

PENERAPAN ALGORITMA *RANDOM FOREST* UNTUK IDENTIFIKASI DEHIDRASI BERBASIS CITRA URINE

Niyalatul Muna¹, Faisal Lutfi Afriansyah², Ameng Bagus Suprayogy³

^{1,3}Rekam Medik, Kesehatan, Politeknik Negeri Jember

²Manajemen Informatika, Teknologi Informasi, Politeknik Negeri Jember
¹niyalatul@polije.ac.id, ²faisal.lutfi@polije.ac.id, ³amengbagus@gmail.com

Abstrak

Tingkat dehidrasi tidak hanya bisa dirasakan secara langsung akan tetapi dapat diamati dan dilihat secara fisik berbasis visual. Secara visual salah satu gejala dari dehidrasi dapat dilihat dari warna urine. Gejala ini biasanya tidak begitu diperhatikan dan dianggap biasa. Padahal gejala hipohidrasi atau dehidrasi merupakan dampak yang merugikan dari asupan air yang tidak memadai sehingga mempengaruhi warna urine yang dihasilkan. Kesulitan panca indra manusia membedakan gejala dehidrasi dan melihat perbedaan warna urine secara visual sering diterjemahkan berbeda-beda, dikarenakan tingkat kemiripan warna yang dihasilkan. Beberapa penelitian menunjukkan adanya pemanfaatan teknologi kamera dengan sistem cerdas dapat membantu kesulitan dan keterbatasan panca indra manusia. Penelitian ini menggunakan citra urine diambil dari sample orang dewasa yang dikelompokkan berdasarkan kategori warna urine hasil penelitian terdahulu. Pengambilan fitur dari setiap citra urine diambil nilai warna dari YCbCr. Model warna yang dihasilkan dari setiap sampel akan diidentifikasi menggunakan algoritma *Random Forest* dengan *cross-validation*. Hasil dari percobaan yang dilakukan menunjukkan akurasi 90% dari 30 dataset yang diujikan dengan nilai *precision* 90.2%, *recall* 90%, *Mean absolute error* 0.2473, dan *Root mean squared error* sebesar 0.3208.

Kata kunci : Citra Urine, *Cross-validation*, Dehidrasi, *Random Forest*, YCbCr.

1. Pendahuluan

Kekurangan cairan dalam tubuh akan mengakibatkan dampak yang buruk untuk metabolisme dan kesehatan. Agar terhindar dari hal tersebut, dibutuhkan asupan cairan yang dapat memenuhi kebutuhan setiap tubuh, Grandjean (2004). Fraser menyatakan tidak tercapainya asupan dan pengeluaran cairan dari dalam tubuh disebut dengan hidrasi. Hasil penelitian dari *The Indonesian Regional Hydration Study (THIRST)* menunjukkan bahwa 49,1% subyek yang diteliti mengalami kurang air atau hipovolemia ringan, pada orang dewasa sebesar 42,5%. Penelitian dari Kenefick menjelaskan bahwa ketika individu berada dalam keadaan terhidrasi atau dehidrasi, namun persepsi rasa haus tak terasa hampir 40% dalam suhu dingin dibandingkan dengan suhu lingkungan kerja yang normal, Ernovitania and Sumarni (2017).

Pengeluaran cairan sebagian besar melalui urine atau saluran pencernaan, serta sebagian lainnya melalui kulit dan paru-paru. Cairan yang hilang dan tidak diganti menyebabkan volume plasma menurun dan terjadi penurunan kemampuan fisik dan kognitif. Oleh karena itu, dengan adanya hidrasi yang baik merupakan salah satu cara yang efektif untuk menjaga kesehatan dan meningkatkan produktivitas. Ratih S, Annisa (2017).

Dehidrasi atau terhidrasinya seseorang yang paling mudah dapat dilihat dari perbedaan warna urine. Panca indra atau vision manusia dapat digunakan untuk mengamati dan mengidentifikasi suatu objek yang terlihat. Melalui mata setiap perbedaan objek dapat terlihat. Sehingga manusia dapat mengklasifikasi dan mengenali suatu objek. Akan tetapi, permasalahan yang biasa ditemukan pada vision manusia adalah muncul kesulitan membedakan warna ketika warna tersebut memiliki kemiripan antara satu warna dengan warna yang lain. Disamping itu kemampuan mengenali warna setiap orang dapat saja berbeda, sehingga akan terjadi perbedaan dalam mengidentifikasi warna setiap objek, Andrizal *et al.*(2018). Objek Citra yang memiliki kemiripan akan mempunyai nilai piksel yang sedikit berbeda dan ada juga yang sama pada titik-titik tertentu disetiap piksel. Setiap objek yang dibandingkan dapat mengalami pergeseran nilai yang sedikit berbeda sehingga warna yang dihasilkan bisa mirip atau hampir sama, Afriansyah *et al.* (2019).

Identifikasi warna dari citra pada umumnya memiliki model warna yang terdiri atas 3 warna dasar utama yang disebut RGB dengan kepanjangan *Red* (merah), *Green* (hijau), dan *Blue* (biru) (Blue). Dari tiga warna dasar dapat menghasilkan warna-warna lain dan model warna lainnya yang disesuaikan dengan warna dominan yang dihasilkan oleh sebuah citra. Penelitian ini menggunakan model warna yang

diambil dari ruang warna YCbCr. Model warna ini akan memisahkan citra ke dalam komponen luminance dan komponen warna. Komponen Y akan merepresentasikan luminance, dan komponen Cb dan Cr akan menginformasikan warna, Kolkur *et al.*, (2017).

Model warna YCbCr yang digunakan pada penelitian mengambil setiap komponen nilai dari setiap citra urine yang diambil, Suprayogy *et al.*, (2017). Berdasarkan selisih antara komponen *blue* dengan suatu nilai referensi yang termasuk dalam komponen Cb, dan selisih antara komponen *red* dengan suatu nilai referensi yang termasuk kedalam komponen Cr. Teknik ini dapat diperoleh dari ruang RGB dengan matrik sederhana, Kolkur *et al.*, (2017).

Dari hasil nilai ruang warna YCbCr akan diidentifikasi sesuai dengan kelasnya. Identifikasi kelas yang dilakukan menggunakan Algoritma *Random Forest* dengan teknik *cross-validation*. Algoritma yang digunakan merupakan salah satu algoritma klasifikasi dengan tingkat akurasi yang termasuk baik. Penelitian sebelumnya menunjukkan hasil yang dapat dijadikan pertimbangan dengan mengolah data maupun atribut diambil secara acak sehingga memungkinkan akan menghasilkan berbagai model pohon keputusan, Dewi *et al.*, (2011). Oleh karena itu, penelitian ini akan mengidentifikasi dehidrasi dan terhidrasi berbasis citra urine untuk menghasilkan akurasi yang sesuai dari berbagai model pohon keputusan yang diujicobakan.

2. Identifikasi Dehidrasi Berbasis Citra Urine

Citra Urine merupakan hasil citra dari akuisisi data yang diambil menggunakan kamera dari penampang yang berisi cairan urine. Citra diambil setelah pengambilan sampel urine dari masing-masing responden yang disimpan di wadah tertutup bening, proses pengambilan langsung dilakukan agar warna urine yang diperoleh sesuai dan tidak mengalami perubahan dikarenakan waktu yang lama untuk menyimpan sampel-sampel urine tersebut. Urine merupakan cairan yang keluar dari dalam tubuh sisa dari metabolisme tubuh yang diekskresikan oleh ginjal, Ramadhan dan Rismayanthi (2005). Cairan ini dikeluarkan untuk menjaga homeostasis cairan tubuh. Dari urine yang dikeluarkan dapat mendeteksi beberapa jenis penyakit ataupun mengetahui gejala kekurangan cairan dalam tubuh yang dapat berdampak buruk, Annisa Ratih S (2017).

Urine yang dihasilkan dari setiap individu bisa menghasilkan warna yang berbeda-beda atau warna yang mirip. Citra Urine yang sudah diambil dapat menunjukkan tingkat perbedaan dari setiap komponen warna yang dimiliki. Salah satu metode yang dapat digunakan untuk melihat ruang warna dari citra Urine adalah YCbCr.

Citra Urine yang didapatkan akan dilihat nilai RGB nya. Kemudian dari nilai RGB akan diubah

menjadi ruang warna YCbCr. Dimana persamaan yang dari penerapan YCbCr dapat dilihat pada persamaan (1).

$$\begin{aligned} Y &= 0.299R + 0.287G + 0.11B \\ Cr &= R - Y \\ Cb &= B - Y \end{aligned} \quad (1)$$

Informasi pencahayaan disimpan sebagai komponen tunggal (Y), dan informasi chrominance disimpan sebagai dua komponen perbedaan warna (Cb dan Cr). Komponen Cb mewakili perbedaan antara komponen biru dan nilai referensi. Komponen Cr mewakili perbedaan antara komponen merah dan nilai referensi.

2.1 Algoritma *Random Forest*

Pengklasifikasian dari algoritma *Random Forest* (RF) dengan cara pengelompokan ensemble yang menggunakan satu set CARTs untuk membuat prediksi. Pohon-pohon yang dibuat dengan menggambar subset sampel pelatihan melalui penggantian, Sugianto (2015). Penerapan yang dilakukan menunjukkan sampel yang sama bisa dipilih beberapa kali, sementara yang lain mungkin tidak dipilih sama sekali. Sekitar dua pertiga dari sampel (disebut sebagai sampel dalam kantong) digunakan untuk melatih pohon dengan sisa sepertiga digunakan dalam internal teknik *cross-validation* untuk memperkirakan seberapa baik hasilnya Performa model *Random Forest*, Breiman, (2001).

RF adalah klasifikasi yang terdiri dari kumpulan klasifikasi pohon yang terstruktur yaitu $\{h(x, \theta_k), k = 1, \dots\}$ di mana θ_k adalah vektor acak independen yang terdistribusi secara identik dan masing-masing pohon memberikan unit pilih kelas yang paling populer di input x .

Tahapan yang diterapkan pada algoritma adalah, Dewi *et al.*, (2011):

1. Melakukan penarikan sampel acak berukuran a dengan memulihkan rangkaian atau himpunan data. Tahapan ini biasanya disebut sebagai *bootstrap*.
2. Membangun pohon dengan sendirinya tidak dipaksa hingga maksimal. Pada setiap ikatan, memilah dan melakukan pemilihan dengan memilih b pengganti secara acak, dimana $b \ll p$. Pemilah terbaik dipilih dari b . Pengganti yang jelas pada tahapan ini adalah *random feature selection*.
3. Lakukan perulangan pada langkah 1 dan 2 sebanyak k kali, sehingga membentuk *Forest* terdiri dari k pohon.

Respons dari pengamatan diprediksi dengan menggabungkan hasil prediksi k pohon. Klasifikasi yang dilakukan prosesnya didasari dari melihat suara terbanyak.

2.2 Pengukuran Kinerja Algoritma Random Forest

Setiap sistem yang dibuat untuk melakukan identifikasi diharapkan menghasilkan akurasi yang terbaik. Akan tetapi, tidak semua sistem yang diujicobakan atau dijalankan akan memberikan hasil yang sesuai dengan harapan (mencapai keberhasilan 100%). Oleh karena itu, perlu diukur kinerja sebuah sistem dalam melakukan proses identifikasi atau klasifikasi. Pengukuran kinerja pada sebuah sistem umumnya menggunakan matriks konfusi. Pendekatan yang umum digunakan adalah mencatat hasil kinerja apakah sesuai atau tidak, E. Prasetyo (2012).

Untuk perhitungan dari identifikasi jumlah data dari masing-masing kelas yang diprediksi secara benar dapat menggunakan perhitungan nilai (m11 + m00), dan data yang classnya tidak sesuai, adalah (m01+m10). Dapat dihitung kualitas matrik konfusi yang dapat dihitung. Untuk mengetahui akurasi dapat digunakan formula (2) sebagai berikut:

$$\begin{aligned}
 \text{Akurasi} &= \frac{\text{Jumlah data yang diprediksi dengan benar /sesuai}}{\text{Jumlah prediksi yang dilakukan (banyaknya data)}} \\
 &= \frac{m11 + m00}{m00 + m01+m10 + m11} \quad (2)
 \end{aligned}$$

Untuk menghitung nilai error atau ketidaksesuaian hasil prediksi dapat menggunakan formula berikut:

$$\begin{aligned}
 \text{Error} &= \frac{\text{Jumlah data yang diprediksi tidak sesuai (salah)}}{\text{Jumlah prediksi yang dilakukan (banyaknya data)}} \\
 &= \frac{m01 + m10}{m00 + m01+m10 + m11} \quad (3)
 \end{aligned}$$

Ada beberapa metode yang biasa digunakan untuk klasifikasi berbasis statistika membuat penanganan dalam mengatasi masalah *overfitting* salah satunya adalah *Cross-validation*. Metode *cross validation* memiliki dasar dalam memvalidasi secara silang, pada setiap set data yang dilatih dan diuji harus saling *cross-over* sehingga setiap data mempunyai kesempatan untuk dilakukan validasi ulang atau pengecekan kembali. Bentuknya adalah *k-fold* merupakan validasi silang bentuk lain yang melibatkan ulangan perhitungan dari *k-fold cross-validation*, Rodriguez, et al (2010).

3. Metodologi

Bagian ini akan membahas metode yang digunakan dalam indentifikasi dehidrasi berbasis citra urine.

3.1 Akuisisi Citra

Sample data yang digunakan dalam penelitian adalah citra urine diambil menggunakan kamera dengan resolusi 48MP yang dapat menghasilkan citra 8.000x6.000 piksel. Citra urine diambil dari sample

orang dewasa yang mengkonsumsi air putih secara rutin atau cukup dan yang mengalami gejala dehidrasi atau kekurangan cairan dalam tubuh. Sampel tersebut diambil sejumlah 40 sample, dan citra urine yang didapatkan dikelompokkan sesuai standar warna urine dari penelitian terdahulu yang sudah terstandarisasi dari *Cleveland Clinic*, Andrizal, dkk (2018).

Dari 40 sampel yang ada yang diuji cobakan sebanyak 15 sampel dehidrasi dan 15 sampel terhidrasi. Citra Urine yang dihasilkan akan disamakan ukurannya menjadi 316x302 pixel. Proses pengambilan citra urine dari sampel cup diambil dengan jarak yang sama dan intensitas cahaya yang sama. Sehingga dalam proses akuisisi citra dapat mengurangi resiko data yang rusak atau berbeda sehingga akan mempengaruhi proses selanjutnya.

3.2 Pengolahan Data

Data yang akan diolah akan dilakukan proses preproesing awal. Tahap *Preproesing* dilakukan untuk memperbaiki akuisisi citra diawal yang perlu dilakukan standarisasi dan persamaan ukuran pixel dan persiapan dalam proses perubahan model warna yang akan diambil fitur nilai warnanya.

Setelah proses awal dilakukan Model Warna awal dari citra urine RGB akan diubah ke YCbCr untuk diambil pola warna yang dimiliki setiap sampel dari citra dehidrasi dan terhidrasi yang sudah diberi label. Kemudian akan dilakukan identifikasi menggunakan sistem dengan penerapana Algoritma *Random Forest*. Akurasi dari sistem akan dihitung berdasarkan sampel yang diujicobakan.

Adapun Hasil Pengolahan Data dari Model warna yang digunakan dapat dilihat pada tabel 1.

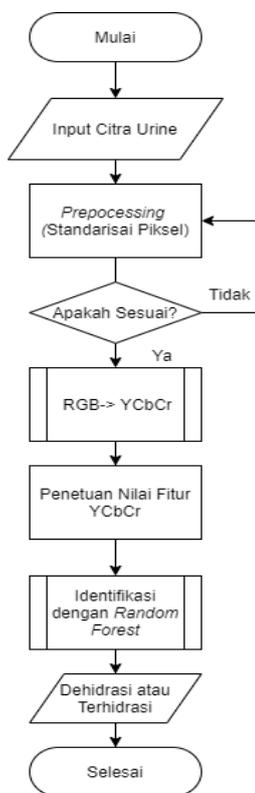
Tabel 1. Pengolahan Data Nilai Model Warna YCbCr

Y	Cb	Cr	Keterangan
77	76	152	Dehidrasi
108	56	163	Dehidrasi
118	55	159	Dehidrasi
123	47	163	Dehidrasi
127	45	158	Dehidrasi
134	47	165	Dehidrasi
149	44	163	Dehidrasi
125	45	156	Dehidrasi
93	38	158	Dehidrasi
130	45	158	Dehidrasi
112	49	156	Terhidrasi
129	41	155	Terhidrasi
133	36	156	Terhidrasi
136	45	147	Terhidrasi
126	45	154	Terhidrasi

Y	Cb	Cr	Keterangan
149	39	149	Terhidrasi
106	63	145	Terhidrasi
93	64	143	Terhidrasi
116	65	145	Terhidrasi
144	88	134	Terhidrasi

3.3. Perancangan Sistem

Alur utama dari sistem yang dibuat dalam penelitian dapat dilihat pada gambar 1. Secara rinci akan dijelaskan tahap-tahap sistem yang dibuat pada bagian ini.

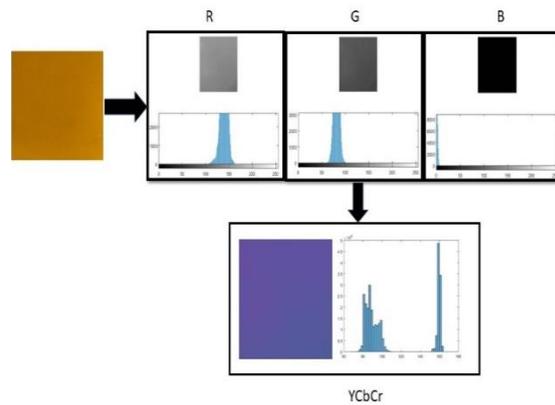


Gambar 1. Flowchat Sistem Identifikasi Dehidrasi

Input citra diambil dari proses akuisisi citra urine pada tahap pertama. Kemudian dilakukan proses perbaikan citra dengan melakukan standarisasi dan menghilangkan noise dari citra asli yang sudah diambil sehingga memiliki ukuran yang sama dan mengurangi ketidak jelasan citra. Akan dilakukan pengecekan pada preprocessing apakah sudah sesuai dengan standar. Jika sudah sesuai maka akan dilakukan penerapan ruang warna RGB dari masing-masing citra ke model warna YcbCr.

Proses perubahan model warna akan dilakukan penentuan nilai fitur dari masing-masing komponen warna YCbCr. Adapun contoh hasil perubahan model warna dapat dilihat pada gambar 2.

Dari Gambar 2 terlihat adanya perubahan warna dan nilai dari masing-masing komponen warna



Gambar 2. Perubahan Model Komponen Warna YCbCr

pada setiap Histogram warna. Setiap Nilai dari Histogram Warna YCbCr dari citra urine yang diujicobakan akan diambil penentuan fitur nilai dari setiap komponen warna.

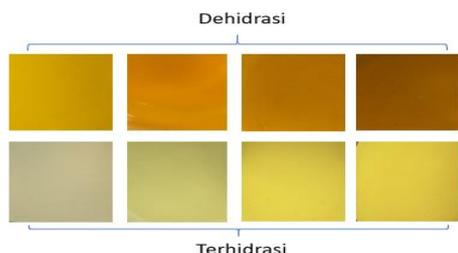
Setelah mendapatkan nilai dari setiap komponen warna akan dilakukan identifikasi terhadap kelas dehidrasi dan terhidrasi menggunakan algoritma *Random Forest*. Langkah-langkah dalam penerapan algoritma dengan metode yang dijelaskan pada 2.1. Setiap sampel data dari masing-masing komponen warna yang memiliki fitur nilai akan dibuat sampel acak berukuran 30x3 dengan rincian 30 baris dan 3 kolom yang akan dibentuk himpunan data dan membangun pohon dengan sendirinya tidak dipaksa hingga maksimal. Pada setiap ikatan atau himpunan, memilah dan melakukan pemilihan disesuaikan dengan besaran data hingga membentuk *Forest* terdiri dari 99 pohon.

Respons dari hasil pengamatan diidentifikasi menggabungkan hasil prediksi k pohon. Identifikasi yang dilakukan prosesnya didasari dari melihat suara terbanyak dari pengelompokan masuk ke kelas yang mana dehidrasi atau terhidrasi dari pengenalan hutan acak yang dibangun.

4. Hasil dan Pembahasan

Implementasi yang diujicobakan pada penelitian menggunakan 30 dataset yang terdiri dari 15 sampel dehidrasi dan 15 sampel terhidrasi. Adapun contoh beberapa sampel citra urine yang digunakan dapat dilihat pada gambar 3. Pengujian dilakukan dengan memvalidasi secara silang, setiap set data pelatihan dan data uji harus saling *cross-over* sehingga setiap data memiliki kesempatan untuk divalidasi ulang. Adapun hasil identifikasi dari dataset yang diujikan berdasarkan *10-folds cross validation* dapat dilihat pada tabel 1.

Dari hasil percobaan terdapat 3 dataset tidak sesuai dengan kelasnya dan 27 dataset yang sesuai dengan kelasnya. Hasil dapat dilihat pada tabel 2. Untuk Hasil Grafik Persebaran Algoritma *Random Forest* yang diterapkan dengan identifikasi dehidrasi dan terhidrasi dapat dilihat pada Gambar 4.



Gambar 3. Sampel Citra Urine

Tabel 2. Hasil Konfusi Matrik dengan 10-folds cross-validation

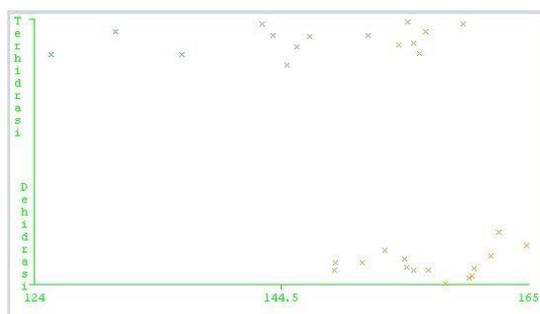
	Dehidrasi	Terhidrasi
Dehidrasi	13	2
Terhidrasi	1	14

Adapun dataset yang tidak sesuai terjadi dikarenakan ada beberapa sampel nilai yang mirip dan hampir sama antara fitur dehidrasi dan terhidrasi sehingga fitur yang diidentifikasi tidak sesuai dengan sampel yang telah dikodekan. Hasil dari pengujian akurasi sistem dapat dilihat pada tabel 2.

Tabel 3. Pengujian Citra Urine dari Sampel Dehidrasi dan Terhidrasi

Identifikasi	Accuracy	Precision	Recall
Dehidrasi	86.7%	92.9%	86.7%
Terhidrasi	93.3%	87.5%	93.3%

Dari tabel 2 dan 3 hasil pengujian memberikan nilai *Mean absolute error* sebesar 0.2473 dan nilai *Root mean squared error* sebesar 0.3208. Hasil ini menunjukkan adanya nilai relative error yang rendah yaitu *Relative absolute error* sebesar 48.9045 % dan *Root relative squared error* dengan presentase 63.4066 %.



Gambar 4. Grafik Persebaran Nilai Fitur Penerapan Algoritma Random Forest

5. Simpulan dan Saran

5.1 Simpulan

Hasil pengujian dataset dari fitur yang diperoleh memiliki nilai kemiripan yang berbeda. Pada kelas Dehidrasi komponen Y dan Cb memiliki rata-rata nilai fitur yang lebih kecil dari pada kelas

terhidrasi. Komponen Cr pada kelas dehidrasi memiliki rata-rata nilai yang lebih besar dari kelas terhidrasi.

Random Forest yang terbentuk dalam proses identifikasi dari setiap kelasnya dapat memberikan nilai akurasi yang baik, dimana hasil identifikasi untuk rata-rata mencapai 90%, nilai *precision* 90.2%, dan *recall* 90%. Dataset yang memiliki identifikasi paling baik dapat mengidentifikasi terhidrasi dengan nilai akurasi 93.3%. Hasil identifikasi dari dataset yang diujikan berdasarkan percobaan pada 10-folds *cross validation* dengan iterasi sebesar 100 dalam waktu 0.12 detik.

5.2 Saran

Saran untuk penelitian ini adalah perlu dilakukan percobaan lain dalam menentukan fitur dengan Model ruang warna yang lainnya. Penentuan fitur bisa dikombinasi dengan ruang warna yang berbeda menggunakan metode Pengolahan Citra Digital. Identifikasi dapat diujicobakan dengan metode Perhitungan Linier untuk proses klasifikasi dan hasilnya dapat dibandingkan untuk melihat hasil yang lebih baik.

Daftar Pustaka:

Afriansyah, F. L. et al. (2019) "Image Mapping Detection of Green Areas Using Speed Up Robust Features," *Proceedings - 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering, ICOMITEE 2019*. IEEE, 1, hal. 165–168.

Andrizal, A. et al. (2018) "Pembuatan Histogram Dan Pola Data Warna Urin Berdasarkan Urinalisis Menggunakan Mini PC," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 2(3), hal. 722–727.

Annisa Ratih S, F. F. D. (2017) "Hubungan Konsumsi Cairan Dengan Status Hidrasi Pekerja Di Suhu Lingkungan Dingin," *Journal Of Nutrition College*, 6(1), hal. 76–83.

Breiman, L. (2001) "Random forests," *Machine Learning, Kluwer Academic Publishers, Manufactured in The Netherlands.*, 45, hal. 5–32.

Dewi, N. K. et al. (2011) "Penerapan Metode Random Forest Dalam Driver Analysis," *Forum Statistika Dan Komputasi*, 16(1), hal. 35–43.

Ernovitania, Y. dan Sumarni, S. (2017) "Air Dengan Status Hidrasi Pada Siswi SMP Unggulan Bina Insani," *The Indonesian Journal of Public Health*, 12(August), hal. 276–285.

Kolkur, S. et al. (2017) "Human Skin Detection Using RGB, HSV and YCbCr Color Models," *Advances in Intelligent Systems Research, Published by Atlantis Press*, 137, hal. 324–332.

Ramadhan, R. I. dan Rismayanthi, C. (2005)

“Hubungan Antara Status Hidrasi Serta Konsumsi Cairan Pada Atlet Bola Basket,” *Medikora*, VX(April), hal. 53–69.

Rodriguez, Juan Diego, Aritz Pérez, and J. A. L. (2010) “Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3), hal. 569–575.

Sugianto, C. A. (2015) “Analisis Komparasi Algoritma Klasifikasi Untuk Menangani Data Tidak Seimbang Pada Data,” *Techno.COM*, 14(4), hal. 336–342.

Suprayogy, A. B. *et al.* (2017) “Analisis Nilai RGB dan YCBCR Pada Urine Untuk Mengetahui Tingkat Dehidrasi,” *Prosiding Seminar Rekam Medik Dan Informasi Kesehatan*, (1).