

PERFORMA RANDOM FOREST GROUP UNTUK KLASIFIKASI PENYAKIT BUSUK PANGKAL BATANG YANG DISEBABKAN OLEH *Ganoderma boninense* PADA PERKEBUNAN KELAPA SAWIT

PERFORMANCE OF RANDOM FOREST GROUP FOR BASAL STEM ROT DISEASE CLASSIFICATION CAUSED BY *Ganoderma boninense* IN OIL PALM PLANTATION

Heri Santoso

Abstrak Pemanfaatan teknik penginderaan jauh sudah banyak dimanfaatkan untuk kajian penyakit busuk pangkal batang (BPB) di perkebunan kelapa sawit. Performa klasifikasi dapat ditingkatkan dengan memanfaatkan *machine learning algorithms*. Identifikasi sejak awal serangan penyakit BPB merupakan kunci dalam penanganan secara kultur teknis yang ditujukan untuk memperpanjang umur produksi tanaman. Pemilihan model klasifikasi yang memanfaatkan *machine learning algorithms* akan mempengaruhi akurasi dan waktu proses klasifikasi. *Random forest* (RF) dari hasil beberapa penelitian menunjukkan model klasifikasi dengan performa akurasi yang tinggi dalam klasifikasi penyakit BPB ini. Model klasifikasi RF saat ini memiliki 15 varian yang dikembangkan oleh para peneliti. Penelitian ini bertujuan untuk membandingkan akurasi dan kecepatan *fitting model* dari 15 metode klasifikasi *random forest group* dalam mengidentifikasi penyakit BPB di perkebunan kelapa sawit. penelitian ini menggunakan analisis deskriptif berdasarkan hasil akurasi dan kecepatan *fitting model* dari 15 metode klasifikasi. Hasil penelitian ini menunjukkan model klasifikasi *oblique random forest* (ORF) dengan metode *partial least squares* (PLS) mempunyai akurasi dan nilai Kappa terbaik (0,914 dan 0,815) dan model *random ferns* (Rferns) sebagai model klasifikasi dengan akurasi dan nilai Kappa yang terendah (0,657

dan 0,334). Model RF *rule-based model* (rlb) memerlukan waktu terlama dalam *fitting model* yaitu 12.993,23 detik dan model *parallel random forest* (PRF) memerlukan waktu tersingkat yaitu 13,54 detik. Pencapaian akurasi dan waktu yang diperlukan untuk *fitting model* menjadi pertimbangan dalam pemilihan model klasifikasi untuk tanaman sehat dan tanaman terinfeksi penyakit BPB karena jamur *Ganoderma*.

Kata kunci: busuk pangkal batang, kelapa sawit, PLS, PRF

Abstract Remote sensing technologies have been used for monitoring basal stem rot (BSR) diseases in oil palm plantations. The classification performance could be improved with machine learning algorithms. The success key of BSR treatment in culture technic treatments to prolong oil palm productivity ages is early detection of BSR disease. The selection of classification models of machine learning algorithms will be affecting accuracy and classification processes time. Random forest (RF) of model classification based on several previous research has a high performance of accuracy for BSR disease classification in the oil palm plantation. Today, there were fifteen models of the variance of the random forest algorithm. The research goals were to compare the accuracy and time consumed for the fitting model from 15 model's classification of RF group in case of identity and classify the BSR disease in the oil palm plantations. Descriptive analysis was adopted to explain the difference accuracy and time for the fitting model of fifteen model classification. The results showed the highest accuracy and Kappa value (0,914 and 0,815) was the oblique random forest (ORF) with partial least squares (PLS)

Penulis yang tidak disertai dengan catatan kaki instansi adalah peneliti pada Pusat Penelitian Kelapa Sawit

Heri Santoso (✉)
Pusat Penelitian Kelapa Sawit
Jl. Brigjen Katamso No. 51 Medan, Indonesia
Email: hs_jmp@yahoo.com

method and random ferns (Rferns) was the method with the lowest accuracy and Kappa value (0,657 dan 0,334). The highest time consumed for the fitting model was RF rule-based model (rlb) with 12.993,23 seconds and the shortest time was parallel random forest (PRF) with 13,54 seconds. The accuracy and Kappa values and time consumed for the fitting model as a consideration to selecting model classification of healthy and infected BSR in oil palm plantations.

Keywords: basal stem rot disease, oil palm, PLS, PRF

PENDAHULUAN

Machine learning algorithms sudah banyak dimanfaatkan dalam kajian penginderaan jauh baik yang memanfaatkan data dari hasil perekaman sistem aktif maupun pasif (Behmann, Mahlein, Rumpf, Römer, & Plümer, 2015; Crawford, Tuia, & Yang, 2013; Kim et al., 2014; Omer, Mutanga, Abdel-Rahman, & Adam, 2016; Santoso, Tani, Wang, Prasetyo, & Sonobe, 2018; Uhlmann, Kiranyaz, & Gabbouj, 2014; Wieland & Pittore, 2014). Teknik penginderaan jauh dan *machine learning algorithms* juga dimanfaatkan dalam identifikasi dan pemetaan sebaran penyakit busuk pangkal batang (BPB) di perkebunan kelapa sawit yang disebabkan oleh jamur *Ganoderma boninense* (Lelong et al., 2010; Liaghat et al., 2014; Santoso, Tani, & Wang, 2017; Santoso et al., 2018). Penyakit BPB ini masih menjadi penyakit utama di perkebunan kelapa sawit sampai saat ini belum ada penanganan yang efektif dalam mengontrol penyebaran penyakit tersebut. Identifikasi dan pemetaan yang cepat sangat diperlukan untuk mengetahui status tanaman kelapa sawit secara individu termasuk dalam status sehat atau sakit. Identifikasi status tanaman yang sakit sejak awal dapat digunakan dalam penanganan secara kultur teknis untuk memperpanjang umur tanaman (Hushiaran, Yusof, & Dutse, 2013; Priwiratama, Prasetyo, & Susanto, 2014).

Teknik penginderaan jauh dan *machine learning algorithms* dalam identifikasi dan pemetaan sebaran tanaman yang sehat dan terinfeksi penyakit BPB memberikan akurasi yang baik serta hasil pemetaan tanaman sakit dan terinfeksi penyakit BPB yang baik mengingat sebaran tanaman yang terinfeksi BPB berada di sekitar tanaman yang sudah mati karena infeksi penyakit BPB tersebut (Santoso et al., 2017, 2018). *Random forest* (RF) merupakan *machine learning algorithms* yang memberikan performa akurasi tinggi dalam identifikasi serta pemetaan tanaman yang sehat

dan terinfeksi penyakit BPB menurut Santoso et al., (2018, 2017). RF merupakan model klasifikasi hirarki yang dibangun dari banyak *decision tree* (DT) dari *ensemble* atau *learning algorithms* yang disurvei, setiap *tree* akan memberikan hasil klasifikasi (klas) dari variabel yang dimasukan sebagai input dan membentuk sebuah *forest*, dan *forest* akan memilih klas mayoritas dari seluruh *tree* yang ada di *forest* tersebut -(Breiman, 2001). Seiring dengan perkembangan pemanfaatan *machine learning* dalam analisis data, beberapa model klasifikasi yang pengembangannya didasarkan pada model klasifikasi RF telah dihasilkan oleh para peneliti seperti yang disampaikan oleh Kuhn (2019).

Beberapa metode klasifikasi yang didasarkan model klasifikasi RF menurut Kuhn (2019) meliputi *oblique random forest* (ORF) dengan metode *partial least squares* (PLS), *regularized random forest* (RRF) dengan metode RRF, *parallel random forest* (PRF) dengan metode parRF, *random forest by randomization* (RFr) dengan metode extraTrees, *weighted subspace random forest* (WSRF) dengan metode wsrF, *conditional inference random forest* (CIRF) dengan metode cforest, *regularized random forest* (RRFg) dengan metode RRFglobal, *oblique random forest* (ORFI) dengan metode ORFlog, *random forest ranger* (RFrg) dengan metode ranger, *random forest* (RFrb) dengan metode Rborist, *oblique random forest* (ORFrg) dengan metode ORFridge, *random forest* (RFrlb) dengan metode *rule-based model* (rfRules), *oblique random forest* (ORFsvm) dengan metode ORFsvm, dan *random ferns* (rFerns) dengan metode rFerns. Masing-masing model klasifikasi tersebut tentunya akan mempengaruhi waktu dalam *fitting model* dan akurasi klasifikasinya. Penelitian ini bertujuan untuk membandingkan akurasi dan kecepatan *fitting model* klasifikasi *random forest group* dalam mengidentifikasi penyakit BPB di perkebunan kelapa sawit.

BAHAN DAN METODE

Data dan Lokasi Penelitian

Data penelitian ini menggunakan data dari penelitian Santoso et al. (2017) dengan total data sebanyak 144 pohon sampel yang terdiri dari 99 tanaman sehat dan 45 tanaman yang diidentifikasi terserang penyakit BPB. Penelitian dilaksanakan di salah satu kebun di Sumatera Utara dengan obyek berupa tanaman kelapa sawit yang sehat dan terinfeksi penyakit BPB yang terbagi dalam 4 lokasi

Tabel 1. Lokasi penelitian
Table 1. Research locations

Lokasi penelitian	Umur (tahun)	Luas (ha)
Lokasi 1	16	150,16
Lokasi 2	21	139,56
Lokasi 3	15 dan 18	130,49
Lokasi 4	10	128,99

penelitian. Tabel 1 menunjukkan detail lokasi penelitian terkait dengan umur tanaman dan luas masing-masing lokasi penelitian. Citra satelit yang digunakan adalah citra QuickBird hasil perekaman pada 4 Agustus 2008 dengan 4 saluran (band) sebagai variabel yaitu band merah (630-690 nm), band hijau (520-600 nm), band biru (450-520 nm) dan band inframerah dekat (760-900 nm).

Model Klasifikasi Random Forest Group

Model klasifikasi RF didasarkan pada hasil *voting* dari *tree* yang terbentuk dengan akhir keputusan (*decision*) berdasarkan suara terbanyak. Proses pemecahan (*split*) diterapkan dengan membagi data berdasarkan jenis atribut yang digunakan. Dalam pembangunan *tree* dalam RF didasarkan pada metode pemilihan *split* vektor independen secara acak (*random split selection*) dan melakukan klasifikasi pada semua *tree* (Breiman, 2001). RF didasarkan pada banyak DT, sedangkan perbedaan tipe pemecahan data (*split*) pada DT meliputi *orthogonal*, *oblique*, dan *rotation* memberikan beberapa turunan model klasifikasi RF. Model klasifikasi tersebut adalah RF, ORF, dan *rotation random forest* (rotRF) dengan model pemecahan (*split*) data berupa *ridge regression*, *partial least squares* (PLS), *support vector machine* (SVM), dan lainnya (Breiman, 2001; Menze, Kelm, Splitthoff, Koethe, & Hamprecht, 2011; Poona, van Niekerk, & Ismail, 2016).

Perbandingan antar model klasifikasi ini dilakukan di perangkat lunak (*software*) R dengan *caret package*. Dalam proses *fitting model*, digunakan parameter *cross-validation* (CV) dengan jumlah iterasi (perulangan) *resampling* sebanyak 10 kali dan jumlah CV sebanyak 10 dan *preprocessing* data berupa *scale*

(masing-masing data yang dibagi dengan nilai standar deviasi) dan *center* (masing-masing data dikurangi nilai rata-rata seluruh data). Penerapan CV ini dimaksudkan untuk menghindari *overfitting* (Kuhn, 2008, 2016). Pembagian data untuk *training data* sebanyak 75% dan *testing data* sebanyak 25% (Liaghat et al., 2014). Model klasifikasi RF group yang digunakan dalam penelitian ini adalah adalah 15 model klasifikasi dengan perbedaan utama dalam tipe *split* dalam DT dan model pemecahan node (*the node splitting models*) dengan parameter masing-masing model klasifikasi (Tabel 2) sebagai berikut:

- Oblique random forest* (ORF) dengan metode ORFpls

Model klasifikasi RF dengan tipe *split* DT *oblique* dan *split* data dengan pendekatan PLS. Parameter *tuning* model berupa *mtry* yaitu jumlah dari predictor terpilih (variable). Model ini dijalankan dengan memanfaatkan *package obliqueRF* pada *package caret* di R (Kuhn, 2008, 2016).

- Random forest* (RF) dengan metode RF

Model klasifikasi RF dengan tipe *split* DT *orthogonal* dan *split* data dengan pendekatan random. Parameter *tuning* model berupa *mtry* dengan memanfaatkan *package random forest* pada *package caret* di R (Kuhn, 2008, 2016).

- Regularized random forest* (RRF) dengan metode RRF

Model klasifikasi RRF menekankan pada penanganan pada pengulangan variable yang telah digunakan dalam pembangunan *tree* dan tidak akan menggunakan variable tersebut pada pembangunan *tree* selanjutnya. Parameter *tuning*

model berupa *mtry*, *coefReg* yaitu koefisien dari *regularization* (semakin kecil koefisien akan menyebabkan subset fitur yang lebih kecil), dan *coefImp* (*importance coefficient*) di *setting TRUE*. Proses RRF ini memerlukan *package random forest* pada *package caret* di R (Kuhn, 2008, 2016). d. *Parallel random forest* (PRF) dengan metode parRF

Model klasifikasi PRF merupakan model klasifikasi RF yang dalam prosesnya dilakukan parallel dalam hal *split* data dan training model (Chen et al., 2016). Parameter *tuning* model berupa *mtry* dengan memanfaatkan *package e1071*, *random forest*, *foreach*, dan *import* pada *package caret* di R (Kuhn, 2008, 2016).

- e. *Random forest by randomization* (RFr) dengan metode extraTrees

Model ini merupakan model RF yang dilakukan dengan metode extraTrees. Dalam extraTrees tidak dilakukan resample dalam observasi ketika membangun *tree* dan tidak menggunakan *split* terbaik (Geurts, Ernst, & Wehenkel, 2006). Parameter *tuning* model berupa *mtry* dan *numRandomCuts* yaitu jumlah pemotongan acak untuk setiap fitur yang dipilih secara acak. Model ini memanfaatkan *package extraTrees* pada *package caret* di R (Kuhn, 2008, 2016).

- f. *Weighted subspace random forest* (WSRF) dengan metode wsrf

Model ini menggunakan metode pembobotan variabel untuk pemilihan sub-ruang variabel menggantikan metode sampling variabel secara acak (Zhao, Williams, & Huang, 2017). Parameter *tuning* model berupa *mtry* dengan memanfaatkan *package wsrf* pada *package caret* di R (Kuhn, 2008, 2016).

- g. *Conditional inference random forest* (CIRF) dengan metode cforest

Model cforest membangun sebuah *forest* dari *tree* inferensi bersyarat (*conditional inference trees*) dengan hanya subset variabel yang tersedia untuk dipecah di setiap *node* (Hothorn, Hornik, & Zeileis, 2006). Parameter *tuning* model berupa *mtry* dengan memanfaatkan *package party* pada *package caret* di R (Kuhn, 2008, 2016).

- h. *Regularized random forest* (RRFg) dengan metode

RRFglobal

Model ini seperti yang dijelaskan sebelumnya yaitu model RRF dan tidak ada perbedaan dengan RRFglobal pada tuning parameter. RRFglobal membangun RRF menggunakan paket honomim dengan parameter *mtry* dan *coefReg*. Parameter *tuning* model berupa *mtry* dan *coefReg* dengan memanfaatkan *package RRF* pada *package caret* di R (Kuhn, 2008, 2016).

- i. *Oblique random forest* (ORFI) dengan metode ORFlog

Model ini seperti yang dijelaskan sebelumnya dengan *split* data menggunakan *logistic regression*. Parameter *tuning* model berupa *mtry* dengan memanfaatkan *package obliqueRF* pada *package caret* di R (Kuhn, 2008, 2016).

- j. *Random forest ranger* (RFrg) dengan metode ranger

Model ranger merupakan salah satu cara yang cepat dalam implementasi RF (Wright & Ziegler, 2017). Parameter *tuning* model berupa *mtry*, *splitrule*, dan *min.node.size* dengan memanfaatkan *package e1071*, *ranger*, dan *dplyr* pada *package caret* di R (Kuhn, 2008, 2016).

- k. *Random forest* (RFrb) dengan metode Rborist

Model ini merupakan implementasi terukur dari RF. Parameter *tuning* model berupa *predFixed* yaitu jumlah trial predictor untuk sebuah *split* (berfungsi seperti *mtry*), dan *minNode* yaitu jumlah minimum ferensi baris yang berbeda untuk membagi sebuah node (Seligman, 2019) dengan memanfaatkan *package Rborist* pada *package caret* di R (Kuhn, 2008, 2016).

- l. *Oblique random forest* (ORFrg) dengan metode ORFridge

Model ini seperti yang dijelaskan sebelumnya dengan *split* data menggunakan kombinasi *linear*. Parameter *tuning* model berupa *mtry* dengan memanfaatkan *package obliqueRF* pada *package caret* di R (Kuhn, 2008, 2016).

- m. *Random forest* (RFrlb) dengan metode *rule-based model* (rfRules)

Model RF yang didasarkan pada serangkaian aturan “if-then” dalam proses pengklasifikasian yang efektif dan secara umum mirip dengan

pendekatan DT tetapi dengan sedikit aturan (*rules*) dan berisi informasi kontekstual dalam aturan tersebut (Berhane et al., 2019). Parameter *tuning* model berupa *mtry* dan *maxdepth* yaitu maksimum level dalam setiap DT (sebagai contoh *depth* 0 berarti hanya ada satu titik cabang, sedangkan *depth* 1 berarti satu internal titik/node ditambah 2 titik cabang) dengan memanfaatkan *package randomForest*, *inTrees*, dan *plyr* pada *package caret* di R (Kuhn, 2008, 2016).

n. *Oblique random forest* (ORFsvm) dengan metode ORFsvm

Model ini seperti yang dijelaskan sebelumnya dengan *split* data menggunakan *linear support vector machine* (SVM). Parameter *tuning* model berupa *mtry* dengan memanfaatkan *package obliqueRF* pada *package caret* di R (Kuhn, 2008, 2016).

o. *Random ferns* (rFern) dengan metode rFerns

Tabel 2. Model, metode, dan parameter *tuning*
Table 2. Models, method, and tuning parameter

Model	Metode	Parameter <i>tuning</i>
<i>Oblique random forest</i> (ORF)	ORFpls	<i>mtry</i>
<i>Random forest</i> (RF)	RF	<i>mtry</i>
<i>Regularized random forest</i> (RRF)	RRF	<i>mtry</i> , <i>coefReg</i> , dan <i>coefImp</i>
<i>Parallel random forest</i> (PRF)	parRF	<i>mtry</i>
<i>Random forest by randomization</i> (RFr)	extraTrees	<i>mtry</i> dan <i>numRandomCuts</i>
<i>Weighted subspace random forest</i> (WSRF)	wsrf	<i>mtry</i>
<i>Conditional inference random forest</i> (CIRF)	cforest	<i>mtry</i>
<i>Regularized random forest</i> (RRFg)	RRFglobal	<i>mtry</i> dan <i>coefReg</i>
<i>Oblique random forest</i> (ORFI)	ORFlog	<i>mtry</i>
<i>Random forest ranger</i> (RFrg)	ranger	<i>mtry</i> , <i>splitrule</i> , dan <i>min.node.size</i>
<i>Random forest</i> (RFrb)	Rborist	<i>predFixed</i> dan <i>minNode</i>
<i>Oblique random forest</i> (ORFrg)	ORFridge	<i>mtry</i>
<i>Random forest</i> (RFrlb)	<i>rule-based model</i> (rfRules)	<i>mtry</i> dan <i>maxdepth</i>
<i>Oblique random forest</i> (ORFsvm)	ORFsvm	<i>mtry</i>
<i>Random ferns</i> (rFern)	rFerns	<i>fern depth</i>

Model ini menggunakan klasifikasi naïve bayes (Kursa, 2014). Parameter *tuning* model berupa *fern depth* yaitu jumlah fitur biner dengan memanfaatkan *package rFerns* pada *package caret* di R (Kuhn, 2008, 2016).

Indikator Performa

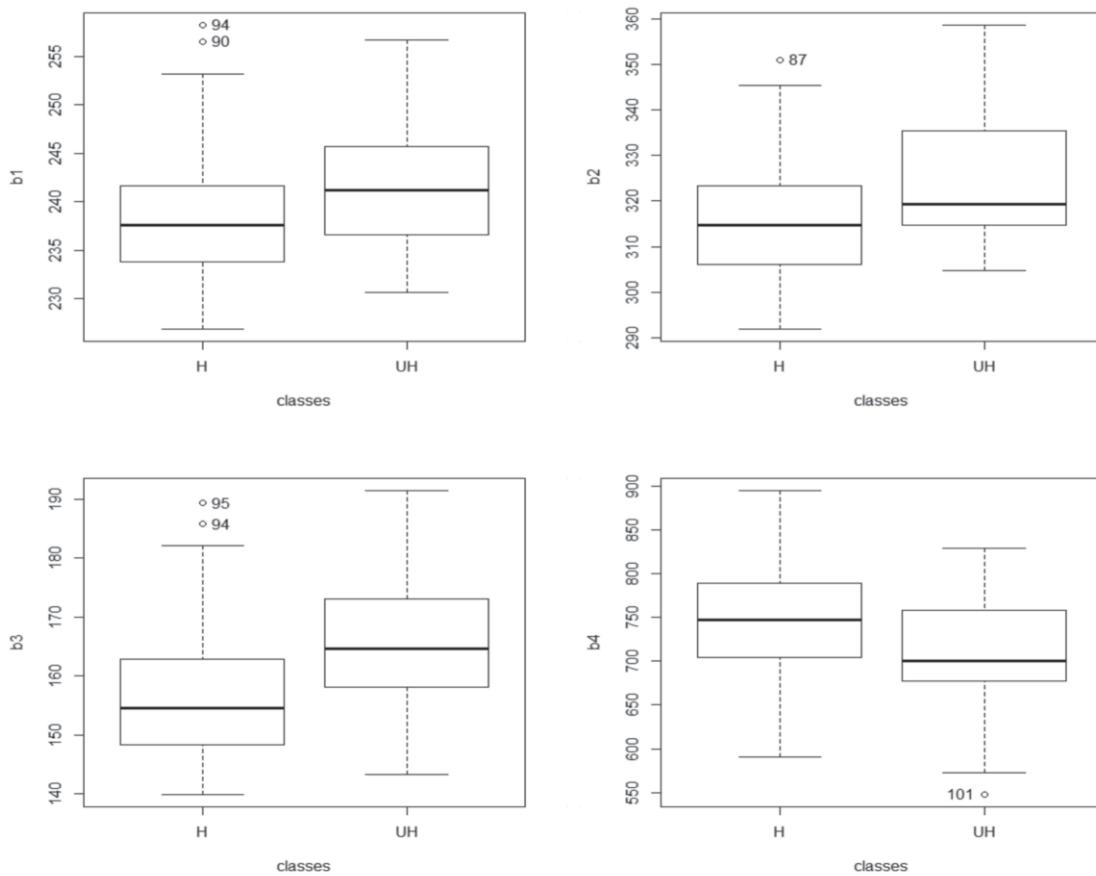
Performa model yang digunakan adalah akurasi interpretasi menggunakan metode *confusion matrix* (Congalton, 1991, 2005; Santoso et al., 2017, 2018; Short, 1982; Viera & Garrett, 2005) serta nilai kappa untuk melihat ukuran reliabilitas dan konsistensi klasifikasi (Cohen, 1960). Selain itu, dievaluasi waktu yang diperlukan dalam *fitting* model dan total proses klasifikasi (*training* dan *testing*) dari 15 model dengan memanfaatkan *package tictoc* (Izrailev, 2016).

HASIL DAN PEMBAHASAN

Penyakit BPB ini dapat digolongkan dalam penyakit *asymptomatic* atau penyakit pada tanaman kelapa sawit yang menyerang tanpa gejala walaupun di lapangan dapat diidentifikasi secara visual melalui dampak dari serangan penyakit BPB tersebut terutama di kanopi dan batang tanaman kelapa sawit (Ahmadi, Muharam, Ahmad, Mansor, & Seman, 2017; Paterson, 2007). Dampak serangan penyakit BPB yang teridentifikasi pada kanopi kelapa sawit dapat dideteksi dan diamati dengan memanfaatkan teknologi penginderaan jauh (Santoso, Gunawan, Jatmiko, Darmosarkoro, & Minasny, 2011; Santoso et al., 2017). Data yang diekstrak dari kanopi tanaman kelapa sawit sampel pada citra QuickBird dari masing-masing band berupa nilai pixel ditunjukkan pada Gambar 1.

Parameter Tuning

Parameter *tuning* dimungkinkan dilakukan pada package *caret* yang ditujukan untuk mencari parameter terbaik berdasarkan akurasi tertinggi. Hasil tuning parameter setiap model klasifikasi tanaman sehat dan terinfeksi penyakit BPB disajikan pada Tabel 3. Model klasifikasi yang menggunakan *mtry* sebagai tuning parameter menunjukkan jumlah variable yang digunakan bervariasi dari satu sampai dengan empat. Model klasifikasi yang hanya menggunakan satu variable dalam proses tuning parameter adalah ORF, RRF, WSRF, RRFg, dan ORF1. Sedangkan nilai *mtry* 2 hanya model ORFrg dan nilai *mtry* 3 yaitu pada model PRF, RFr, RFrg, dan ORFsvm. Sedangkan nilai *mtry* 4 yaitu RF, CIRF, dan RFrlb. Model RFrb menggunakan 2 variable (*predFixed* = 2) dalam tuning parameter.



Gambar 1. Box plot data nilai pixel kanopi pada tanaman sehat (H) dan tidak sehat (UH) pada citra QuickBird band 1 (kiri atas), band 2 (kanan atas), band 3 (kiri bawah), dan band 4 (kanan bawah)

Figure 1. Box plot of pixel values of canopy on healthy (H) and infected (UH) palms from QuickBird imagery of band 1 (top left), band 2 (top right), band 3 (bottom left), and band 4 (bottom right)

Tabel 3. Tuning parameter setiap model

Table 3. Parameter tuning of all models

Model	Metode	Tuning parameter
ORF	ORFpls	<i>mtry</i> = 1
RF	Rf	<i>mtry</i> = 4
RRF	RRF	<i>mtry</i> = 1, <i>coefReg</i> = 4 and <i>coefImp</i> = TRUE
PRF	parRF	<i>mtry</i> = 3
RFr	extraTrees	<i>mtry</i> = 3 and <i>numRandomCuts</i> = 2
WSRF	Wsrif	<i>mtry</i> = 1
CIRF	Cforest	<i>mtry</i> = 4
RRFg	RRFglobal	<i>mtry</i> = 1 and <i>coefReg</i> = 1
ORFI	ORFlog	<i>mtry</i> = 1
RFrg	Ranger	<i>mtry</i> = 3, <i>splitrule</i> = <i>extratrees</i> , and <i>min.node.size</i> =10
RFrb	Rborist	<i>predFixed</i> = 2 dan <i>minNode</i> = 10
ORFrg	ORFridge	<i>mtry</i> = 2
RFrlb	rfRules	<i>mtry</i> = 4 and <i>maxdepth</i> = 2
ORFsvm	ORFsvm	<i>mtry</i> = 3
Rferns	rFerns	<i>Fern depth</i> =1

Akurasi Interpretasi dan Nilai Kappa

Akurasi interpretasi dan nilai Kappa dari tuning parameter masing-masing model disampaikan pada Tabel 4 dan visualisasi akurasi seluruh model ditunjukkan di Gambar 2 dan nilai Kappa ditunjukkan pada Gambar 3. Terdapat lima kelompok model klasifikasi dengan memperhatikan Tabel 4 dan Gambar 2 yaitu kelompok pertama dengan akurasi 0,914 antara lain model ORF dengan metode PLS, RF, RRF, dan PRF. Kelompok kedua dengan akurasi 0,886 antara lain RFr, WSRF, CIRF, dan RRFg. Kelompok ketiga dengan akurasi 0,857 yaitu ORFI, RFrg, dan RFrb. Kelompok keempat dengan akurasi 0,829 yaitu ORFrg, RFrlb, dan ORFsvm. Sedangkan kelompok kelima yaitu Rferns dengan akurasi 0,657. Dengan demikian kelompok pertama sebagai kelompok model klasifikasi dengan akurasi tertinggi dan terendah adalah Rferns. Model ORF dengan metode PLS merupakan model dengan nilai Kappa tertinggi (Tabel 4 dan Gambar 3) yaitu 0,815 sedangkan RF dan RRF sebagai model dengan nilai Kappa tertinggi kedua (nilai Kappa = 0,806) dan Rferns sebagai model dengan nilai Kappa terendah (nilai Kappa = 0,335).

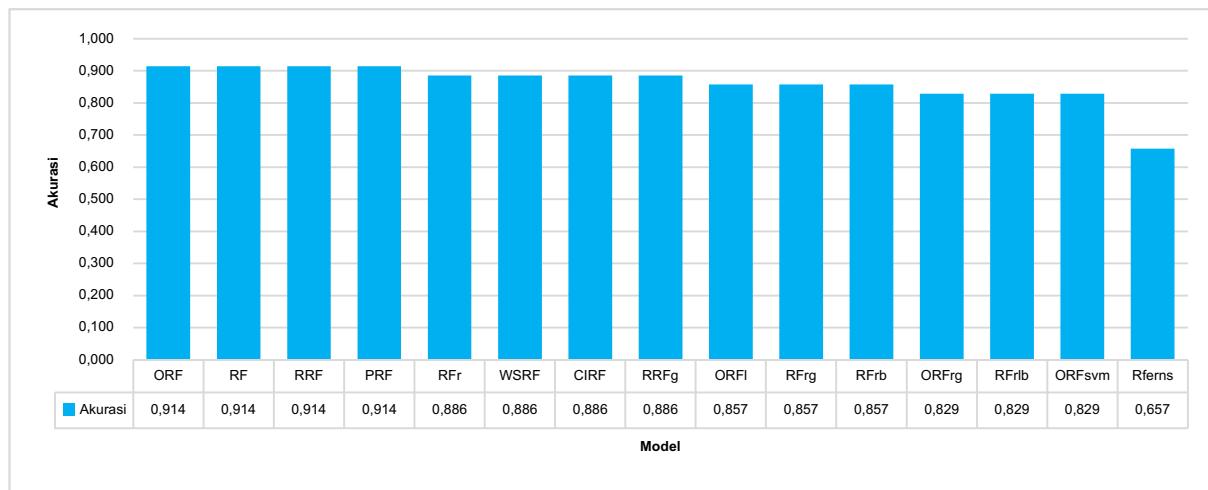
Waktu *Fitting* Model dan Total Proses

Waktu yang diperlukan untuk *fitting* model dan total seluruh proses klasifikasi dari seluruh model klasifikasi disampaikan pada Tabel 5 dan visualisasinya ditunjukkan oleh Gambar 4. Waktu untuk *fitting* model dan total proses terlama terjadi pada model klasifikasi RFrlb dengan waktu berturut-turut yaitu 12.993,23 dan 12.993,30 detik. Model klasifikasi PRF merupakan model dengan waktu tersingkat untuk *fitting* model dan total proses dengan waktu berturut-turut adalah 13,54 dan 13,59 detik. Model klasifikasi dengan waktu *fitting* model dan total proses > 2.000 detik (dari tinggi ke rendah) adalah RFrlb, RFrb, ORFsvm, dan ORFI. Model klasifikasi yang memerlukan waktu 100 sampai dengan 950 detik (dari tinggi ke rendah) adalah ORF, ORFrg, Rferns, dan RRF. Model dengan waktu *fitting* model dan total proses 13,54 sampai dengan 70 detik adalah RFr, WSRF, CIRF, RRFg, RF, dan PRF. Sedangkan model klasifikasi RF sebagai model klasifikasi utama memerlukan waktu *fitting* model dan total proses 15,29 dan 15,36 detik atau memerlukan waktu tersingkat kedua setelah PRF.

Tabel 4. Akurasi klasifikasi dan nilai Kappa setiap model

Table 4. Classification of accuracy and Kappa value of all models

Model	Metode	Akurasi	Kappa
ORF	ORFpls	0,914	0,815
RF	Rf	0,914	0,806
RRF	RRF	0,914	0,806
PRF	parRF	0,914	0,796
RFr	extraTrees	0,886	0,759
WSRF	Wsrf	0,886	0,747
CIRF	Cforest	0,886	0,735
RRFg	RRFglobal	0,886	0,735
ORFI	ORFlog	0,857	0,691
RFrg	Ranger	0,857	0,677
RFrb	Rborist	0,857	0,677
ORFrg	ORFridge	0,829	0,638
RFrlb	rfRules	0,829	0,638
ORFsvm	ORFsvm	0,829	0,602
Rferns	rFerns	0,657	0,335



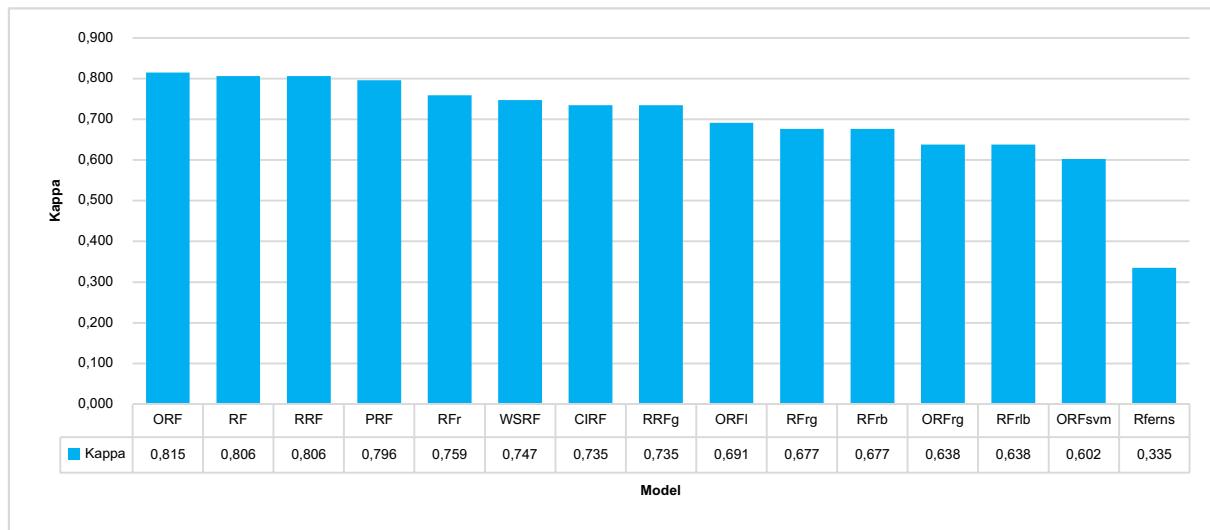
Gambar 2. Akurasi klasifikasi masing-masing model

Figure 2. Classification accuracy of all models

Pembahasan

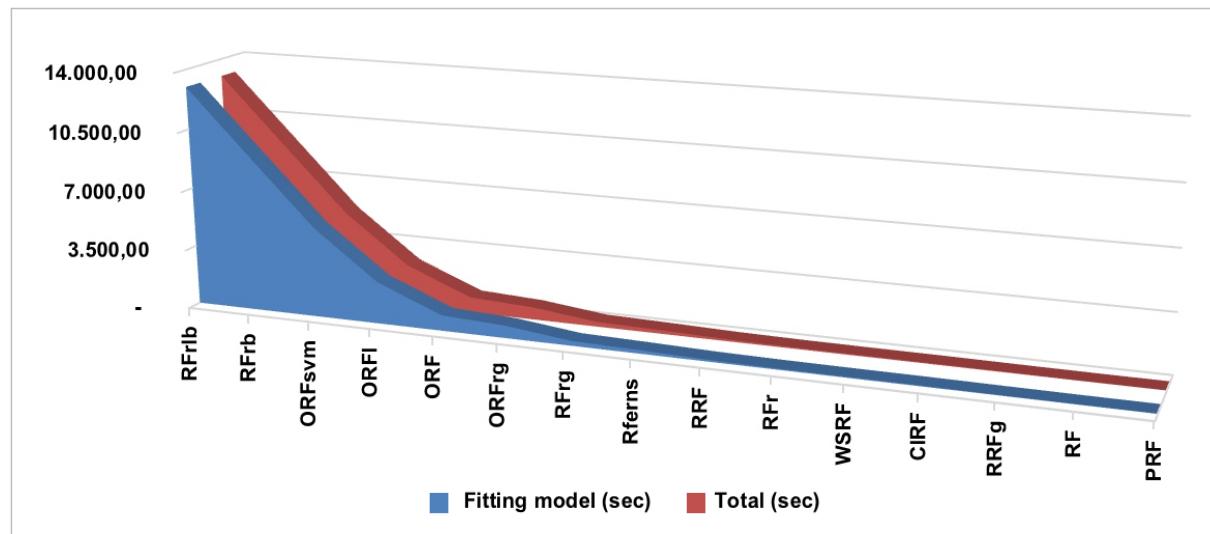
Klasifikasi tanaman kelapa sawit yang sehat dan terinfeksi penyakit BPB dengan menggunakan *machine learning* sangat dipengaruhi oleh variable yang digunakan untuk *fitting* model. Distribusi data setiap kelas (H dan UH) pada masing-masing variable dalam kajian ini berupa empat band dari citra QuickBird (Gambar 1) memberikan gambaran relatif sulit membedakan tanaman yang sehat (H) dan

terinfeksi penyakit BPB (UH). Data *boxplot* nilai pixel citra QuickBird pada penelitian ini lebih baik jika dibandingkan dengan penelitian Santoso et al., (2018), sehingga akurasi klasifikasi lebih baik pada penelitian ini. *Scatterplot* matrik pada Gambar 5 menunjukkan data yang saling overlap antara tanaman kelapa sawit H dan UH, sehingga dapat digunakan untuk menguji model klasifikasi terkait performa akurasi klasifikasi seperti yang dilakukan pada penelitian ini.



Gambar 3. Nilai Kappa dari seluruh model

Figure 3. Kappa value of all models



Gambar 4. Waktu untuk *fitting* model dan total seluruh proses setiap model

Figure 4. Time consumed for fitting model and total processes of all models

Tabel 5. Waktu untuk *fitting* model dan total proses setiap model

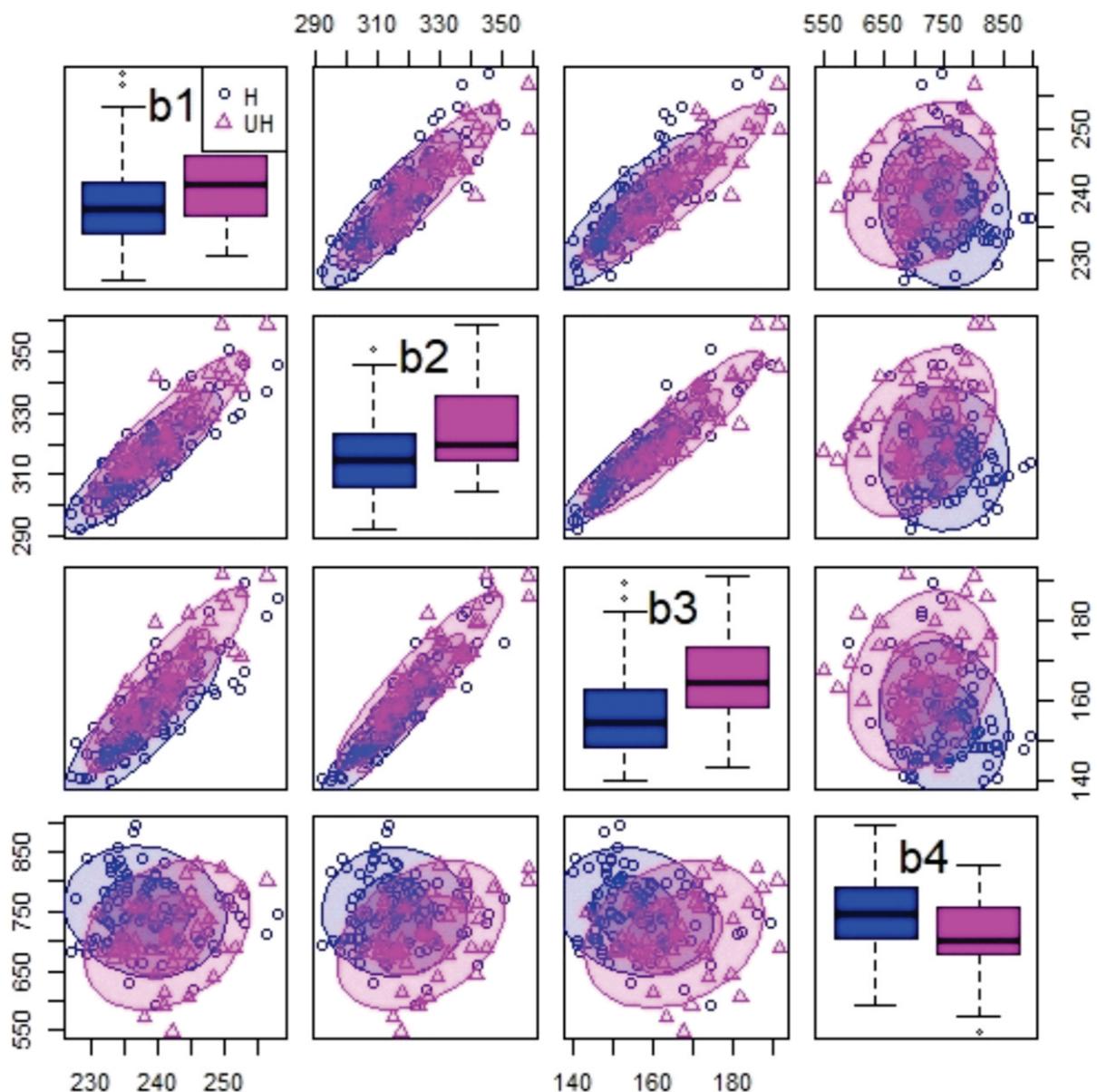
Table 5. Time consumed for fitting model and total processes of all models

Model	Metode	Fitting model (detik)	Total (detik)
ORF	ORFpls	920,59	920,75
RF	Rf	15,29	15,36
RRF	RRF	107,75	107,82
PRF	parRF	13,54	13,59
RFr	extraTrees	67,58	67,64
WSRF	Wsrf	52,79	52,86
CIRF	Cforest	48,03	48,18
RRFg	RRFglobal	27,20	27,27
ORFI	ORFlog	2.453,02	2.453,14
RFrg	Ranger	281,62	281,72
RFrb	Rborist	9.134,47	9.134,55
ORFrg	ORFridge	726,58	726,71
RFrlb	rfRules	12.993,23	12.993,30
ORFsvm	ORFsvm	5.294,05	5.294,19
Rferns	rFerns	203,50	203,55

Pencapaian akurasi klasifikasi (Tabel 4 dan Gambar 2), nilai Kappa (Tabel 4 dan Gambar 3), dan waktu yang diperlukan untuk *fitting* model serta total proses (Tabel 5 dan Gambar 4) dapat dijadikan pertimbangan dalam pemilihan model klasifikasi. Berdasarkan pencapaian akurasi klasifikasi dan nilai Kappa, ORF dengan metode PLS sebagai model *split* dalam DT sebagai model terbaik walaupun diperlukan waktu 920,59 detik untuk *fitting* model. Akurasi klasifikasi pada model RF dan RRF sama dengan model ORF namun nilai Kappa yang lebih kecil 0,009 dibandingkan nilai Kappa pada model ORF. Namun, waktu untuk *fitting* model pada ORF > RRF > RF yang lamanya berturut-turut sebesar 920,59, 107,75, dan 15,29 detik. Model ORF secara efisien memilih satu variable dalam *fitting* model dengan metode PLS yang secara fungsinya mengurangi dimensi data yaitu menghilangkan data yang tidak relevan dan pengulangan penggunaan variable serta memperbaiki kualitas data untuk strategi pengorganisasian data yang lebih efisien (Dhage & Raina, 2016). Hal yang sama dilakukan pada model RRF dalam hal pengulangan variable

(*redundant*), sehingga dalam RRF variable yang digunakan hanya 1 variable. Hasil yang berbeda disampaikan oleh Poona et al., (2016) bahwa ORF dengan model SVM memberikan hasil akurasi yang lebih tinggi dibandingkan dengan ORF dengan metode PLS dalam kajian klasifikasi tanaman yang sehat dan terinfeksi jamur pada pembibitan tanaman pinus dengan data hyperspectral. Jumlah variable, data yang *redundant*, dan struktur data akan mempengaruhi akurasi klasifikasi apalagi obyek yang diklasifikasikan berupa penyakit yang *asymptomatic* seperti yang disampaikan Poona et al. (2016) dan (Santoso et al., 2018).

Model Rferns yang secara prinsip menggunakan struktur *flat* dalam proses training dan klasifikasi dengan kecepatan training dan klasifikasi yang lebih cepat dibandingkan dengan model dengan struktur hirarki DT (Oberweger, 2014). Hasil penelitian menunjukkan hal yang berbeda, model Rferns memerlukan waktu 203,50 detik untuk *fitting* model dan mempunyai akurasi klasifikasi yang paling rendah dari seluruh model yang digunakan. Hal ini



Gambar 5. Scattterplot matrik seluruh data
Figure 5. Matrix scatterplot of the data

kemungkinan disebabkan oleh data yang overlap di masing-masing variable (Gambar 5) dan struktur flat tidak sesuai dengan data yang digunakan. Seluruh model dengan struktur hirarki dalam penelitian ini mempunyai akurasi lebih besar dari 0,80 dan

menurut Sutanto (2010) dengan mempertimbangkan kesulitan dalam klasifikasi tanaman yang sehat dan terinfeksi penyakit BPB yang masuk dalam penyakit *asymptomatic*, maka akurasi $\geq 0,80$ masih dapat diterima dari perspektif ilmu penginderaan jauh.

KESIMPULAN

Hasil penelitian ini menunjukkan model klasifikasi ORF dengan metode PLS merupakan model dengan akurasi dan nilai Kappa terbaik (0,914 dan 0,815), sedangkan model Rferns sebagai model klasifikasi dengan akurasi dan nilai Kappa yang terendah (0,657 dan 0,334). Model RFrlb memerlukan waktu untuk *fitting* model terlama (12.993,23 detik), sedangkan model PRFmemerlukan waktu tersingkat (13,54 detik). Pencapaian akurasi dan waktu yang diperlukan untuk *fitting* model menjadi pertimbangan dalam pemilihan model klasifikasi tanaman sehat dan tanaman terinfeksi penyakit BPB yang disebabkan oleh jamur *Ganoderma*.

DAFTAR PUSTAKA

- Ahmadi, P., Muharam, F. M., Ahmad, K., Mansor, S., & Seman, I. A. (2017). Early detection of ganoderma basal stem rot of oil palms using artificial neural network spectral analysis. *Plant Disease*, 101(6), 1009–1016. <https://doi.org/10.1094/PDIS-12-16-1699-RE>
- Behmann, J., Mahlein, A.-K., Rumpf, T., Römer, C., & Plümer, L. (2015). A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. *Precision Agriculture*, 16(3), 239–260. <https://doi.org/10.1007/s11119-014-9372-7>
- Berhane, T. M., Lane, C. R., Wu, Q., Autry, B. C., Anenkhonov, O. A., Chepinoga, V. V., & Liu, H. (2019). Decision-Tree, Rule-Based, and Random Forest Classification of High-Resolution Multispectral Imagery for Wetland Mapping and Inventory. *Remote Sens (Basel)*, 10(4), 580. <https://doi.org/10.3390/rs10040580>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1201/9780429469275-8>
- Chen, J., Li, K., Tang, Z., Bilal, K., Yu, S., Weng, C., & Li, K. (2016). A Parallel Random Forest Algorithm for Big Data in a Spark Cloud Computing Environment. *IEEE Transactions on Parallel and Distributed System*, 27(1–1), 1–15. <https://doi.org/10.1109/TPDS.2016.2603511>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35–46. [https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B)
- Congalton, R. G. (2005). Thematic and Positional Accuracy Assessment of Digital Remotely Sensed Data. *2005 Proceedings of the Seventh Annual Forest Inventory and Analysis Symposium*, (1993), 149–154.
- Crawford, M. M., Tuia, D., & Yang, H. L. (2013). Active Learning: Any Value for Classification of Remotely Sensed Data? *PROCEEDINGS OF THE IEEE. PREPRINT DOWNLOADED FROM WIKI.EPFL.CK/EO-ADAPT*. Retrieved from <http://lasig.epfl.ch/>
- Dhage, S. N., & Raina, C. K. (2016). A review on Machine Learning Techniques. *International Journal on Recent and Innovation Trends in Computing and Communication*, 4(3), 395–399. Retrieved from <http://www.ijritcc.org>
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Mach Learn* (). <https://doi.org/10.1007/s10994-006-6226-1>
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3), 651–674. <https://doi.org/10.1198/106186006X133933>
- Hushiaran, R., Yusof, N. A., & Dutse, S. W. (2013). Detection and control of Ganoderma boninense: strategies and perspectives. *SpringerPlus*, 2(1), 555. <https://doi.org/10.1186/2193-1801-2-555>
- Izrailev, S. Functions for timing R scripts, as well as implementations of Stack and List structures., 1.0 Package ‘tictoc’ Version 1.0 1–8 (2016). Retrieved from <https://cran.r-project.org/web/packages/tictoc/tictoc.pdf>
- Kim, Y. H., Yoo, S. J., Gu, Y. H., Lim, J. H., Han, D., & Baik, S. W. (2014). Crop Pests Prediction Method Using Regression and Machine Learning Technology: Survey. *IERI Procedia*, 6, 5 2 – 5 6 . <https://doi.org/10.1016/j.ieri.2014.03.009>

- Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal Of Statistical Software*, 28 (5), 1 – 26 . <https://doi.org/10.1053/j.sodo.2009.03.002>
- Kuhn, M. (2016). “caret - Classification and Regression Training” R Package. <https://github.com/topepo/caret/>. Retrieved from <https://github.com/topepo/caret/>
- Kuhn, M. (2019). The caret Package. Retrieved from <http://topepo.github.io/caret/train-models-by-tag.html#random-forest>
- Kursa, M. B. (2014). rFerns: An implementation of the random ferns method for general-purpose machine learning. *Journal of Statistical Software*, 61 (10) , 1 – 13 . <https://doi.org/10.18637/jss.v061.i10>
- Lelong, C. C. D., Roger, J. M., Brégand, S., Dubertret, F., Lanore, M., Sitorus, N. a., ... Caliman, J. P. (2010). Evaluation of oil-palm fungal disease infestation with canopy hyperspectral reflectance data. *Sensors*, 10, 734–747. <https://doi.org/10.3390/s100100734>
- Liaghat, S., Ehsani, R., Mansor, S., Shafri, H. Z. M., Meon, S., Sankaran, S., & Azam, S. H. M. N. (2014). Early detection of basal stem rot disease (Ganoderma) in oil palms based on hyperspectral reflectance data using pattern recognition algorithms. *International Journal of Remote Sensing*, 35 (10) , 3427 – 3439 . <https://doi.org/10.1080/01431161.2014.903353>
- Menze, B. H., Kelm, B. M., Splitthoff, D. N., Koethe, U., & Hamprecht, F. A. (2011). On oblique random forests. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6912 LNCS(PART 2), 453–469. https://doi.org/10.1007/978-3-642-23783-6_29
- Oberweger, M. (2014). *Embeddings for Random Ferns Classification*. Graz University of Technology, Austria.
- Omer, G., Mutanga, O., Abdel-Rahman, E. M., & Adam, E. (2016). Empirical prediction of leaf area index (LAI) of endangered tree species in intact and fragmented indigenous forests ecosystems using WorldView-2 data and two robust machine learning algorithms. *Remote Sensing*, 8(4), 1–26. <https://doi.org/10.3390/rs8040324>
- Paterson, R. R. M. (2007). Ganoderma disease of oil palm-A white rot perspective necessary for integrated control. *Crop Protection*, 26(9), 1 3 6 9 – 1 3 7 6 . <https://doi.org/10.1016/j.cropro.2006.11.009>
- Poona, N., van Niekerk, A., & Ismail, R. (2016). Investigating the utility of oblique tree-based ensembles for the classification of hyperspectral data. *Sensors (Switzerland)*, 16(11). <https://doi.org/10.3390/s16111918>
- Priwiratama, H., Prasetyo, A. E., & Susanto, A. (2014). Pengendalian Penyakit Busuk Pangkal Batang Kelapa Sawit secara Kultur Teknis. *Jurnal Fitopatologi Indonesia*, 10(51), 1–7. <https://doi.org/10.14692/jfi.10.1.1>
- Santoso, H., Gunawan, T., Jatmiko, R. H., Darmosarkoro, W., & Minasy, B. (2011). Mapping and identifying basal stem rot disease in oil palms in North Sumatra with QuickBird imagery. *Precision Agriculture*, 12(2), 233–248. <https://doi.org/10.1007/s11119-010-9172-7>
- Santoso, H., Tani, H., & Wang, X. (2017). Random Forest classification model of basal stem rot disease caused by Ganoderma boninense in oil palm plantations. *International Journal of Remote Sensing*, 38 (16) , 4683 – 4699 . <https://doi.org/10.1080/01431161.2017.1331474>
- Santoso, H., Tani, H., Wang, X., Prasetyo, A. E., & Sonobe, R. (2018). Classifying the severity of basal stem rot disease in oil palm plantations using WorldView-3 imagery and machine learning algorithms. *International Journal of Remote Sensing*, 00(00) , 1–23 . <https://doi.org/10.1080/01431161.2018.1541368>
- Seligman, M. (2019). Package “Rborist” extensible, parallelizable implementation of the random forest. *Package “Rborist.”* Retrieved from <https://cran.r-project.org/web/packages/Rborist/Rborist.pdf>
- Short, N. M. (1982). *The landsat tutorial workbook-Basics of satellite remote sensing*. Washington, D.C.: National Aeronautics and Space Administration, Scientific and Technical Information Branch.
- Sutanto. (2010). REMOTE SENSING RESEARCH: A User’s Perspective. *Indonesian Journal of*

- Geography*, 42(2), 129–142.
<https://doi.org/10.22146/indo.j.geog,2287>
- Uhlmann, S., Kiranyaz, S., & Gabbouj, M. (2014). Semi-Supervised Learning for Ill-Posed Polarimetric SAR Classification. *Remote Sens*, 6, 4801–4830.
<https://doi.org/10.3390/rs6064801>
- Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, 37(5), 360–363.
<https://doi.org/Vol. 37, No. 5>
- Wieland, M., & Pittore, M. (2014). Performance Evaluation of Machine Learning Algorithms for Urban Pattern Recognition from Multi-spectral Satellite Images. *Remote Sens*, 6, 2912–2939.
<https://doi.org/10.3390/rs6042912>
- Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77(1), 1–17.
<https://doi.org/10.18637/jss.v077.i01>
- Zhao, H., Williams, G. J., & Huang, J. Z. (2017). Wsrf: An R package for classification with scalable weighted subspace random forests. *Journal of Statistical Software*, 77(3).
<https://doi.org/10.18637/jss.v077.i03>