**Laporan Tugas Akhir MK. Pengantar Pengolahan Citra Digital (KOM332), Semester Genap 2016/2017**

**Indentifikasi Jenis Buah Berdasarkan Histogram Warna Buah dengan Menggunakan Algoritma SVM**

Faishal Hanin (G64150025), Muhammad Harits Arrazi (G64150092), Juan Josua (G64150088), Raymundus Jati Primanda (G64150106), M Daffa Rangga Putra (G64150109)

Kelompok: 10

**PENDAHULUAN**

**Latar Belakang**

Hortikultura berasal dari kata “*hortus*” (kebun) dan “*colere*” (budidaya). Secara harfiah istilah hortikultura diartikan sebagai usaha membudidayakan tanaman buah-buahan, sayuran dan tanaman hias. Golongan tanaman yang termasuk hortikultura secara umum adalah tanaman kebun yang mencakup buah, sayur, dan tanaman hias.

Buah merupakan tubuh reproduksi tumbuhan berbiji (*Spermatophyta*), yang merupakan hasil proses pembuahan; di dalam buah biasanya terdapat satu biji atau lebih; pada umumnya buah terdiri atas kulit buah, daging buah, kulit biji, dan biji yang mengandung lembaga. Buah-buahan biasanya dibedakan berdasarkan struktur dinding buah.

Setiap tipe buah, memiliki karakteristiknya masing-masing olreh karena itu untuk membedakan serta mengidentifikasi buah tersebut, terutama pada jenis-jenis buah dalam satu genus. Hal ini dikarenakan terdapat kemiripan antara satu jenis buah dan buah lainnya.

Pengolahan citra merupakan alternatif untuk persoalan ini. Teknologi pengolahan citra digital (*Digital Image Processing*) berkembang semakin pesat dengan adanya aplikasi dari teknologi di berbagai bidang. Pada umumnya, pengolahan citra dapat dikelompokkan dalam dua jenis kegiatan, yaitu perbaikan kualitas dalam citra, serta pengolahan informasi yang terdapat dalam citra. Untuk itu dalam penelitian ini dilakukan identifikasi citra buah dengan cara mengolah informasi warna yang terdapat dalam citra.

Pada proses pengolahan citra ini, dilakukan proses pembersihan *noise* pada citra sehingga didapat citra yang memiliki warna yang akurat untuk dapat diolah lebih lanjut menggunakan perhitungan histogram Hue Saturation and Intensity Value (HSV) dan penggunaan *Support Vector Machine* untuk melakukan klasifikasi terhadap buah apel berdasarkan hasil perhitungan histogram.

**Perumusan Masalah**

Rumusan permasalahan pada penelitian ini adalah bagaimana mengidentifikasi citra buah dengan melakukan identifikasi pada fitur warna buah dengan mencari sebaran nilai warna pada setiap citra gambar dan seberapa akurat identifikasi citra buah (*Apple Grannysmith*, *Apple Red*, *Apple*  *Golden*, *Lemon* dan *Mandarin Orange*) dengan menggunakan algoritme SVM.

**Tujuan**

Tujuan dari penelitian ini adalah melakukan identifikasi pada citra buah (*Apple Grannysmith*, *Apple Red*, *Apple Golden*, *Lemon* dan *Mandarin Orange*) menggunakan identifikasi dan klasifikasi SVM dengan sebaran warna *hue* dan *saturation* pada citra sehingga dapat mengetahui jenis buah yang tampak pada sebuah citra dari setiap jenis buah.

**Ruang Lingkup Penelitian**

Lingkup dari penelitian ini, yaitu:

1. Penerapan identifikasi jenis buah (*Apple Grannysmith*, *Apple Red*, *Apple Golden*, *Lemon* dan *Mandarin Orange*) dengan menggunakan *dataset* gambar yang sudah disediakan dengan jumlah 490 gambar pada setiap jenis buah.
2. Penerapan identifikasi jenis buah hanya dengan mengolah warna citra buah pada sebuah citra.
3. Buah yang dijadikan objek penilitian hanya terbatas pada jenis buah apel dan jeruk (*Apple Grannysmith*, *Apple Red*, *Apple Golden*, *Lemon* dan *Mandarin Orange*).

**TINJAUAN PUSTAKA**

**Median Filtering**

Ada dua cara yang dilakukan dalam peningkatan kualitas citra. Pertama adalah dengan *point processing*, dan kedua adalah *mask processing*. Jika pada *point processing* dilakukan operasi terhadap masing-masing piksel, maka pada *mask processing* dilakukan operasi terhadap suatu jendela ketetanggaan pada citra. Selanjutnya diterapkan (melakukan konvolusi) suatu mask terhadap jendela tersebut.

Salah satu penerapan *mask processing* adalah *median filtering*. *Median Filtering* merupakan salah satu teknik peningkatan kualitas citra dalam domain spasial. Metode ini termasuk dalam kategori *non linear filtering*. Cara kerjanya hampir sama dengan *mean filtering*. Pada *mean filtering* dalam setiap piksel output diatur ke tingkat rata-rata dari nilai-nilai piksel dalam mask yang ditentukan. Namun dengan *median filtering*, nilai piksel output ditentukan oleh median dari lingkungan mask yang ditentukan. (Muslim, 2009)

*Median filtering* merupakan *order-statistics filter* yang paling dikenal. Cara kerja dari filter ini dirumuskan pada gambar 1 berikut:



*Gambar 1 rumus median filtering*

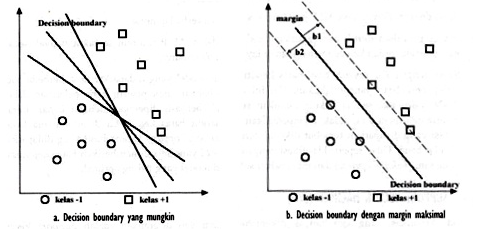
*Median filtering* mengambil area tertentu pada citra sesuai dengan ukuran *mask* yang telah ditentukan (umumnya 3×3), kemudian dilihat setiap nilai piksel pada area tersebut, dan nilai tengah pada area diganti dengan nilai *median*. Cara memperoleh nilai median adalah dengan menggunakan nilai keabuan dari titik-titik pada matriks diurutkan dari nilai terkecil hingga yang terbesar, kemudian ditentukan nilai yang berada di tengah dari deret piksel. Untuk tiep-tipe *noise* tertentu, *filter* ini memberikan kemampuan reduksi *noise* yang sangat baik, dengan *blurring* yang lebih sedikit daripada *linear smoothing filter* untuk ukuran citra yang sama. *Median filtering* memberikan hasil yang sangat bagus untuk citra yang terkena *noise impulse bipolar* dan *unipolar*.



*Gambar 2 Contoh median filtering*

**Support Vector Machine (SVM)**

Konsep klasifikasi dengan SVM dapat dijelaskan secara sederhana sebagai usaha untuk mencari *hyperplane*terbaik yang berfungsi sebagai pemisah dua buah kelas data pada ruang input. *Hyperplane* terbaik antara kedua kelas dapat ditemukan dengan mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya. *Margin* adalah jarak antara *hyperplane* tersebut dengan data terdekat dari masing-masing kelas. Data yang paling dekat ini disebut *support vector*. Garis pada gambar sebelah kanan menunjukkan *hyperplane* terbaik, yaitu yang terletak tepat pada tengah-tengah kedua kelas, sedangkan data lingkaran dan bujur sangkar yang dilewati garis batas *margin* (garis putus-putus) adalah *support vector*. Usaha untuk mencari lokasi *hyperplane* ini merupakan inti dari proses pelatihan pada SVM. (Nugroho, 2003)



*Gambar 3 Hyperplane*

**Color Histogram**

*Color histogram* adalah representasi penyebaran warna pada suatu citra. Pada citra digital, *color histogram* merepresentasikan banyaknya piksel dengan sebuah *color range*, yang meliputi *color space* pada sebuah citra. *Color histogram* dapat dibangun pada jenis *color space* apapun sepert RGB atau HSV. *Color Histogram* hanya berfokus pada proporsi dari jenis-jenis warna yang berbeda, terlepas dari lokasi spasial warna suatu pixel. (Bashkov, 2002)

**METODE**

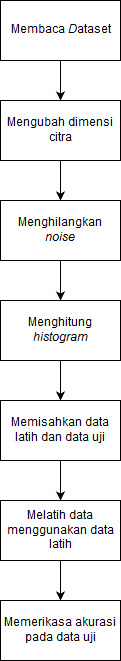
**Data**

Data yang digunakan pada penelitian ini adalah gambar dari lima jenis buah yang berbeda (*Apple Grannysmith*, *Apple Red*, *Apple Golden*, *Lemon* dan *Mandarin Orange*) dengan total 300 gambar untuk masing-masing jenis buah. Data ini dapat diperoleh dari GitHub milik Horea94 dengan repositories bernama *fruit-image-dataset* (<https://github.com/Horea94/Fruit-Images-Dataset>), data ini dikumpulkan oleh Horea Muresan dan Mihai Oltean. Salah satu contoh citra dataset yang digunakan disajikan pada Gambar 4.



*Gambar 4 Contoh gambar dari dataset fruit-images-dataset*

**Tahapan Penelitian**

****

*Gambar 4 Tahapan penelitian*

* *Pre-processing*

Tahapan *pre-processing* yang dilakukan adalah penghilangan *background* citra, penghilangan *noise* dan perubahan dimensi citra. *Dataset* citra yang digunakan pada penelitian ini sudah dilakukan penghilangan *bacground sebelumnya* sehingga *pre-processing* yang dilakukan hanya merubah dimensi citra*,* serta penghilangan *noise.* Pada penelitian ini, penghilangan *noise* dilakukan dengan metode *median filter*.

* Ekstraksi fitur

Selanjutnya, citra dimasukkan ke dalam tahap ekstraksi fitur. Ekstraksi fitur dilakukan agar mendapatkan nilai nilai khusus yang mewakili karakteristik unik objek yang akan diteliti. Ekstraksi fitur penelitian ini ada dua yaitu penghitungan *hue histogram* dan *saturation histogram*.

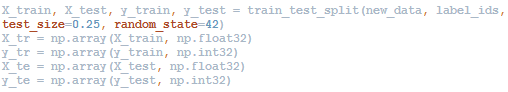
Penghitungan *hue histogram* dan *saturation histogram* menggunakan *cv2 library* pada PyCharm.



*Gambar 4 fungsi hue dan saturation histogram*

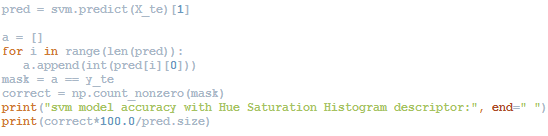
* Klasifikasi

Klasifikasi adalah suatu proses pengelompokan sebuah data dengan data lain yang memiliki karakteristik yang sama. Proses klasifikasi penelitian ini menggunakan algoritme SVM. Penelitian ini membandingkan hasil akurasi proses klasifikasi berdasarkan algoritme SVM. Semua data nantinya digunakan sebagai data latih dan data uji.



*Gambar 5 fungsi pembagian data train dan data test*

Setelah data terbagi menjadi menjadi data latih dan data uji, selanjutnya adalah melakukan uji pada data uji dan menghitung tingkat akurasi.

****

*Gambar 6 melakukan uji pada data latih dan menghitung akurasi*

**Lingkungan Pengembangan**

1. **Perangkat Keras :**

* Laptop Asus model X465UF (Intel Core i5, RAM 4GB)

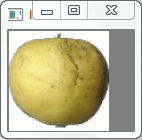
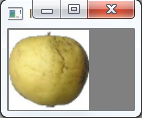
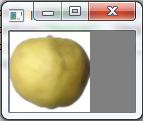
1. **Perangkat Lunak :**

* Sistem Operasi: Windows 10
* Python2.7
* OpenCV 3.4
* PyCharm Edu 4.0
* Django 1.11
* Atom

**HASIL DAN PEMBAHASAN**

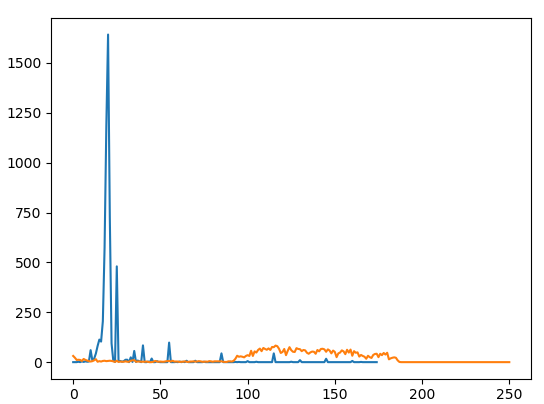
Dataset yang digunakan pada penelitian ini berasal dari GitHub milik Horea94 dengan repositories bernama *fruit-image-dataset* dengan mengambil 5 jenis buah yaitu *Apple Grannysmith*, *Apple Red*, *Apple Golden*, *Lemon* dan *Mandarin Orange* dengan total gambar 490 gambar pada setiap jenis buah.

Sebagai contoh citra yang ditampilkan pada hasil penelitian ini adalah citra buah dari *Apple Golden*. Citra yang digunakan pada penelitian ini keadaannya berlatar putih, sehingga tidak perlu pemisahan citra dengan latar. Karena tidak ada tahap pemisahan objek dengan latar, maka proses pengolahan citra dilanjutkan dengan perubahan dimensi citra menjadi 80\*80 p*ixel*, kemudian dilakukan *filtering* menggunakan *median filtering*. Penggunaan *filtering* bertujuan untuk mengurangi *noise* pada gambar, sehingga ekstraksi *hue histogram*  dan *saturation* *histogram* bisa memunculkan hasil yang lebih baik.

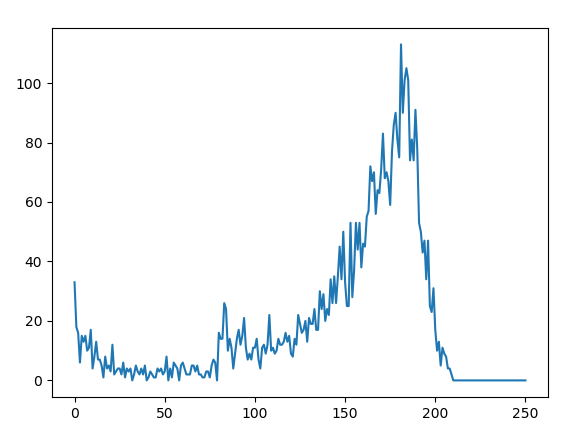
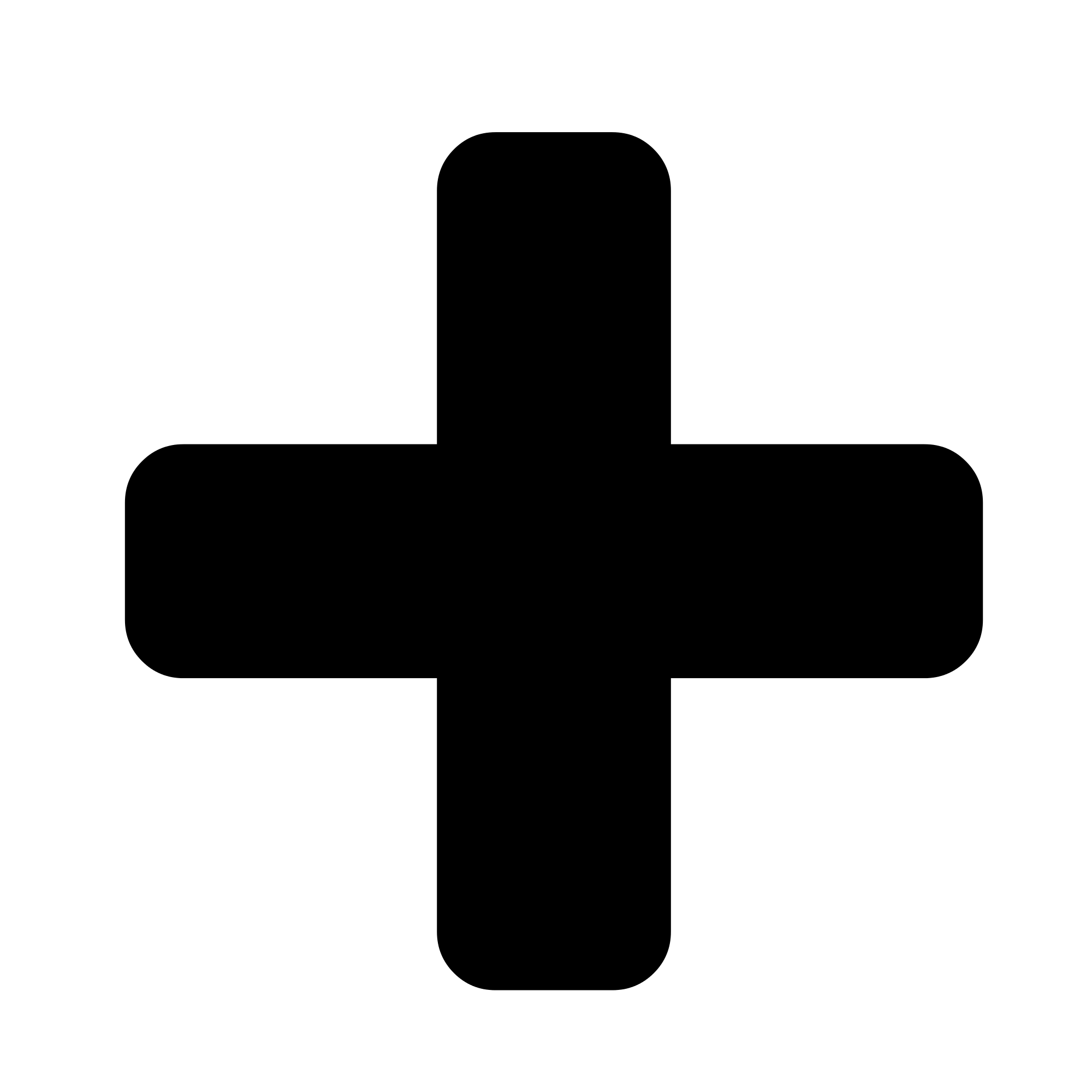
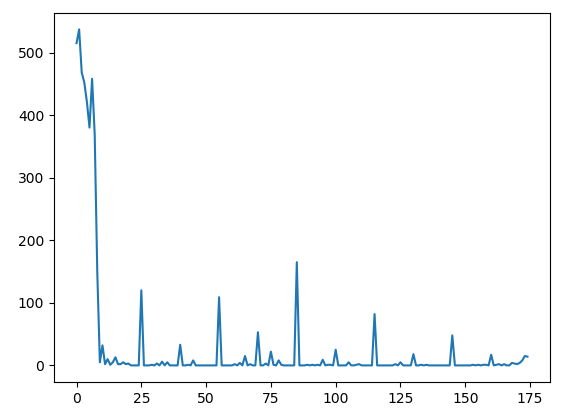
***Gambar 7 Citra awal (paling kiri), citra hasil resize (tengah), dan citra setelah dilakukan proses filtering menggunakan median filtering (paling kanan)***

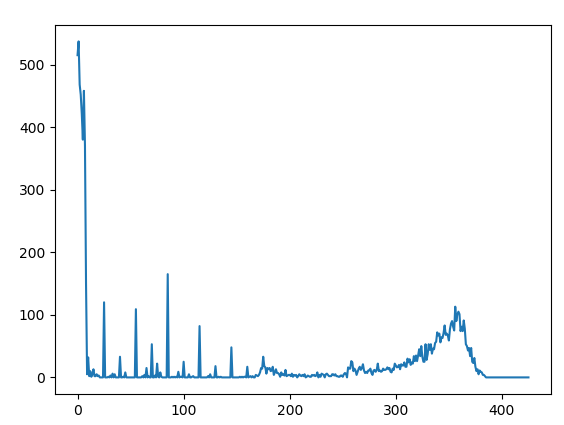
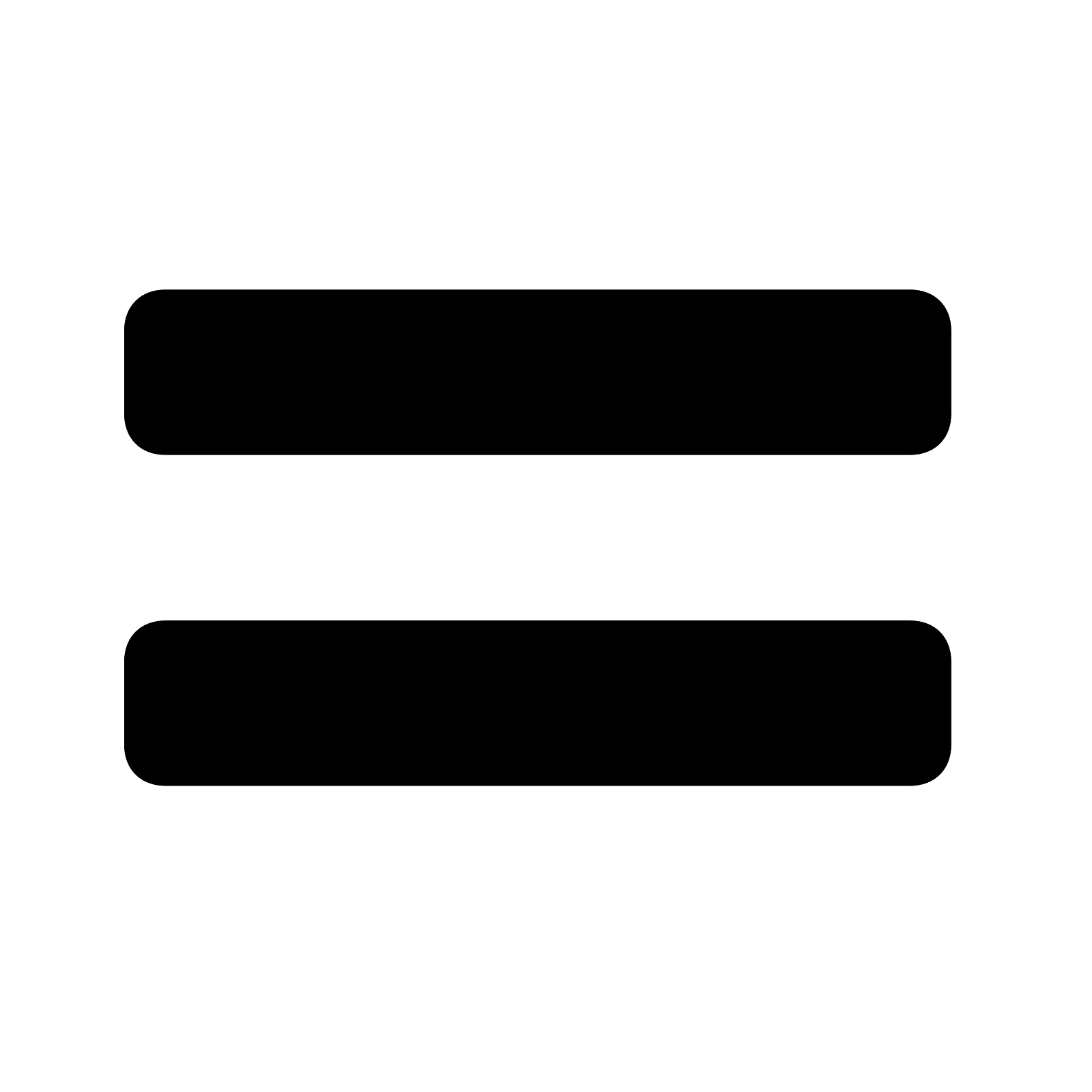
Tahap selanjutnya adalah menghitung nilai *hue histogram* dan *saturation histogram.* Warna pada citra diekstrak nilai *hue* dan nilai *saturation-*nyasehingga didapatkan nilai yang nantinya dapat digunakan untuk proses klasifikasi menggunakan algoritme SVM. Contoh hasil dari perhitungan nilai *hue histogram* dan *saturation histogram* untuk buah *Apple Golden* dapat dilihat pada gambar 8.



***Gambar 8 Contoh hasil perhitungan nilai hue histogram dan saturation histogram***

Perhitungan nilai *hue histogram* dan *saturate histogram* dilakukan untuk semua data, sehingga memunculkan nilai-nilai ekstraksi dari tiap buah. Lalu dilakukan penyambungan pada dua histogram tersebut(*concatenation*). Penyambungan dilakukan agar data dapat dimasukkan ke dalam algoritma SVM yang dimiliki opencv. Contoh penyambungan histogram untuk salah satu gambar *apple red* ditunjukan oleh gambar 9.





***Gambar 9 contoh hasil concatenation antara hue histogram dan saturation histogram***

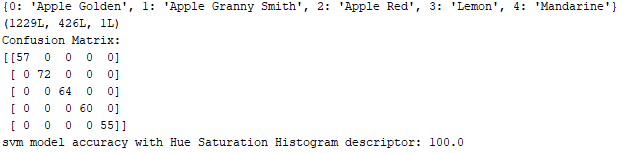
Tahap selanjutnya adalah pembagian data hasil ekstraksi tersebut menjadi data uji dan data latih dengan proporsi 75% sebagai data latih dan 25% sebagai data uji. Pemisahan data menggunakan fungsi *train\_test\_split* yang disediakan oleh library scikit-learn Potongan kode untuk memisah data menjadi data test dan train ditunjukan pada gambar 10.



***Gambar 10 Memisah data menjadi data Train dan Test***

Pada tahap pembuatan model, data yang sudah dibagi ini dimasukkan ke dalam fungsi SVM yang terdapat pada library opencv. Algoritma SVM akan membuat *decision boundary* dengan dimensi yang sesuai dengan panjang vector masukan(dalam hal ini 426) yang akan digunakan untuk membagi data antar kelas. Pada algoritma SVM multiclass pada opencv, pendekatan yang dilakukan berbasis *one vs one*, dimana tidak terdapat satu *classifier* tunggal untuk membagi data, tetapi N(N-1)/2 buah classifier dimana N merupakan jumlah kelas yang ada. Pendekatan *one vs one* maksudnya adalah masing-masing *classifier* akan menentukan sebuah data merupakan satu kelas tertentu atau satu kelas tertentu lainnya (mis. apple red atau lemon), lalu hasil atau output akan ditentukan melalui voting dari masing-masing *classifier*.

Dari hasil perhitungan algoritma SVM didapatkan model klasifikasi dengan akurasi 100% jika diuji menggunakan data uji yang sudah dipisah sebelumnya. Gambar *Confusion matrix* dan hasil prediksi dapat dilihat pada gambar 11.



*Gambar 11 hasil akurasi model ketika diuji menggunakan data uji yang sudah dipisahkan sebelumnya.*

Setelah berhasil diuji pada data uji asli dari dataset tersebut, kemudian dicoba untuk pengujian menggunakan data citra buah lain yang didapat dari luar dataset dengan jenis buah yang sama untuk melakukan pengujian ulang. Kami melakukan beberapa pengujian dengan mengambil gambar buah dengan latar putih dari *google* lalu memasukkannya pada model yang sudah dibuat. Saat dilakukan pengujian tersebut, ternyata klasifikasi kelas yang muncul cenderung merujuk kepada Apel *GrannySmith*. Contoh salah pengklasifikasian pada gambar dari luar ditunjukan pada gambar 12(kelas seharusnya adalah 0 atau *apple gold*).



*Gambar 12 contoh salah pengklasifikasian oleh model pada data yang didapat dari luar dataset*

**KESIMPULAN DAN SARAN**

**Kesimpulan**

Berdasarkan hasil analisis yang dilakukan, dapat diketahui nilai akurasi yang didapat bila model klasifikasi dikenakan pada data uji sebesar 100%. Namun saat model diujikan pada data baru yang berasal dari luar, hasil kelas yang didapatkan adalah kebanyakan Apel *GrannySmith*. Yakni menunjukkan bahwa error model yang didapatkan terlalu besar jika model digunakan pada data dari luar.

Hal ini dapat dikatakan bahwa dataset yang digunakan tidak bagus atau tidak direkomendasikan untuk dipakai dalam mengklasifikasi buah dengan meng-ekstrak fitur warna pada gambar. Dapat dikatakan juga bahwa jenis ekstraksi citra (*Color Histogram*) ini tidak cocok untuk digunakan pada dataset tersebut.

**Saran**

Untuk penelitian berikutnya, disarankan mengambil dataset yang bagus dan bervariasi buahnya pada setiap jenis buah atau membuat dataset secara manual dengan mengambil gambar sendiri tiap jenis buah yang digunakan. Penelitian berikutnya dapat mencoba algoritme klasifikasi lain selain SVM untuk membandingkan nilai akurasi antara kedua algoritme tersebut. Serta dapat juga dicoba untuk meng-ekstrak fitur citra lain selain *Color Histogram*, untuk melihat apakah ada jenis ekstraksi fitur yang layak digunakan pada studi kasus ini.

**DAFTAR PUSTAKA**

Bashkov E.A, Shozda N.S. 2002. *Content-Based Image Retrieval Using Color Histogram Correlation*. Graphicon proceedings(RU). Donetsk National Technical University.

Muresan H, Mihai Oltean. 2017. *Fruit recognition from images using deep learning*. Cluj-Napoca(RO). Babes-Bolyai University.

Muslim Krisna. 2009. *Median Filtering* [Internet]. Tersedia pada: <http://belajar-share.blogspot.com/2009/02/median-filtering.html>

Nugroho A, Witarto A, Handoko Dwi. 2003. *Support Vector Machine - Teori dan Aplikasinya dalam Bioinformatika* [Internet].[diunduh 2018 Juni 2]. Tersedia pada: <http://asnugroho.net/papers/ikcsvm.pdf>