.1%** sensitivity. On the other hand, XGBoost demonstrated superior performance with **98.2%** sensitivity, **97.8%** PPV, and an AUC of **85.3%**. Although the Cine-CNNTrans model achieved the highest AUC (88%), XGBoost provided a more optimal balance of metrics for detecting malignant cases. These findings confirm that the combination of U-Net and XGBoost with CSL is more efficient and clinically relevant in handling data imbalance for thyroid nodule risk classification.

Kata Kunci: Ultrasound, SVM, XGBoost, Cost-sensitive Learning, Thyroid Classification