

Penerapan Machine Learning Dalam Analisis Stadium Penyakit Hati Untuk Proses Diagnosis dan Perawatan

Jimmy*, Lili Dwi Yulianto, Eni Heni Hermaliani, Laela Kurniawati

Fakultas Teknologi Informasi, Universitas Nusa Mandiri, Jakarta, Indonesia

Email: ^{1,*}14210178@nusamandiri.ac.id, ²14210161@nusamandiri.ac.id, ³enie_h@nusamandiri.ac.id, ⁴laela@nusamandiri.ac.id

Email Penulis Korespondensi: 14210178@nusamandiri.ac.id

Abstrak-Penyakit hati merupakan penyakit yang sudah lama ada dan cukup umum di masyarakat. Penyakit ini terjadi dikarenakan organ hati tidak dapat bekerja dengan optimal dikarenakan adanya peradangan atau virus. Oleh karena itu salah satu cara yang digunakan untuk mengetahui penyakit hati adalah dengan melakukan tes darah di laboratorium sehingga mendapatkan informasi berupa kadar enzim, akan tetapi tes darah di laboratorium memerlukan biaya yang cukup mahal sehingga prediksi menggunakan machine learning sangat diperlukan untuk kasus ini, karena gejala pada penyakit hati perlu penanganan secara cepat. Data rekam medis dan hasil laboratorium menghasilkan banyak fitur sedangkan fitur yang terlalu banyak dapat mengurangi nilai akurasi pada machine learning, sehingga diperlukan model features selection untuk mengetahui fitur yang paling berpengaruh pada machine learning. Pada penelitian ini menggunakan tiga model features selection yaitu Random Forest Importance, Chi Square Test dan Recursive Features Elimination dan berhasil mendapatkan dua fitur tertinggi yaitu SGOT dan SGPT. Hasil akurasi akan dibandingkan antara dua fitur dengan sebelas fitur menggunakan K-fold Cross Validation, dan melakukan perbandingan menggunakan model Features Extraction dengan menggunakan Principal Component Analysis (PCA). Perhitungan akurasi dilakukan menggunakan algoritma Random Forest, Decision Tree, Naive Bayes, Logistic Regression, Support Vector Machine, KNN, Gradient Boosting dan Artificial Neural Network, Hasil perhitungan akurasi menggunakan algoritma Random Forest dengan PCA antara sebelas dan dua fitur mengalami penurunan sebesar 0,6%, sedangkan menggunakan features selection meningkat sebesar 0,7%, didapati akurasi tertinggi menggunakan algoritma Random Forest dengan 2 fitur sebesar 72,2%.

Kata Kunci : Penyakit Hati; Machine Learning; Features Selection; Features Extraction; PCA; SGOT; SGPT

Abstract-Liver disease is a disease that has existed for a long time and is quite common in society. This disease occurs because the liver cannot work optimally due to inflammation or viruses. Therefore, one of the ways used to determine liver disease is to do a blood test in the laboratory so as to obtain information in the form of enzyme levels, but blood tests in the laboratory require a fairly expensive so that predictions using machine learning is needed for this case, because the symptoms of liver disease need to be handled quickly. Medical record Data and laboratory results produce many features while too many features can reduce the value of accuracy in machine learning, so the features selection model is needed to determine the most influential features in machine learning. in this research that using three models of features selection, namely Random Forest Importance, Chi Square Test and Recursive Features Elimination and managed to get the two highest features, namely SGOT (Serum Glutamic Oxaloacetic Transaminase) and SGPT (Serum Glutamic Pyruvic Transaminase). Accuracy results will be compared between two features with eleven features using K-fold Cross Validation, and perform comparison using Features Extraction model using Principal Component Analysis (PCA). Accuracy calculation is done using Random Forest algorithm, Decision Tree, Naive Bayes, Logistic Regression, Support Vector Machine, KNN, Gradient Boosting and Artificial Neural Network, the calculation accuracy using Random Forest algorithm with PCA between Eleven and two features decreased by 0.6%, while using features selection increased by 0.7%, found the highest accuracy using Random Forest algorithm with 2 features of 72.2%.

Keywords : Liver Disease; Machine Learning; Features Selection; Features Extraction; PCA; SGOT; SGPT

1. PENDAHULUAN

Di dalam tubuh kita terdapat beberapa organ penting yang masing-masing fungsinya sangat berguna, salah satunya adalah hati. Hati merupakan organ di dalam tubuh yang mempunyai ukuran paling besar dan memiliki fungsi sangat penting[1][2][3]. Dengan adanya organ hati maka tubuh kita akan terhindar dari berbagai racun yang mampu mengganggu kesehatan. Fungsi organ hati yaitu sebagai tempat untuk penyimpanan glikogen, membantu proses pembentukan dan sekresi empedu, sintesa urea, berperan sebagai metabolisme kolesterol dan lemak serta fungsi utama hati sebagai detoksifikasi racun atau penetral racun[4]. Hati secara luas dilindungi oleh tulang iga dengan berat 1500 gram atau 2,5% berat tubuh pada orang dewasa normal. Setiap gangguan yang terjadi pada organ hati dapat menyebabkan peradangan akut atau kronis yaitu melemahnya fungsi hati lebih parahnya lagi dapat merusak organ lain di dalam tubuh[5][6].

Penyakit liver merupakan penyakit hati yang mengakibatkan hati tidak mampu berfungsi bahkan menyebabkan kematian dan sudah lama ada dan cukup umum di masyarakat[7][8]. Menurut data WHO (World Health Organization) memaparkan bahwa sekitar 1,2 juta orang per-tahun di wilayah di Asia Tenggara dan Afrika mengalami kematian akibat dari penyakit hati[9][10]. Apabila organ hati tidak berfungsi dengan baik maka kemampuan untuk menetralisir racun yang masuk ke tubuh kita, hal tersebut akan membahayakan tubuh jika tidak segera ditangani[11].

Fase pertama penyakit hati ditandai dengan adanya peradangan, sel-sel jaringan menjadi lunak dan membengkak pada organ hati, fase kedua ditandai dengan gangguan pernafasan akibat terbentuknya jaringan parut di organ paru-paru (fibrosis)[12][13][14], pada fase ketiga ditandai dengan terjadinya sirosis[15][16][17][18], yaitu kerusakan parah pada hati akibat penumpukan jaringan parut, pada fase keempat akan berakibat gagal hati karena fungsi hati hilang secara keseluruhan (kronis)[2][6]. Pengobatan yang disarankan yaitu transplantasi hati[19]. Untuk melakukan diagnosis penyakit hati, dokter akan melakukan beberapa pengecekan salah satunya dengan melakukan tanya jawab mengenai gejala-gejala

yang dirasakan oleh pasien, kedua akan melihat riwayat kesehatan dan faktor risiko yang diderita oleh pasien[20][21], seperti halnya Riwayat mengkonsumsi obat-obatan yang sudah diminum atau jumlah alkohol yang sudah dikonsumsi setiap harinya[22]. Tahap ketiga dokter akan melakukan pemeriksaan secara fisik dan menyeluruh[23], dan juga memperhatikan perubahan warna kulit dan mata, pengecekan apakah adanya rasa nyeri ketika perut ditekan, serta adanya pembengkakan di bagian perut dan kaki[24]. Hampir semua orang mengalami keterlambatan penanganan dikarenakan terkendala biaya dan gejala yang sudah cukup parah[25][26]. [tambahkan referensi]

Pemeriksaan oleh dokter dapat dilakukan di laboratorium menggunakan sample darah untuk mengetahui kondisi peradangan hati[27], hasil uji tersebut memberikan informasi berupa jumlah bilirubin[28], enzim alkaline phosphatase[29][28], enzim SGOT (Serum Glutamic Oxaloacetic Transaminase), enzim SGPT (Serum Glutamic Pyruvic Transaminase)[29], protein dan kadar albumin[30]. SGOT dan SGPT merupakan enzim yang dapat ditemukan pada organ hati, jantung, ginjal dan otak dan memiliki peran untuk membantu proses pencernaan protein dan metabolisme di dalam tubuh[5][29]. Kadar enzim ini harus dalam kadar normal jika tidak maka enzim ini akan keluar dari sel organ tersebut lalu masuk ke pembuluh darah sehingga nilai SGOT dan SGPT menjadi tinggi, kadar SGOT dan SGPT memiliki peran penting untuk mengetahui kerusakan dan peradangan pada organ hati[31].

Tes laboratorium untuk mendapatkan sample darah secara lengkap memerlukan langkah dan persiapan khusus seperti mencatat apa saja obat-obatan yang sedang dikonsumsi, dan juga membutuhkan biaya yang tidak sedikit[32]. Oleh karena itu machine learning dapat membantu mengetahui tingkat peradangan dan kerusakan pada organ hati dengan menggunakan dua informasi kadar enzim yaitu SGOT, enzim SGPT. Kedua enzim tersebut memiliki pengaruh kuat terhadap kerusakan organ hati dan juga mendapatkan peringkat empat teratas jika dibandingkan dengan fitur lain pada sample darah yang didapat, cara kami memberikan peringkat tersebut dengan menggunakan model features selection[33][34][35], model tersebut memiliki tujuan untuk menyeleksi fitur apa saja yang memiliki pengaruh terhadap prediksi pada machine learning, kami menggunakan tiga jenis features selection yaitu random forest importance, chi square test, dan recursive features elimination[36][37][38][39].

Untuk memastikan prediksi yang kami lakukan memiliki akurasi yang cukup baik, kami melakukan pengujian tambahan menggunakan features extraction, model yang kami gunakan yaitu Principal Component Analysis (PCA)[40]. Tujuan dari model ini adalah untuk mengurangi kompleksitas hubungan timbal-balik antara sejumlah besar variabel yang diamati ke sejumlah relatif kecil dari kombinasi linearnya, yang biasa disebut dengan komponen utama[41][42].

Penelitian yang dilakukan oleh Davide Chicco et al bertujuan untuk melakukan prediksi kelangsungan hidup pasien yang menderita gagal jantung dengan cara membandingkan pengujian menggunakan 2 fitur saja yaitu serum creatinine dan ejection fraction dan dibandingkan dengan menggunakan semua fitur dari dataset, untuk menghasilkan 2 fitur tertinggi penulis menggunakan feature ranking dengan menggunakan metode Pearson correlation coefficients (PCC), Shapiro–Wilk tests, dan Chi squared test menghasilkan 2 fitur yang memiliki pengaruh besar terhadap proses prediksi yaitu Serum creatinine dan ejection fraction sehingga penelitian tersebut menghasilkan akurasi yang tinggi tanpa harus menggunakan data rekam medis yang lengkap, dapat dibuktikan bahwa dengan menggunakan serum creatinine dan ejection fraction dapat menghasilkan akurasi 0.75 menggunakan algoritma random forests, 0.750 dengan menggunakan gradient boosting dan SVM radial dengan akurasi sebesar 0.720[43].

Penelitian selanjutnya yang dilakukan oleh Jagdeep Singh, et al yang bertujuan untuk melakukan analisis prediksi tingkat keparahan penyakit hati dengan membandingkan akurasi yang dihitung menggunakan semua field yaitu 5 fitur yang didapat dari hasil features selection. Metode features selection pada penelitian ini menggunakan metode feature evaluator and greedy stepwise dengan hasil yang cukup signifikan yaitu dengan menggunakan semua field diapati akurasi sebesar 72,50% menggunakan algoritma logistic regression sedangkan dengan menggunakan 5 fitur akurasi yang didapat meningkat menjadi 74,36% pada penelitian ini menyimpulkan bahwa features selection dapat memberikan pengaruh yang cukup signifikan terhadap nilai akurasi[29].

Berdasarkan latar belakang diatas, penelitian ini memiliki tujuan untuk melakukan analisis prediksi tingkat keparahan penyakit hati apakah termasuk kronis atau tidak kronis dengan melakukan perbandingan menggunakan sebelas fitur dengan dua fitur yang didapat dari features selection dengan menggunakan model random forest importance[9], chi square test dan recursive features elimination[44][45], selanjutnya akan dilakukan prediksi ulang menggunakan features extraction dengan model Principal Component Analysis (PCA)[40]. Algoritma machine learning yang digunakan yaitu Support Vector Machine (VM)[46], K-Nearest Neighbors (K-NN)[1], Logistic Regression (LR)[47], Decision Tree (DT), Artificial Neural Network (ANN)[48][49], Naive Bayes (NB), Gradient Boosting (GB)[50], dan Random Forest (RF)[51], Evaluasi model yang kami gunakan adalah menggunakan k-fold cross validation.

2. METODOLOGI PENELITIAN

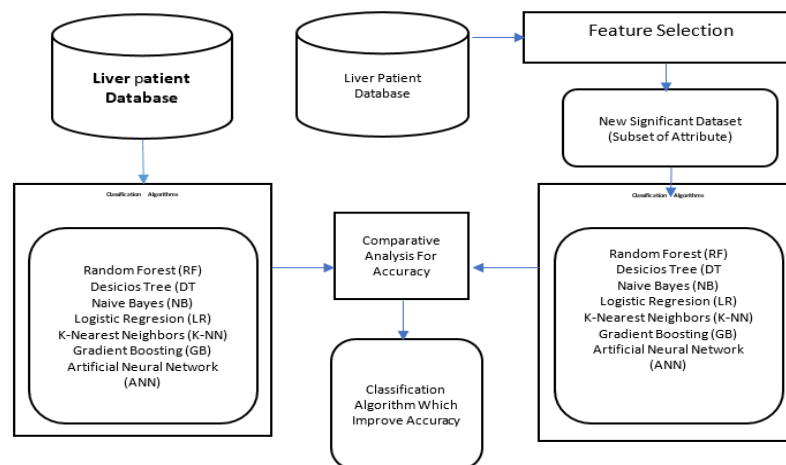
2.1 Dataset

Pada penelitian ini menggunakan dataset yang didapat dari website UCI Machine Learning Repository yaitu Indian Liver Patient Dataset (ILPD). Dataset ini berisi data yang dikumpulkan dari para pasien yang ada di timur laut Andhra Pradesh, India. Dataset berisi 416 pasien penderita liver sedangkan 167 pasien bukan penderita liver, data pasien pria yaitu 441 dan 142 pasien wanita. Dataset ini memiliki 11 atribut dimana 10 atribut sedangkan 1 atribut class atau label yang ditunjukkan pada tabel 1, dan memiliki 583 instance atau isi data. Instance pada dataset ini tidak memiliki nilai missing sehingga semua atribut berisi nilai dalam kondisi baik dan siap diproses[29].

Tabel 1. Dataset menggunakan 10 Fitur

NO	ATRIBUT	TIPE DATA	FUNGSI
1	Usia	Numeric	Menunjukkan Usia (Tahun)
2	Jenis Kelamin	Kategorikal	Menunjukkan Jenis Kelamin pada Pasien (Female/Male)
3	Total Bilirubin (TB)	Numeric	Menunjukkan Sel Darah Merah di Hati (mg/dl)
4	Direct Bilirubin (DB)	Numeric	Bilirubin Bebas (mg/dl)
5	Alkaline Phosphotase (Alkphos)	Numeric	Menunjukkan Enzim dari Usus
6	SGPT Alamine	Numeric	Menunjukkan Enzim didalam Hati
6	SGOT Protiens	Numeric	Menunjukkan Enzim protiens didalam Hati
7	Total Protiens (TP)	Numeric	Serum protein yang terdapat didalam Hati (g/dl)
8	Albumin	Numeric	Sintesa Protein didalam Hati
9	Rasio Albumin dan Globulin (Ratio A/G)	Numeric	Menunjukkan Perbandingan Rasio Albumin dan Globulin
10	Class Variable	Numeric	Positive Liver (1) dan Negative Liver (0)

Dalam melakukan analisis prediksi penyakit hati, beberapa langkah harus dilakukan untuk memperoleh hasil yang terbaik, berikut tahapan yang digunakan pada metode penelitian :



Gambar 1. Metodologi Penelitian

2.2 Feature Selection

Features selection digunakan untuk menghapus beberapa fitur yang dianggap tidak memiliki dampak yang cukup besar terhadap proses klasifikasi dan prediksi[29], banyaknya fitur juga dapat mempengaruhi nilai akurasi dan dampak buruknya dapat menyebabkan masalah overfitting, overfitting merupakan keadaan dimana data yang digunakan untuk data latih bernilai “baik” sehingga ketika dilakukan menggunakan data baru akan berdampak pada penurunan nilai akurasi[52][34].

2.1.1 Random Forest Importance

Merupakan model perangkaian fitur yang menggunakan konsep pohon keputusan, Setiap cabang pohon yang dihasilkan dari model random forest akan dihitung untuk mendapatkan fitur-fitur yang penting dari sebuah dataset ini berkaitan dengan cara kerja classification and regression trees (CART)[14]. Data training yang digunakan, $X = x_1, \dots, x_n$ and $Y = y_1, \dots, y_n$ untuk sebuah keluaran dengan rumus :

$$f = \frac{1}{B} \sum_{b=1}^B f_b(x') \tag{1}$$

Nilai X' akan dicocokkan dengan f_b dan kemudian semua hasil pohon yang dihasilkan yaitu B akan dirata ratakan, semakin banyak cabang pohon, semakin penting fitur tersebut untuk digunakan[53].

2.1.2 Chi Square Test

Metode ini menghitung Chi kuadrat antara setiap fitur dan target dan memilih jumlah fitur yang diinginkan dengan skor Chi kuadrat terbaik.

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (2)$$

O_i merupakan nilai yang akan diobservasi sedangkan E_i merupakan nilai yang akan dicari. chi square sangat berguna untuk menguji hubungan antara dua buah variabel dan mengukur kuatnya hubungan antara variabel[44]

2.1.3 Recursive Features Elimination

Metode features selection ditawarkan dalam library sklearn python adalah recursive features elimination yaitu melakukan perangkangan dengan cara penghapusan fitur yang tidak perlu berdasarkan variant dan korelasi[54], dengan cara mempertimbangkan secara rekursif sebuah dataset yaitu fitur yang lebih kecil, dengan cara melatih estimator berdasarkan kumpulan fitur[23].

2.2 Klasifikasi

Klasifikasi merupakan salah satu teknik yang digunakan untuk mengelompokkan data berdasarkan kesamaan label, dengan cara menemukan suatu pola yang dapat membedakan kelas-kelas tertentu agar dapat digunakan untuk prediksi [4], klasifikasi dapat menentukan model yang menggambarkan dan membedakan kelas dari fitur yang terdapat di dalam dataset, model tersebut diturunkan berdasarkan analisis dari pembagian dataset yang digunakan, yaitu data latih dan data uji[41][31]. Kita dapat menggunakan entropy untuk menentukan percabangan di dalam pohon keputusan dengan rumus :

$$Entropy = \sum_{i=1}^C - P_i * \log_2(P_i) \quad (3)$$

f_i merupakan frekuensi dari label i pada sebuah node, C merupakan label dari angka unik[53]. kekurangan yang dimiliki oleh algoritma ini adalah strukturnya akan sangat kompleks untuk data dengan jumlah variabel yang banyak, sehingga beresiko overfitting[1].

2.3.2 Support Vector Machine (SVM)

Algoritma ini pada proses perhitungannya menggunakan prinsip Structural Risk Minimization (SRM). SVM kernel linear bekerja dengan sangat baik dibandingkan SVM kernel polynomial, kernel linear memiliki rumus :

$$F(X) = w^T * X + b \quad (4)$$

Pada persamaan diatas W merupakan nilai vektor yang ingin di minimalkan, X merupakan data yang ingin di klasifikasikan, dan b merupakan koefisien linier yang diestimasi.

SVM Kernel polynomial memiliki rumus :

$$F(X1, X2) = (a + X1^T * X2) ^ b \quad (5)$$

Merupakan kernel polynomial yang lebih sederhana dengan $f(X1, X2)$ mewakili Batasan sebuah keputusan polynomial sedangkan $X1$ dan $X2$ merupakan data yang akan digunakan[11][55].

2.3.3 Decision Tree (DT)

Decision Tree menggunakan struktur pohon keputusan pada proses perhitungan[11], decision tree bekerja dengan membagi menjadi dua atau lebih sebuah regresi model dalam word bentuk sebuah pohon, dengan cara memecah pohon inti ke dalam himpunan bagian yang lebih kecil hingga pohon keputusan tersebut akan berkembang secara bertahap.

$$E(S) = \sum_{i=1}^C - P_i \log_2 P_i \quad (6)$$

Dimana S merupakan kondisi awal, i merupakan set kelas pada S yaitu seperti yes, no P_i merupakan peluang suatu kejadian S . Manfaat dari decision tree adalah melakukan break down proses pengambilan keputusan yang kompleks menjadi simpel sehingga pengambilan keputusan akan lebih efektif[45].

2.3.4 Naïve Bayes (NB)

Naive Bayes merupakan algoritma klasifikasi sederhana menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai[20]. Naive bayes mengasumsikan bahwa kehadiran fitur tertentu di kelas tidak terkait dengan kehadiran fitur lainnya. Algoritma ini mudah dibuat dan sangat berguna apabila dihadapkan dengan dataset yang besar[1]. Naive Bayes adalah pendekatan statistik untuk inferensi yang diinduksi pada masalah klasifikasi[21].

$$P(x) = \frac{P(c) P(c)}{P(x)} \quad (7)$$

$P(c|x)$: merupakan probabilitas kelas (target) yang diberikan predictor (atribut)

$P(c)$: merupakan peluang pada kelas sebelumnya

$P(x|c)$: merupakan kemungkinan probabilitas kelas yang diberikan prediktor

$P(x)$: probabilitas sebelumnya dari predictor

2.3.5 Logistic Regression (LR)

Merupakan sebuah algoritma yang digunakan untuk mengukur hubungan antara variabel independen dan variabel dependen kategorik. Hal yang membedakan model regresi logistik dengan model regresi linier sederhana adalah pada

regresi logistik, variabel keluaran bersifat biner[52]. Logistic Regression dikategorikan menjadi dua, yaitu single logistic regression, dimana terdapat satu jenis variabel input, dan multiple logistic regression, di mana terdapat dua atau lebih variabel input, untuk memprediksi nilai variabel output. Dalam penelitian ini, model regresi yang digunakan adalah regresi logistik umum[56].

2.3.6 K-Nearest Neighbors (K-NN)

K-Nearest Neighbor merupakan algoritma untuk melakukan klasifikasi terhadap atribut berdasarkan data training yang menggunakan prinsip jarak tetangga terdekat dan memiliki kemiripan. Contoh pengujian pada data uji memiliki atribut yang paling mirip dengan atribut pada data latih dengan mempertimbangkan jarak terdekat. Saat proses klasifikasi variabel yang memiliki kemiripan akan dihitung untuk dijadikan data uji dan diambil sejumlah nilai 'K'[57].

$$dis(x_1, x_2) = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2} \quad (8)$$

Formula di atas dapat digunakan jika jumlah independent variabel hanya satu variabel, jika ada lebih dari satu maka menggunakan rumus di bawah :

$$dis = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2 + \dots} \quad (9)$$

KNN mempunyai karakteristik yang sederhana dan dapat bekerja berdasarkan jarak terpendek dari query instance ke data training sample untuk menentukan ketetangaan bagian-bagian tersebut akan berbentuk titik dan ditandai dengan kelas-kelas tertentu[58].

2.3.7 Gradient Boosting (GB)

Merupakan algoritma berbasis decision tree yang digunakan untuk proses klasifikasi dengan teknik boosting. ilustrasi algoritma ini dengan cara memasang setiap pohon dalam urutan ke residu pohon sebelumnya, seperti :

1. Sesuaikan Pohon keputusan dengan data : $F_1(x) = y$
2. Kemudian sesuaikan pohon keputusan dengan residu sebelumnya $h_1(x) = y - F_1(x)$
3. Tambahkan Pohon baru ke dalam algoritma : $F_2(x) = F_1(x) + h_1(x)$
4. Sesuaikan pohon keputusan berikutnya dengan residual dari $F_2 : H_2(x) + h_2(x)$
5. Lanjutkan proses ini sampai beberapa pohon hingga proses terhenti

Rumus yang dihasilkan menjadi :

$$f(x) = \sum_{b=1}^B f^b(x) \quad (10)$$

Rumus diatas menjelaskan yaitu predictor(x) memiliki hubungan gelombang sinus yang benar (garis biru). Tujuan dari ensemble adalah mengatasi kelemahan yang terdapat pada classifier yang satu dengan menggunakan kelebihan dari classifier yang lain[59][60].

2.3.8 Artificial Neural Network (ANN)

Artificial neural network (ANN) merupakan algoritma yang mengikuti konsep struktur sistem saraf makhluk hidup (neural systems), sistem saraf pada otak manusia mempunyai struktur dasar yang mirip seperti algoritma ini, yaitu terdiri dari beberapa simpul (nodes), yang bisa sebagai variabel masukan (input) dan keluaran (output)[47] Artificial neural network (ANN) dapat disimpulkan sebagai sebuah kesatuan fungsi yang dapat mengelompokkan variabel input ke variabel output dengan menggunakan bobot sebagai parameter penghubungnya.

2.4 Evaluasi Model

Evaluasi model yang digunakan yaitu model K-Fold Cross Validation memiliki tujuan untuk mengevaluasi kinerja model dimana dataset di pisah menjadi dua subset yaitu data uji dan data latih dimana algoritma yang digunakan akan diujikan dan dilakukan validasi, data tersebut dibagi menjadi himpunan bagian k dengan jumlah yang sama, lalu model akan diuji dan dilatih sebanyak nilai k, pada setiap pengulangan salah satu himpunan bagian akan digunakan sebagai data uji dan data latih. k fold cross validation dapat berpengaruh dalam mengurangi waktu komputasi dengan tetap menjaga keakuratan proses klasifikasi dan prediksi[53][54].

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Hasil akurasi features extraction terhadap 10 fitur menggunakan K-Fold Cross Validation sebanyak K=10 berikut hasil dari perhitungan menggunakan model K-Fold Cross Validation :

Tabel 2. Pengujian 10 field menggunakan K-Fold Cross Validation dengan nilai K = 10

Algoritma	Akurasi	Precision	Recall
Random Forest	71,5%	100%	71,5%
Decision Tree	63,2%	72,9%	73,9%

Algoritma	Akurasi	Precision	Recall
Naïve Bayes	55,4%	38,8%	95,3%
Logistic Regression	72%	93,3%	74,6%
Support Vector Machine	70,1	87,6%	88,5%
KNN	65,3%	78,1%	74,3%
Gradient Boosting	68%	83,4%	75,3%
Artificial Neural Network	68,4%	81,3%	75,6%

Berdasarkan tabel 2, dapat dilihat bahwa hasil ke delapan algoritma untuk menyelesaikan masalah klasifikasi penyakit liver terhitung belum begitu baik. Algoritma Logistic Regression menghasilkan akurasi yang lebih baik dari algoritma Random Forest, Decision Tree, Naive Bayes, Support Vector Machine, K-Nearest Neighbor Gradient Boosting dan Artificial Neural Network. Decision Tree memiliki akurasi sebesar 72%, selain akurasi logistic regression lebih tinggi ternyata algoritma logistic regression juga mampu mengklasifikasikan pasien yang menderita liver dengan jumlah yang lebih banyak dibanding algoritma lainnya.

Kemudian pada pengujian selanjutnya terhadap features extraction dengan model Principal Component Analysis (PCA) terhadap 10 fitur menggunakan K-Fold Cross Validation sebanyak K=10 berikut hasil dari perhitungan menggunakan model K-Fold Cross Validation :

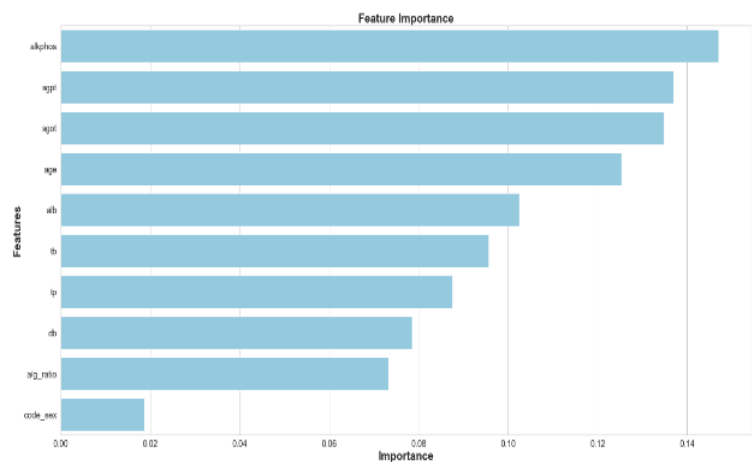
Tabel 3. Pengujian 10 field menggunakan features extraction model Principal Component Analysis (PCA)

Algoritma	Akurasi	Precision	Recall
Random Forest	70,9%	98,8%	71,5%
Decision Tree	61,8%	74,4%	72,6%
Naive Bayes	71,5%	99,3%	71,9%
Logistic Regression	71,3%	95,3%	72,8%
Support Vector Machine	69,4%	78,5%	80,2%
KNN	64,9%	81,4%	72,4%
Gradient Boosting	68,5%	88,8%	72,9%
Artificial Neural Network	71,3%	99,7%	71,4%

Dari pengujian di atas dapat disimpulkan bahwa penggunaan features extraction model PCA tidak dapat meningkatkan nilai akurasi

3.1.1 Random Forest Importance

Berdasarkan hasil akurasi yang didapat, terlihat dalam tabel ataupun grafik dengan menggunakan fitur yang ditentukan bahwa 10 fitur mencapai akurasi tertinggi pada data penyakit liver ini. Adapun grafik hasil nilai skor setiap fitur dari 10 yang digunakan. Hasil dari proses ini ditampilkan pada gambar 2. berikut hasil dari proses Random Forest Importance :



Gambar 2. Random Forest Importance

Table 4. Hasil menghitung peringkat menggunakan random forest importance

No	Features	Gini-Importance
1	Alkphos	0.146983
2	Age	0.136919
3	SGPT	0.134807
4	SGOT	0.125331
5	DB	0.102482
6	TP	0.095634

No	Features	Gini-Importance
7	ALB	0.087494
8	TB	0.078475
9	A/G ratio	0.073269
10	Code Sex	0.018606

Berikut perolehan hasil pada proses feature importance akan di observasi mengenai fitur atau atribut mana yang berpengaruh dan tidak berpengaruh terhadap tingkat peradangan dan kerusakan pada organ hati dengan berdasarkan kadar enzim. Berdasarkan dataset yang dimiliki, diperoleh atribut Alkaline Phosphatase, sgot dan sgpt yang merupakan feature tertinggi dari hasil feature importance.

3.1.2 Chi Square Test

Hipotesis awal menyatakan bahwa term t independen terhadap kategori c. Sedangkan hipotesis akhir menyatakan bahwa term t dependen terhadap kategori c. berikut hasil dari Chi Square Test:

Tabel 5. Hasil menghitung peringkat menggunakan Chi Square

No	Features	P-Value
1	Alkphos	0.000000e+00
2	SGPT	0.000000e+00
3	SGOT	0.000000e+00
4	TB	1,67526E-55
5	DB	1,85E-29
6	Age	6,40E-05
7	Alb	7,44E+04
8	A/G ratio	2,10E+05
9	Code Sex	3,57E+05
10	TP	6,36E+05

Daftar hasil perhitungan berdasarkan nilai seleksi fitur Chi Square yang berdasarkan hipotesis independensi, dengan hipotesis awal menyatakan bahwa term t independen terhadap kategori c. Apabila nilai seleksi fitur Chi Square lebih besar daripada nilai signifikan, sehingga penolakan hipotesis awal akan terpenuhi. Hipotesis akhir yang diperoleh menyatakan bahwa term t dependen terhadap kategori c.

3.1.3 Recursive Features Elimination

Sedangkan, dengan Recursive Feature Elimination (RFE) Seleksi fitur dilakukan dengan membuang peringkat terendah dari fitur. Kemudian, menggunakan 10 fitur untuk di evaluasi. Seperti yang dapat di lihat pada Tabel 6. berikut hasil dari Recursive Features Elimination :

Tabel 6. Hasil perhitungan peringkat menggunakan Recursive Features Elimination

No	Feature	Weight
1	Alkphos	0.163632
2	Age	0.136660
3	SGPT	0.136140
4	SGOT	0.135765
5	DB	0.112809
6	TP	0.079206
7	ALB	0.078133
8	TB	0.073955
9	A/G ratio	0.068741
10	Code Sex	0.014958

Dari hasil perhitungan features selection di atas kami menyimpulkan didapati dua fitur yang menempati ranking 4 teratas yaitu fitur SGOT dan SGPT, berikut hasil perhitungan akurasi menggunakan dua fitur tersebut. Evaluasi model menggunakan 2 field dengan K-Fold Cross Validation sebanyak K=10. Berikut hasil dari perhitungan :

Tabel 7. Pengujian 2 Dengan evaluasi model K-Fold Cross Validation menggunakan nilai K = 10

Algoritma	Akurasi	Precision	Recall
Random Forest	72,2%	99%	72,2%
Decision Tree	63,3%	74,8%	74,0%
Naïve Bayes	47,8%	28,3%	93,6%
Logistic Regression	72,1%	94,1%	71,5%
Support Vector Machine	71,4%	89,3%	86,2%

Algoritma	Akurasi	Precision	Recall
KNN	68,7%	84,5%	74,5%
Gradient Boosting	68,2%	86,4%	73,1%
Artificial Neural Network	71,5%	100%	71,5%

Berdasarkan Tabel 7 dapat dijelaskan bahwa nilai akurasi yang dihasilkan dari penggunaan 2 fitur model K-fold cross validation terhadap beberapa algoritma klasifikasi ini menghasilkan Random Forest nilai akurasi yang lebih baik daripada algoritma lainnya. Pada pengujian menggunakan 4 fitur dengan model K-fold cross validation RF memperoleh nilai akurasi sebesar 72,2%.

Tabel 8. Pengujian 2 menggunakan features extraction model Principal Component Analysis (PCA)

Algoritma	Akurasi	Precision	Recall
Random Forest	71,6%	99%	71,9%
Decision Tree	65,1%	76%	75%
Naïve Bayes	48,1%	29,3%	93,1%
Logistic Regression	71,5%	100%	71,5%
Support Vector Machine	70,2%	87,7%	85,3%
KNN	66,5%	83,9%	72,9%
Gradient Boosting	67,7%	86,4%	72,8%
Artificial Neural Network	71,5%	100%	71,5%

Berdasarkan Tabel 7 dapat dijelaskan bahwa nilai akurasi yang dihasilkan dari penggunaan 2 fitur model PCA terhadap beberapa algoritma klasifikasi ini menghasilkan Random Forest nilai akurasi yang lebih baik daripada algoritma lainnya. Pada pengujian menggunakan 2 fitur dengan model PCA RF memperoleh nilai akurasi sebesar 71,6%.

Berikut ini merupakan hasil perbandingan Hasil perhitungan akurasi menggunakan 10 field dengan 2 field menggunakan evaluasi model K-Fold Cross Validation :

Tabel 9. Perbandingan Akurasi 10 Field dengan 2 field K-Fold Cross Validation

Algoritma	Akurasi K Fold Cross Validation					
	Semua Field			2 Field		
	NO PCA	PCA	↑↓	NO PCA	PCA	↑↓
Random Forest	71,5%	70,9%	↓	72,2%	71,6%	↓
Decision Tree	63,2%	61,8%	↓	63,3%	65,1%	↑
Naïve Bayes	55,4%	71,5%	↑	47,8%	48,1%	↑
Logistic Regression	72%	71,3%	↓	72,1%	71,5%	↓
Support Vector Machine	70,1	69,4%	↓	71,4%	70,2%	↓
KNN	65,3%	64,9%	↓	68,7%	66,5%	↓
Gradient Boosting	68%	68,5%	↑	68,2%	67,7%	↓
Artificial Neural Network	68,4%	71,3%	↓	71,5%	71,5%	=

Tabel 10. Perbandingan akurasi 10 field dengan 2 field K Fold Cross Validation menggunakan model Principal Component Analysis (PCA)

Algoritma	Akurasi K Fold Cross Validation					
	No PCA			PCA		
	11 Field	2 Field	↑↓	11 Field	2 Field	↑↓
Random Forest	71,5%	72,2%	↑	70,9%	71,6%	↓
Decision Tree	63,2%	63,3%	↑	61,8%	65,1%	↑
Naïve Bayes	55,4%	47,8%	↓	71,5%	48,1%	↓
Logistic Regression	72%	72,1%	↑	71,3%	71,5%	↑
Support Vector Machine	70,1	71,4%	↑	69,4%	70,2%	↑
KNN	65,3%	68,7%	↑	64,9%	66,5%	↑
Gradient Boosting	68%	68,2%	↑	68,5%	68,7%	↓
Artificial Neural Network	68,4%	71,5%	↑	71,3%	71,5%	↑

Hasil penelitian kami menggunakan features selection mengalami peningkatan akurasi cukup signifikan dari beberapa algoritma yang digunakan sedangkan menggunakan features extraction dengan metode PCA akurasi yang didapat mengalami penurunan jika menggunakan model evaluasi K-Fold Cross Validation artinya untuk melakukan analisis prediksi perlu menggunakan 10 field yang terdapat di dalam dataset.

Jagdeep Singha et al pada penelitian mereka melakukan analisis prediksi penyakit hati menggunakan dataset yang sama, akan tetapi hasil dari features selection yang mereka gunakan mendapatkan 5 fitur penting diantaranya TB , DB, Alkphos, SGPT dan SGOT, Algoritma Logistic Regression mendapatkan akurasi dengan menggunakan 10 field sebesar

72,50%, dimana kelebihan pada penelitian kami dari hasil feature selection menggunakan 2 field yaitu SGOT dan SGPT mendapatkan akurasi tertinggi menggunakan algoritma Logistic Regression menghasilkan 72%, sedangkan jika menggunakan algoritma Random Forest menghasilkan akurasi tertinggi sebesar 72,5%, sedangkan penambahan pada penelitian kami dengan menerapkan features extraction menggunakan model Principal Component Analysis (PCA) dan hasil yang didapatkan adalah penurunan nilai akurasi.

4. KESIMPULAN

Berdasarkan penelitian yang sudah kami lakukan untuk memprediksi tingkat kronis penyakit menggunakan model evaluasi K-Fold Cross Validation menghasilkan akurasi tertinggi menggunakan algoritma Logistic Regression dengan menggunakan 10 field yaitu 72%, jika dilakukan features extraction dengan menggunakan PCA akurasi menurun menjadi 71,3% sedangkan menggunakan 2 field yaitu SGOT dan SGPT hasil dari features selection didapati akurasi meningkat menjadi 72,1% jika menggunakan PCA akurasi menurun menjadi 71,5%. Sedangkan akurasi tertinggi jika menggunakan 2 field yaitu algoritma Random Forest tanpa menggunakan PCA menghasilkan akurasi sebesar 72,2%. Hasil penelitian yang diusulkan bahwa teknik features selection tidak terlalu signifikan karena hasil akurasi yang didapat mengalami sedikit penurunan jika dibandingkan menggunakan 10 field untuk melakukan analisis prediksi penyakit hati.

REFERENCES

- [1] S. Kumar dan S. Katyal, "Effective Analysis and Diagnosis of Liver Disorder by Data Mining," Proc. Int. Conf. Inven. Res. Comput. Appl. ICIRCA 2018, hal. 1047–1051, 2018, doi: 10.1109/ICIRCA.2018.8596817.
- [2] S. Ambesange, A. Vijayalaxmi, R. Uppin, S. Patil, dan V. Patil, "Optimizing Liver disease prediction with Random Forest by various Data balancing Techniques," Proc. - 2020 IEEE Int. Conf. Cloud Comput. Emerg. Mark. CCEM 2020, hal. 98–102, 2020, doi: 10.1109/CCEM50674.2020.00030.
- [3] L. Meng, C. Wen, dan G. Li, "Support vector machine based liver cancer early detection using magnetic resonance images," 2014 13th Int. Conf. Control Autom. Robot. Vision, ICARCV 2014, vol. 2014, no. December, hal. 861–864, 2014, doi: 10.1109/ICARCV.2014.7064417.
- [4] G. Shaheamlung, H. Kaur and M. Kaur, "A Survey on machine learning techniques for the diagnosis of liver disease," 2020 Int. Conf. Intell. Eng. Manag. (ICIEM), London, UK, vol. pp. 337-34, no. International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 2020.
- [5] A. T. I. B. Siregar, "Gamabaran Kadar Serum Glutamic Pyruvic Transaminase (SGPT) Pada Perokok Berat Usia Diatas 40 Tahun," J. Chem. Inf. Model., hal. 25, 2018.
- [6] D. Zhang dan Y. Gong, "The Comparison of LightGBM and XGBoost Coupling Factor Analysis and Prediagnosis of Acute Liver Failure," IEEE Access, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3042848.
- [7] A. Sebastian dan S. M. Varghese, "Fuzzy logic for Child-Pugh classification of patients with cirrhosis of liver," Proc. - 2016 Int. Conf. Inf. Sci. ICIS 2016, hal. 168–171, 2017, doi: 10.1109/INFOSCI.2016.7845320.
- [8] K. Hamid, A. Asif, W. Abbasi, D. Sabih, dan F. U. A. A. Minhas, "Machine Learning with Abstention for Automated Liver Disease Diagnosis," Proc. - 2017 Int. Conf. Front. Inf. Technol. FIT 2017, vol. 2017-Janua, hal. 356–361, 2017, doi: 10.1109/FIT.2017.00070.
- [9] S. Thaiparnit, N. Chumuang, dan M. Ketcham, "A Comparitive Study of Clasification Liver Dysfunction with Machine Learning," 2018 Int. Jt. Symp. Artif. Intell. Nat. Lang. Process. iSAI-NLP 2018 - Proc., vol. 283, hal. 1–4, 2018, doi: 10.1109/iSAI-NLP.2018.8692808.
- [10] S. P. Deng dan W. L. Guo, "Identifying Key Genes of Liver Cancer by Networking of Multiple Data Sets," IEEE/ACM Trans. Comput. Biol. Bioinforma., vol. 16, no. 3, hal. 792–800, 2019, doi: 10.1109/TCBB.2018.2874238.
- [11] A. Sivasangari, B. J. Krishna Reddy, A. Kiran, dan P. Ajitha, "Diagnosis of liver disease using machine learning models," Proc. 4th Int. Conf. IoT Soc. Mobile, Anal. Cloud, ISMAC 2020, hal. 627–630, 2020, doi: 10.1109/I-SMAC49090.2020.9243375.
- [12] M. Sasso, V. Miette, dan L. Sandrin, "Novel Controlled Attenuation Parameter for the evaluation of fatty liver disease," hal. 2256–2259, 2010, doi: 10.1109/ulsym.2009.5441681.
- [13] I. Zubair dan B. Wajid, "Comparison of APRI, FIB-4 and fibro test in prediction of fibrosis and cirrhosis in patients with hepatitis C," Proc. 2018 15th Int. Bhurban Conf. Appl. Sci. Technol. IBCAST 2018, vol. 2018-Janua, hal. 222–227, 2018, doi: 10.1109/IBCAST.2018.8312227.
- [14] N. H. Lu, C. M. Kuo, dan H. J. Ding, "Automatic ROI segmentation in B-mode ultrasound image for liver fibrosis classification," Proc. - 2013 Int. Symp. Biometrics Secur. Technol. ISBAST 2013, hal. 10–13, 2013, doi: 10.1109/ISBAST.2013.4.
- [15] V. M. Bostan dan B. Pantelimon, "Creating a model based on artificial neural network for liver cirrhosis diagnose," 2015 9th Int. Symp. Adv. Top. Electr. Eng. ATEE 2015, hal. 295–298, 2015, doi: 10.1109/ATEE.2015.7133783.
- [16] K. Fujino et al., "A note of liver cirrhosis classification on M-mode ultrasound images by higher-order local auto-correlation features," 2013 Int. Conf. Soft Comput. Pattern Recognition, SoCPar 2013, hal. 50–53, 2013, doi: 10.1109/SOCPAR.2013.7054099.
- [17] W. Jung, C. Li, D. S. Kim, dan C. H. Ahn, "A sensing tube with an integrated piezoelectric flow sensor for liver transplantation," Proc. 31st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. Eng. Futur. Biomed. EMBC 2009, hal. 4469–4472, 2009, doi: 10.1109/IEMBS.2009.5333823.
- [18] T. M. Hassan, M. Elmogy, dan E. Sallam, "A classification framework for diagnosis of focal liver diseases," Proc. - 2015 10th Int. Conf. Comput. Eng. Syst. ICCES 2015, hal. 395–401, 2016, doi: 10.1109/ICCES.2015.7393083.
- [19] D. G. Melvin, M. Niranjana, R. W. Prager, A. K. Trull, dan V. F. Hughes, "Neuro-computing versus linear statistical techniques applied to liver transplant monitoring: A comparative study," IEEE Trans. Biomed. Eng., vol. 47, no. 8, hal. 1036–1043, 2000, doi: 10.1109/10.855930.
- [20] Rhyzoma Grannata Rafsanjani, N. Hidayat, dan R. K. Dewi, "Diagnosis Penyakit Hati Menggunakan Metode Naive Bayes Dan

- Certainty Factor,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, hal. 4478–4482, 2018.
- [21] H. Hartatik, M. B. Tamam, dan A. Setyanto, “Prediction for Diagnosing Liver Disease in Patients using KNN and Naïve Bayes Algorithms,” *2020 2nd Int. Conf. Cybern. Intell. Syst. ICORIS 2020*, hal. 1–5, 2020, doi: 10.1109/ICORIS50180.2020.9320797.
- [22] N. Ramkumar, S. Prakash, S. A. Kumar, dan K. Sangeetha, “Prediction of liver cancer using Conditional probability Bayes theorem,” *2017 Int. Conf. Comput. Commun. Informatics, ICCCI 2017*, hal. 7–11, 2017, doi: 10.1109/ICCCI.2017.8117752.
- [23] L. Guo et al., “CEUS-based classification of liver tumors with deep canonical correlation analysis and multi-kernel learning,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, hal. 1748–1751, 2017, doi: 10.1109/EMBC.2017.8037181.
- [24] T. I. Trishna, S. U. Emon, R. R. Ema, G. I. H. Sajal, S. Kundu, dan T. Islam, “Detection of Hepatitis (A, B, C and E) Viruses Based on Random Forest, K-nearest and Naïve Bayes Classifier,” *2019 10th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2019*, hal. 1–7, 2019, doi: 10.1109/ICCCNT45670.2019.8944455.
- [25] C. Geetha dan A. R. Arunachalam, “Evaluation based Approaches for Liver Disease Prediction using Machine Learning Algorithms,” *2021 Int. Conf. Comput. Commun. Informatics, ICCCI 2021*, hal. 55–58, 2021, doi: 10.1109/ICCCI50826.2021.9402463.
- [26] Y. Bei, P. Hong, F. Ke, Y. Zhou, L. Xu, dan M. Zhu, “Psychiatric insights into liver cirrhosis and their correlations with traditional Chinese medicine diagnostics,” *Proc. - 2013 IEEE Int. Conf. Bioinforma. Biomed. IEEE BIBM 2013*, hal. 279–284, 2013, doi: 10.1109/BIBM.2013.6732503.
- [27] M. Urdaneta dan P. Wahid, “A study on enhanced hyperthermia treatment for liver cancer using magnetic nanoparticles,” *IEEE MTT-S Int. Microw. RF Conf. 2014, IMaRC 2014 - Collocated with International Symp. Microwaves, ISM 2014*, hal. 255–258, 2014, doi: 10.1109/IMaRC.2014.7039021.
- [28] R. Alfiani dan Muljono, “Comparison of Naïve Bayes and KNN algorithms to understand hepatitis,” *Proc. - 2020 Int. Semin. Appl. Technol. Inf. Commun. IT Challenges Sustain. Scalability, Secur. Age Digit. Disruption, iSemantic 2020*, hal. 196–201, 2020, doi: 10.1109/iSemantic50169.2020.9234299.
- [29] J. Singh, S. Bagga, dan R. Kaur, “Software-based Prediction of Liver Disease with Feature Selection and Classification Techniques,” *Procedia Comput. Sci.*, vol. 167, no. 2019, hal. 1970–1980, 2020, doi: 10.1016/j.procs.2020.03.226.
- [30] S. Bahramirad, A. Mustapha, dan M. Eshraghi, “Classification of liver disease diagnosis: A comparative study,” *2013 2nd Int. Conf. Informatics Appl. ICIA 2013*, hal. 42–46, 2013, doi: 10.1109/ICoIA.2013.6650227.
- [31] Y. Sela et al., “fMRI-based hierarchical SVM model for the classification and grading of liver fibrosis,” *IEEE Trans. Biomed. Eng.*, vol. 58, no. 9, hal. 2574–2581, 2011, doi: 10.1109/TBME.2011.2159501.
- [32] A. Rosida, “Pemeriksaan Laboratorium Penyakit Hati,” *Berk. Kedokt.*, vol. 12, no. 1, hal. 123, 2016, doi: 10.20527/jbk.v12i1.364.
- [33] M. Hassoon, M. S. Kouhi, M. Zomorodi-Moghadam, dan M. Abdar, “Rule Optimization of Boosted C5.0 Classification Using Genetic Algorithm for Liver disease Prediction,” *2017 Int. Conf. Comput. Appl. ICCA 2017*, hal. 299–305, 2017, doi: 10.1109/COMAPP.2017.8079783.
- [34] H. Wang, Y. Liu, dan W. Huang, “The application of feature selection in Hepatitis B virus reactivation,” *2017 IEEE 2nd Int. Conf. Big Data Anal. ICBD 2017*, hal. 893–896, 2017, doi: 10.1109/ICBD.2017.8078767.
- [35] P. Saha, S. Patikar, dan S. Neogy, “A correlation - Sequential forward selection based feature selection method for healthcare data analysis,” *2020 IEEE Int. Conf. Comput. Power Commun. Technol. GUCON 2020*, hal. 69–72, 2020, doi: 10.1109/GUCON48875.2020.9231205.
- [36] C. M. Chen, C. Y. Hsu, dan C. H. Bai, “Building a patient oriented treatment decision system for liver cancer,” *Proc. - 2012 3rd Glob. Congr. Intell. Syst. GCIS 2012*, hal. 413–418, 2012, doi: 10.1109/GCIS.2012.11.
- [37] K. Zheng, Y. Geng, M. Jiang, X. He, dan G. Zheng, “Exploring basic treatment principles for liver cancer within Chinese herbal medicine through text mining,” *Proc. - 2014 IEEE Int. Conf. Bioinforma. Biomed. IEEE BIBM 2014*, no. 30825047, hal. 225–228, 2014, doi: 10.1109/BIBM.2014.6999364.
- [38] H. Cindric et al., “Retrospective Study for Validation and Improvement of Numerical Treatment Planning of Irreversible Electroporation Ablation for Treatment of Liver Tumors,” *IEEE Trans. Biomed. Eng.*, vol. 68, no. 12, hal. 3513–3524, 2021, doi: 10.1109/TBME.2021.3075772.
- [39] Z. Xiulian, Z. Shengliang, C. Yanmei, L. Chunfang, dan X. Tingting, “Progress in treatment of gastroesophageal reflux disease with liver-soothing and stomach-harmonizing Chinese herbal compound: A post-fundoplication case,” *Proc. - 2015 7th Int. Conf. Inf. Technol. Med. Educ. ITME 2015*, hal. 146–149, 2016, doi: 10.1109/ITME.2015.70.
- [40] L. Hao dan F. Xu, “An investigation on electronic nose diagnosis of liver cancer,” *Proc. - 2017 10th Int. Congr. Image Signal Process. Biomed. Eng. Informatics, CISP-BMEI 2017*, vol. 2018-Janua, hal. 1–5, 2018, doi: 10.1109/CISP-BMEI.2017.8302211.
- [41] S. Rajesh, N. A. Choudhury, dan S. Moulik, “Hepatocellular Carcinoma (HCC) Liver Cancer prediction using Machine Learning Algorithms,” *2020 IEEE 17th India Counc. Int. Conf. INDICON 2020*, no. C, 2020, doi: 10.1109/INDICON49873.2020.9342443.
- [42] H. Parveen dan S. Pandey, “Sentiment analysis on Twitter Data-set using Naive Bayes algorithm,” *Proc. 2016 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol. iCATccT 2016*, hal. 416–419, 2017, doi: 10.1109/ICATCCCT.2016.7912034.
- [43] D. Chicco dan G. Jurman, “Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone,” *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, hal. 1–16, 2020, doi: 10.1186/s12911-020-1023-5.
- [44] Y. Deta Kirana dan S. Al Faraby, “Sentiment Analysis of Beauty Product Reviews Using the K-Nearest Neighbor (KNN) and TF-IDF Methods with Chi-Square Feature Selection,” *Open Access J Data Sci Appl*, vol. 4, no. 1, hal. 31–042, 2021, doi: 10.34818/JDSA.2021.4.71.
- [45] S. S. Gavankar dan S. D. Sawarkar, “Eager decision tree,” *2017 2nd Int. Conf. Conver. Technol. I2CT 2017*, vol. 2017-Janua, hal. 837–840, 2017, doi: 10.1109/I2CT.2017.8226246.
- [46] H. Zhang et al., “Multi-Source Transfer Learning Via Multi-Kernel Support Vector Machine plus for B-Mode Ultrasound-Based Computer-Aided Diagnosis of Liver Cancers,” *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 10, hal. 3874–3885, 2021, doi: 10.1109/JBHI.2021.3073812.
- [47] K. Thirunavukkarasu, A. S. Singh, M. Irfan, dan A. Chowdhury, “Prediction of liver disease using classification Algorithms,” *2018 4th Int. Conf. Comput. Commun. Autom. ICCCA 2018*, vol. 6, no. 9, hal. 1–3, 2018, doi: 10.1109/CCAA.2018.8777655.
- [48] S. Gupta, G. Karanth, N. Pentapati, dan V. R. B. Prasad, “A Web Based Framework for Liver Disease Diagnosis using Combined

- Machine Learning Models,” Proc. - Int. Conf. Smart Electron. Commun. ICOSEC 2020, no. Icosec, hal. 421–428, 2020, doi: 10.1109/ICOSEC49089.2020.9215454.
- [49] D. S. Reddy, R. Bharath, dan P. Rajalakshmi, “Classification of nonalcoholic fatty liver texture using convolution neural networks,” 2018 IEEE 20th Int. Conf. e-Health Networking, Appl. Serv. Heal. 2018, hal. 1–5, 2018, doi: 10.1109/HealthCom.2018.8531193.
- [50] G. Cao, M. Li, C. Cao, Z. Wang, M. Fang, dan C. Gao, “Primary Liver Cancer Early Screening Based on Gradient Boosting Decision Tree and Support Vector Machine,” ICIIBMS 2019 - 4th Int. Conf. Intell. Informatics Biomed. Sci., hal. 287–290, 2019, doi: 10.1109/ICIIBMS46890.2019.8991441.
- [51] S. Visalakshi dan V. Radha, “A literature review of feature selection techniques and applications: Review of feature selection in data mining,” 2014 IEEE Int. Conf. Comput. Intell. Comput. Res. IEEE ICCIC 2014, no. 1997, 2015, doi: 10.1109/ICCIC.2014.7238499.
- [52] M. Emu, F. B. Kamal, S. Choudhury, dan T. E. Alves De Oliveira, “Assisting the Non-invasive Diagnosis of Liver Fibrosis Stages using Machine Learning Methods,” Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, vol. 2020-July, hal. 5382–5387, 2020, doi: 10.1109/EMBC44109.2020.9176542.
- [53] S. A. J. Zaidi, S. Tariq, dan S. B. Belhaouari, “Future prediction of covid-19 vaccine trends using a voting classifier,” Data, vol. 6, no. 11, 2021, doi: 10.3390/data6110112.
- [54] T. Hendrawati dan C. P. Yanti, “Analysis of Twitter Users Sentiment against the Covid-19 Outbreak Using the Backpropagation Method with Adam Optimization,” J. Electr. Electron. Informatics, vol. 5, no. 1, hal. 1, 2021, doi: 10.24843/jeei.2021.v05.i01.p01.
- [55] X. Z. Wang dan S. X. Lu, “Improved Fuzzy Multicategory Support Vector Machines Classifier,” Proc. 2006 Int. Conf. Mach. Learn. Cybern., vol. 2006, no. August, hal. 3585–3589, 2006, doi: 10.1109/ICMLC.2006.258575.
- [56] X. Zou, Y. Hu, Z. Tian, dan K. Shen, “Logistic Regression Model Optimization and Case Analysis,” Proc. IEEE 7th Int. Conf. Comput. Sci. Netw. Technol. ICCSNT 2019, hal. 135–139, 2019, doi: 10.1109/ICCSNT47585.2019.8962457.
- [57] S. Sun dan R. Huang, “An adaptive k-nearest neighbor algorithm,” Proc. - 2010 7th Int. Conf. Fuzzy Syst. Knowl. Discov. FSKD 2010, vol. 1, no. Fskd, hal. 91–94, 2010, doi: 10.1109/FSKD.2010.5569740.
- [58] Okfalisa, I. Gazalba, Mustakim, dan N. G. I. Reza, “Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification,” Proc. - 2017 2nd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2017, vol. 2018-Janua, hal. 294–298, 2018, doi: 10.1109/ICITISEE.2017.8285514.
- [59] P. Sheng, L. Chen, dan J. Tian, “Learning-based road crack detection using gradient boost decision tree,” Proc. 13th IEEE Conf. Ind. Electron. Appl. ICIEA 2018, hal. 1228–1232, 2018, doi: 10.1109/ICIEA.2018.8397897.
- [60] X. Yu et al., “Load Forecasting Based on Smart Meter Data and Gradient Boosting Decision Tree,” Proc. - 2019 Chinese Autom. Congr. CAC 2019, hal. 4438–4442, 2019, doi: 10.1109/CAC48633.2019.8996810.