



Perbandingan Metode Random Forest, Convolutional Neural Network, dan Support Vector Machine Untuk Klasifikasi Jenis Mangga

Ricky Mardianto¹, Stefanie Quinevera^{2*}, Siti Rochimah³

¹²³Teknik Informatika, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut Teknologi Sepuluh Nopember, Indonesia
¹ricky_mardianto@widyadharma.ac.id, ²stefani_quinevera@widyadharma.ac.id, ³siti@if.its.ac.id

Abstract

Mango is a fruit known as the "King of Fruit" due to its rich flavor, vast variability, and high nutritional value. Classifying mangoes based on their external appearance is the initial step in the process of identifying and categorizing mango types conventionally. The classification process can be performed by examining external features such as fruit color, shape, and size. Classifying different types of mango fruits accurately can assist researchers in developing superior varieties and also aid farmers for cultivation purposes, sales, distribution, and selecting the right varieties for local growth and weather conditions. This research conducts the classification of mango types based on color from mango images using machine learning. The study compares three methods, namely Random Forest, Support Vector Machine (SVM), and Convolutional Neural Network (CNN), to determine the best method for classifying mango types based on their images. The dataset underwent preprocessing, where image sizes were standardized to 300 x 300 pixels, and color was changed to grayscale. The dataset was then divided into training and testing data with a ratio of 70:30. Subsequently, the dataset was processed using three methods, and their accuracy results were compared. The findings indicate that the Random Forest method yielded the highest accuracy compared to the other methods, with an accuracy rate of 96%. The accuracy of the SVM method was 95%, and the accuracy of the CNN method was 33%. From these results, it can be concluded that the Random Forest method is highly effective for classifying mango types based on their image compared to SVM and CNN methods.

Keywords: classification, image processing, mango, Convolutional Neural Network, Random Forest, Support Vector Machine

Abstrak

Mangga merupakan buah yang dinamakan "King of Fruit" karena kekayaan rasanya, variabilitas yang besar, dan nilai gizi yang tinggi. Klasifikasi buah mangga berdasarkan tampilan luar merupakan langkah awal dalam proses mengidentifikasi dan mengklasifikasikan jenis mangga secara konvensional. Proses klasifikasi dapat dilakukan dengan melihat tampilan luar seperti dari warna buah, bentuk, maupun ukurannya. Pengklasifikasian jenis buah mangga dengan tepat dapat membantu peneliti untuk mengembangkan varietas unggul dan juga petani untuk tujuan pembudidayaan, penjualan, distribusi, dan seleksi varietas yang tepat untuk pertumbuhan dan cuaca lokal. Penelitian ini melakukan proses klasifikasi jenis mangga berdasarkan warna dari citra mangga menggunakan *machine learning*. Penelitian ini membandingkan tiga metode, yaitu metode *Random Forest*, *Support Vector Machine* (SVM), dan *Convolutional Neural Network* (CNN) untuk mendapatkan metode yang terbaik dalam melakukan klasifikasi jenis mangga berdasarkan citranya. Dataset dilakukan tahap pra-pemrosesan dimana citra gambar dilakukan penyeragaman ukuran citra menjadi ukuran 300 x 300 *pixel* dan dilakukan perubahan warna menjadi skala abu-abu (*grayscale*). Dataset tersebut kemudian dibagi menjadi data latih dan data uji dengan rasio perbandingan 70:30. Kemudian dataset diproses menggunakan tiga metode dan dibandingkan hasil akurasi. Hasil yang didapat menunjukkan bahwa metode *Random Forest* menghasilkan nilai akurasi terbaik dari metode lain yang dibandingkan, dengan nilai akurasi sebesar 96%. Sedangkan nilai akurasi dari metode SVM sebesar 95% dan nilai akurasi dari metode CNN yang sudah dilakukan normalisasi sebesar 33%. Dari hasil penelitian tersebut dapat diambil kesimpulan bahwa metode *Random Forest* sangat baik untuk melakukan klasifikasi jenis mangga berdasarkan citra gambarnya dibandingkan dengan metode SVM dan CNN.

Kata kunci: citra gambar, klasifikasi, mangga, Convolutional Neural Network, Random Forest, Support Vector Machine

1. Pendahuluan

Mangga merupakan salah satu buah yang memiliki sumber karotenoid terbaik dan memberikan kandungan vitamin A yang tinggi [1][2]. Mangga dinamakan "King of Fruit" karena kekayaan rasanya, variabilitas yang besar, dan nilai gizi yang tinggi [1][2]. Dengan kandungan vitamin dan gizi yang tinggi, mengonsumsi

buah mangga sangat baik untuk tubuh [3]. Buah mangga menjadi salah satu buah yang populer di dunia, salah satunya di negara Pakistan [4]. Pakistan merupakan negara produsen mangga terbesar kelima di dunia dan eksportir terbesar keempat dalam perdagangan mangga global [5]. Beberapa jenis mangga di Pakistan yang terkenal antara lain adalah Dosehri, Chaunsa, Sindhri,



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

Langra, Fajri dan Anwar Ratool [1][4][6][7]. Beberapa jenis mangga memiliki ciri-ciri fisik yang mirip, seperti bentuk dan warna buah yang hampir sama antara jenis yang satu dengan yang lainnya [8].

Klasifikasi mangga berdasarkan tampilan luar merupakan langkah awal yang penting dalam proses mengidentifikasi dan pemilihan jenis buah. Untuk mengenali dan mengklasifikasikan berbagai jenis mangga dapat dilakukan dengan cara konvensional, yaitu dengan melihat bentuk dan warna mangga tersebut. Seiring dengan perkembangan zaman, pengklasifikasian jenis mangga dapat dilakukan pengolahan citra menggunakan *machine learning*. Pemilihan metode yang tepat dapat memberikan hasil pengklasifikasian yang lebih baik. Oleh karena itu, diperlukan pemilihan metode yang tepat untuk menentukan jenis mangga berdasarkan citra dari buah mangga tersebut.

Penelitian tentang klasifikasi mangga menggunakan *machine learning* relatif baru dan mulai mendapatkan sorotan dalam beberapa tahun terakhir [9][10]. Pemanfaatan kecerdasan buatan dan algoritma *machine learning* dapat membantu meningkatkan akurasi klasifikasi mangga berdasarkan gambar [9][10]. Mengidentifikasi jenis mangga dapat dilakukan berdasarkan bentuk, ukuran, dan warna kulitnya [8]. Permasalahan dalam klasifikasi jenis mangga melalui tampilan luar merupakan kesulitan dalam mengidentifikasi jenis mangga yang serupa secara fisik. Beberapa jenis mangga dapat memiliki ciri-ciri yang mirip, seperti bentuk, ukuran, dan warna buah yang hampir sama, sehingga sulit untuk membedakan mereka hanya berdasarkan penampilan luar. Hal ini dapat menjadi masalah bagi peneliti untuk mengembangkan varietas unggul dan petani karena memerlukan identifikasi varietas yang sesuai untuk tujuan pembudidayaan, penjualan, distribusi, dan seleksi varietas yang tepat untuk pertumbuhan dan cuaca lokal.

Kegiatan klasifikasi mangga melibatkan penggunaan teknik pengolahan citra untuk membedakan varietas yang serupa secara fisik dengan lebih akurat, salah satunya berdasarkan warna pada citra. Warna memungkinkan pengenalan dan identifikasi objek dengan jelas, karena mata manusia dapat mengamati berbagai macam warna yang memungkinkan pengklasifikasian objek, seperti ketika membedakan jenis buah mangga. Pengklasifikasian mangga dapat dilakukan dengan mengambil sampel dan menentukannya berdasarkan kesamaan, termasuk dengan memperhatikan warna pada mangga sebagai salah satu faktor yang dijadikan pertimbangan.

Penelitian serupa mengenai klasifikasi jenis mangga berdasarkan bentuk dan tekstur daun menggunakan metode *Convolutional Neural Network* (CNN) yang diteliti oleh Hidayat dengan hasil pengujian ketiga jenis mangga selama 100 kali pengujian (*epoch*), memiliki durasi rata-rata adalah 2 detik dengan akurasi rata-rata

adalah 0,4916 dan terus meningkat pada tes ke-24 serta dengan ketepatan rata-rata 0,5333 dan terus meningkat pada tes ke-24 dengan tingkat akurasi sebesar 88,3% dengan *epoch* sebanyak 95 kali. Perbandingan dilakukan untuk menentukan metode mana yang memiliki performa yang lebih baik. Adapun saran dari penelitian ini bahwa metode akan lebih baik jika pengujian dilakukan lebih dari 50 kali [11].

Penelitian terdahulu mengenai klasifikasi berdasarkan citra menggunakan *Random Forest* untuk mengklasifikasikan motif songket Palembang berdasarkan SIFT dengan nilai rata-rata *overall accuracy* sebesar 92,98% per *class accuracy* 94,07%, *precision* 92,98%, dan *recall* 89,74%. Adapun masukkan yang digunakan pada penelitian ini adalah ekstraksi fitur citra menggunakan metode ekstraksi SIFT untuk mendapatkan informasi hasil fitur *keypoint*. *Keypoint* yang telah didapat bertujuan untuk mengatasi pengaruh *noise*. Pada penelitian ini objek citra tidak seluruhnya dapat diklasifikasi oleh metode *Random Forest* karena motif yang tersimpan pada dataset latih tidak memiliki kesamaan dengan motif yang ada pada data uji. Penelitian tersebut juga menyarankan untuk menggunakan jumlah citra yang lebih banyak agar memperoleh hasil akurasi yang lebih baik atau menggunakan metode klasifikasi lainnya [12].

Penelitian mengenai penentuan tingkat kemanisan buah mangga berdasarkan fitur warna pernah dilakukan dengan mengimplementasi *Support Vector Machine* (SVM) dan *K-Nearest Neighbour* (KNN) sebagai metoda pembandingan untuk klasifikasi. Dalam penelitian ini peneliti menyimpulkan bahwa tingkat akurasi atau kesesuaian sistem pengenalan tingkat kemanisan mangga berdasarkan citra warna buah untuk menggunakan metode *K-Nearest Neighbour* (KNN) memiliki persentase akurasi sebesar 83.3%, sistem mampu mengklasifikasikan citra buah mangga dengan benar berjumlah 20 buah dari 24 buah yang diuji dengan pengaturan parameter $k = 7$, sedangkan akurasi atau sistem pengenalan tingkat kemanisan mangga berdasarkan citra warna buah menggunakan metode *Support Vector Machine* (SVM) memiliki persentase sebesar 87.5%, sistem mampu mengklasifikasikan citra buah mangga dengan benar berjumlah 21 buah dari 24 buah yang diuji. Adapun keluaran dari penelitian yaitu buah mangga tersebut memiliki rasa manis, sedang, dan belum manis. Penelitian ini menyimpulkan bahwa metode *Support Vector Machine* (SVM) memiliki akurasi lebih unggul dalam mengklasifikasi berdasarkan fitur warna pada citra mangga. Namun terlihat nilai akurasi yang dihasilkan belum maksimal atau tidak menyentuh 90%-an keatas [13].

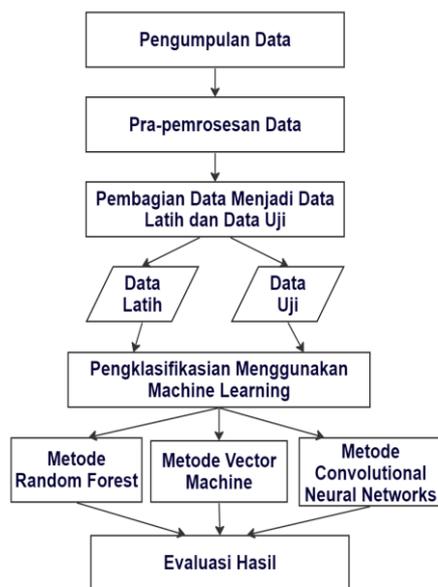
Dalam penelitian yang mengklasifikasikan jenis buah pisang menggunakan metode *Support Vector Machine* (SVM) berdasarkan citra warna menghasilkan akurasi sebesar 89,86%. Adapun masukan yang digunakan dari penelitian ini adalah ekstraksi fitur warna merubah

warna RGB ke BGR kemudian direduksi menggunakan PCA. Penelitian tersebut juga menyarankan untuk menggunakan metode dan model warna lain sebagai perbandingan hasil keakuratan sistem [14].

Berdasarkan hasil penelitian terdahulu yang dijelaskan sebelumnya mengenai klasifikasi berdasarkan citra gambar, urutan metode klasifikasi yang terbaik yaitu *Random Forest*, SVM dan CNN memiliki akurasi yang sangat baik dengan nilai akurasi yang tinggi. Oleh karena itu, dilakukan penelitian mengenai perbandingan metode *Random Forest*, CNN dan SVM untuk membandingkan hasil akurasi terbaik dalam klasifikasi jenis mangga. Penelitian ini melakukan pengolahan data berdasarkan citra mangga.

2. Metode Penelitian

Alur kegiatan penelitian yang dilakukan dapat dilihat pada Gambar 1. Penelitian ini dimulai dari pengumpulan data, pra-pemrosesan data, pembagian data menjadi data latih dan data uji, kemudian melakukan klasifikasi data menggunakan metode *Random Forest*, *Convolutional Neural Network*, dan *Support Vector Machine*, serta melakukan evaluasi hasil.



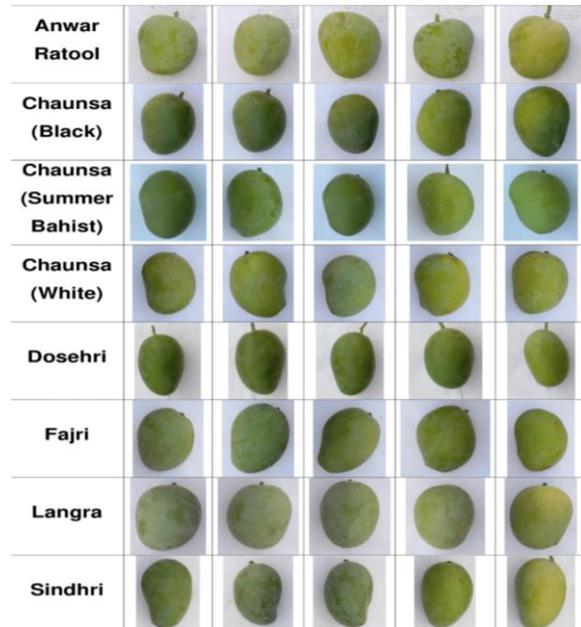
Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini berupa kumpulan gambar untuk beberapa jenis mangga dari Pakistan yang diperoleh dari website Kaggle [15]. Dataset ini diupload oleh Saurabh Shahane pada tahun 2021 dengan judul *Mango Varieties Classification and Grading*. Alasan dataset tersebut digunakan dalam penelitian ini adalah karena dataset memuat kumpulan gambar untuk setiap jenis mangga yang terdapat di Pakistan.

Dalam dataset ini terdapat delapan folder yang mewakili jenis mangga yang berbeda, yaitu mangga jenis Dosehri,

Chaunsa (*White*), Chaunsa (*Summer Bahist*), Chaunsa (*Black*), Sindhri, Langra, Fajri dan Anwar Ratool. Setiap folder terdapat 200 gambar mangga yang berbeda untuk jenis mangga tersebut. Berikut beberapa contoh gambar untuk setiap jenis buah mangga dari dataset yang digunakan yang dapat dilihat pada Gambar 2.



Gambar 2. Kumpulan Gambar Mangga

2.2. Pra-pemrosesan Data

Data yang didapat kemudian dilakukan pra-pemrosesan data agar data siap untuk diproses oleh algoritma dan hasil pemrosesan menjadi lebih efektif. Pra-pemrosesan data merupakan tahap awal yang penting dan bertujuan untuk memastikan bahwa data yang digunakan oleh algoritma klasifikasi memiliki konsistensi dan kualitas yang optimal. Tahapan pertama dalam pra-pemrosesan adalah penyeragaman ukuran citra, di mana setiap citra mangga diubah menjadi format 300 x 300 pixel. Hal ini dilakukan untuk memastikan bahwa seluruh citra memiliki dimensi yang seragam dan memfasilitasi proses pembelajaran algoritma secara efisien.

Tahap kedua dalam pra-pemrosesan adalah transformasi warna citra. Citra mangga yang mungkin memiliki variasi warna yang sangat beragam, diubah menjadi citra *grayscale* (skala abu-abu). Tujuan dari langkah ini adalah untuk menyederhanakan representasi warna, mengurangi kompleksitas data, dan memfokuskan perhatian algoritma pada informasi esensial yang diperlukan untuk klasifikasi mangga. Dengan demikian, proses ini meningkatkan efektivitas algoritma dengan memastikan bahwa citra yang digunakan memiliki karakteristik yang seragam dan relevan untuk proses klasifikasi lebih lanjut.

2.3. Pembagian Dataset

Data hasil pra-pemrosesan selanjutnya akan dibagi menjadi dua jenis, yaitu data latih (*training set*) dan data

uji (*validation set*). Pembagian dataset ini dilakukan untuk masing-masing jenis mangga untuk memastikan keberagaman dan representasi yang baik dengan skala pembagian dilakukan dengan proporsi 70:30.

Pembagian data latih dan data uji dengan proporsi 70:30 menjadi pilihan umum dalam mengevaluasi suatu metode *machine learning*. Dengan menggunakan 70% data untuk pelatihan memungkinkan model untuk mempelajari pola dari data yang diberikan. Sedangkan penggunaan data uji sebesar 30% memberikan sampel yang cukup besar untuk mengevaluasi kinerja model dengan baik dan memberikan perkiraan yang akurat tentang perilakunya pada data yang belum pernah dipelajari sebelumnya. Penggunaan dataset uji yang cukup besar memungkinkan validasi model yang lebih kuat, memastikan bahwa model tidak hanya belajar pola yang spesifik pada data latih tetapi juga mampu menggeneralisasi dengan baik ke data baru. Penggunaan 30% data untuk pengujian juga memastikan sampel yang cukup besar untuk menghasilkan estimasi kinerja model yang andal, mengurangi variabilitas dalam estimasi dan memberikan hasil yang lebih stabil. Dengan demikian, didapatkan 70% data latih dan 30% data uji untuk mencapai keseimbangan yang baik antara kinerja model dan efisiensi waktu dalam proses pembelajaran atau pelatihan.

2.4. Metode Machine Learning

Penelitian ini menggunakan metode *machine learning* yang diujikan dalam klasifikasi citra mangga, antara lain *Random Forest*, *Convolutional Neural Networks* (CNN), dan *Support Vector Machines* (SVM). Tujuan membandingkan kinerja masing-masing metode dalam mengklasifikasikan citra mangga adalah untuk memberikan wawasan yang lebih mendalam tentang penerapan teknik *machine learning* dalam domain klasifikasi citra. Berikut penjelasan mengenai metode *machine learning* yang digunakan dalam penelitian ini.

2.4.1. Random Forest

Random Forest merupakan metode *ensemble* pohon keputusan yang kuat yang dapat digunakan untuk berbagai tugas klasifikasi pola dengan menggunakan sekumpulan pohon klasifikasi sebagai pembelajaran dasar. Metode ini bergantung dengan *bagging* (*bootstrap aggregating*) dan pengambilan sampel subruang secara acak untuk membangun sebuah komite. Oleh karena itu label kelas akhir dari setiap contoh data ditentukan melalui suara mayoritas. Misalkan $\{X, T\}$ menyatakan sekumpulan data pelatihan di mana:

$$X = x_0, x_1, \dots, x_{n-1} \text{ dan } T = t_0, t_1, \dots, t_{n-1}.$$

Misalkan $h(x)$ mempresentasikan sebuah pohon klasifikasi. Untuk setiap pohon individu $h(x)$, model memilih sampel acak dengan mengganti data pelatihan yang telah dikumpulkan dan menggunakan data sampel tersebut untuk melatih $h(x)$. Prosedur ini bertujuan untuk mencapai kinerja model yang lebih baik karena

prosedur ini memiliki kemampuan untuk mengurangi varians model tanpa meningkatkan bias model. Mekanisme *feature bagging* juga digunakan oleh *Random Forest* selain *sample bagging*. Artinya, subset fitur digunakan untuk melatih $h(x)$. Ini adalah proses untuk mengurangi korelasi dari pembelajaran secara keseluruhan di seluruh komite. Biasanya untuk kasus klasifikasi pola, jumlah fitur yang dipilih oleh individu $h(x)$ adalah \sqrt{D} di mana D adalah jumlah total fitur yang tersedia [16]. Penggunaan metode *random forest* ini cocok untuk analisis ukuran dataset berukuran besar [17].

Algoritma *Random Forest* yang digunakan menggunakan pustaka *scikit-learn* di Python. Pertama, objek *Random Forest Classifier* dibuat dengan parameter *default* menggunakan sintaks "modelRF = RandomForestClassifier()". Selanjutnya, model tersebut dilatih dengan data latih (x_{Latih} dan y_{Latih}) menggunakan perintah "modelRF.fit(x_{Latih} , y_{Latih})". Setelah proses pelatihan selesai, model *Random Forest* melakukan prediksi pada data uji (x_{Uji}) menggunakan perintah "yPred = modelRF.predict(x_{Uji})". Akurasi model diukur dengan membandingkan prediksi yang dihasilkan (y_{Pred}) dengan label sebenarnya pada data uji (y_{Uji}) menggunakan metrik akurasi melalui perintah "hasilAkurasi = accuracy_score(y_{Uji} , y_{Pred})". Dengan *Random Forest*, sejumlah pohon keputusan dibangun untuk meningkatkan akurasi dan ketepatan prediksi, menjadikannya algoritma yang kuat untuk tugas klasifikasi pada dataset.

Berikut adalah algoritma *Random Forest* yang digunakan:

Algoritma Random Forest

```
from sklearn.ensemble import
RandomForestClassifier
from sklearn.metrics import accuracy_score

modelRF = RandomForestClassifier() //parameter
default

# Latih model
modelRF.fit(xLatih, yLatih)

# Lakukan prediksi pada data uji
yPred = modelRF.predict(xUji)

hasilAkurasi = accuracy_score(yUji, yPred)
```

2.4.2. Convolutional Neural Network

CNN merupakan salah satu algoritma *deep learning* yang memiliki representasi fitur yang efektif dalam pengenalan citra, deteksi objek dalam skala yang berbeda, dan segmentasi semantik [11]. CNN dirancang untuk mengolah data dua dimensi, seperti citra, dan dapat digunakan untuk tugas klasifikasi [11]. CNN secara otomatis dapat mengekstraksi informasi dari setiap citra sebagai fungsi sebagai metode *supervised learning*. Hal ini terimplementasi melalui algoritma *deep learning*, yang secara spesifik menggunakan *Multilayer Perceptron* (MPL) dan memproses data dalam format dua dimensi.

Keunggulan dari metode CNN ini terlihat ketika menangani citra dengan banyak kelas. Semakin banyak kelas yang ditangani, metode ini akan meningkatkan ketajaman analisis terhadap data dan memberikan model kemampuan untuk memahami data yang lebih kompleks dalam citra [18]. Hal ini membuat metode CNN menjadi salah satu metode yang dipilih untuk melakukan tugas yang melibatkan pengenalan objek.

Berikut adalah algoritma CNN yang digunakan:

```
Algoritma Convolutional Neural Network
from keras.models import Sequential
from keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout
from keras.utils import to_categorical
from sklearn.metrics import accuracy_score

#Membangun model CNN
model = Sequential()
model.add(Conv2D())
model.add(MaxPooling2D())
model.add(Flatten())
model.add(Dense())
model.add(Dropout())
//parameter default

#Compile model
model.compile() //parameter default

# Latih model
model.fit(xLatih, yLatih, epochs=100,
validation_data=(xUji, yUji))

# Evaluasi model
accuracy = model.evaluate(xUji, yUji)

#Lakukan prediksi pada data uji
yPred = model.predict(xUji)

# Mengambil indeks kelas dengan probabilitas tertinggi sebagai prediksi
yPredClasses = np.argmax(y_pred, axis=1)

# Mengambil indeks kelas dengan probabilitas tertinggi dari one-hot encoding y_test
yTrue = np.argmax(y_test, axis=1)

#Laporan klasifikasi
class_names = [int_to_label[i] for i in range(len(unique_labels))]

hasilAkurasi = classification_report(yTrue, yPredClasses, target_names = class_names )
print(hasilAkurasi)
```

Algoritma CNN dalam implementasi ini menggunakan pustaka Keras di Python. Model CNN dibangun secara berurutan dengan lapisan-lapisan seperti *Convolutional*, *MaxPooling*, *Flatten*, *Dense*, dan *Dropout* menggunakan parameter *default*. Setelah membangun model, dilakukan kompilasi dengan perintah "model.compile()" menggunakan parameter *default*. Model kemudian dilatih dengan data latih (xLatih dan yLatih) selama 100 *epochs* dengan validasi menggunakan data uji (xUji dan yUji). Evaluasi model dilakukan untuk mengukur akurasi, dan prediksi pada data uji dilakukan dengan perintah "yPred = model.predict(xUji)". Hasil prediksi diambil dengan mengambil indeks kelas dengan probabilitas tertinggi, dan laporan klasifikasi dihasilkan menggunakan perintah "classification_report". Ini memberikan

ringkasan yang komprehensif tentang performa model, termasuk akurasi dan metrik-metrik lainnya.

2.4.3. Support Vector Machine

SVM merupakan sebuah algoritma yang menggunakan pemetaan *non-linier* untuk mengubah data latih ke dalam dimensi yang lebih tinggi. Dalam penggunaan teknik SVM, metode ini mencoba menemukan fungsi pemisah optimal untuk memisahkan dua kelas yang berbeda. Pendekatan ini bertujuan untuk menemukan *hyperplane* terbaik yang memisahkan dua jenis objek, di mana *hyperplane* tersebut berada di tengah antara dua set objek dari dua kelas yang berbeda.

Hyperplane merupakan garis pembatas yang digunakan untuk mengklasifikasikan titik data, dan titik data yang berada di sisi berlawanan dari *hyperplane* biasanya diklasifikasikan ke dalam kelas yang berbeda. Dimensi dari *hyperplane* tergantung pada jumlah fitur yang ada dalam data [19]. *Hyperplane* yang efektif adalah *hyperplane* yang berada pada posisi pusat di antara dua kelompok objek tersebut [20].

Berikut algoritma SVM yang digunakan:

```
Algoritma Support Vector Machine
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score

modelSVM = SVC() //parameter default

# Latih model
modelSVM.fit(xLatih, yLatih)

# Lakukan prediksi pada data uji
yPred = modelSVM.predict(xUji)

hasilAkurasi = accuracy_score(yUji, yPred)
```

Algoritma SVM yang digunakan menggunakan pustaka scikit-learn di Python. Model SVM dibuat dengan parameter *default* menggunakan sintaks "modelSVM = SVC()". Setelah itu, model dilatih dengan data latih (xLatih dan yLatih) menggunakan perintah "modelSVM.fit(xLatih, yLatih)", dan prediksi dilakukan pada data uji (xUji) dengan perintah "yPred = modelSVM.predict(xUji)". Akurasi model dievaluasi dengan membandingkan hasil prediksi (yPred) dengan label sebenarnya pada data uji (yUji) menggunakan metrik akurasi melalui perintah "hasilAkurasi = accuracy_score(yUji, yPred)". Ini memberikan gambaran singkat tentang performa model SVM dalam tugas klasifikasi pada dataset yang diberikan.

2.4.4. Metrik Perbandingan

Untuk menilai kinerja model klasifikasi mangga yang diusulkan maka dilakukan perbandingan akurasi dari metode-metode yang digunakan. Selain menampilkan hasil akurasi dari setiap metode, ditampilkan juga informasi mengenai *precision*, *recall*, dan *f1-score*.

Akurasi adalah rasio jumlah prediksi yang benar terhadap jumlah semua prediksi oleh pengklasifikasi [21]. Akurasi digunakan untuk mengukur seberapa tepat

model dapat mengklasifikasikan data dengan menghitung rasio jumlah prediksi yang benar (positif dan negatif) terhadap total data. Nilai akurasi yang tinggi tidak selalu menunjukkan performa model yang baik dalam kondisi ketidakseimbangan kelas.

Rumus untuk menghitung akurasi terlihat pada persamaan (1).

$$\text{Akurasi} = \frac{\sum_{k=1}^K (TP_k + TN_k)}{\sum_{k=1}^K (TP_k + TN_k + FP_k + FN_k)} \quad (1)$$

Dalam rumus menghitung akurasi yang dapat dilihat pada persamaan (1), jumlah total kelas disimbolkan sebagai K , jumlah *True Positives* untuk kelas k sebagai TP_k , jumlah *True Negatives* untuk kelas k sebagai TN_k , jumlah *False Positives* untuk kelas k sebagai FP_k , dan jumlah *False Negatives* untuk kelas k sebagai FN_k .

Precision adalah hasil perbandingan dari jumlah data bernilai positif dengan hasil jumlah data benar bernilai positif dan data salah bernilai positif [22]. *Precision* mengukur seberapa akurat prediksi positif model. Ini adalah rasio jumlah positif sebenarnya terhadap jumlah total prediksi positif dan berfokus pada keakuratan prediksi positif. Nilai *precision* yang tinggi mengindikasikan model membuat prediksi positif yang kemungkinan prediksi tersebut benar.

Rumus untuk menghitung *precision* dapat dilihat pada persamaan (2).

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (2)$$

Recall atau *True Positive Rate* (TPR) merupakan perbandingan antara jumlah *True Positives* dan total sampel yang relevan [21]. Ini mencerminkan proporsi contoh yang sebenarnya positif dan diidentifikasi sebagai positif. *Recall* digunakan untuk mengevaluasi kinerja model dalam mengidentifikasi hasil dari perbandingan nilai data benar bernilai positif dengan hasil jumlah data benar yang bernilai positif dan data salah bernilai negatif [22].

Rumus untuk menghitung *recall* dapat dilihat pada persamaan (3).

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (3)$$

F1-Score adalah nilai rata-rata harmonik dan merupakan parameter evaluasi kinerja *retrieval* yang menyatukan *precision* dan *recall*. Nilainya dihitung melalui perkalian *precision* dan *recall*, kemudian hasilnya dibagi dengan penjumlahan *precision* dan *recall*, lalu dikalikan dengan dua [21][22]. Rumus untuk menghitung *F1-Score* dapat dilihat pada persamaan (4).

$$\text{F1 - Score} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (4)$$

Pada persamaan (2), (3), dan (4), *True positive* (TP) adalah keputusan menempatkan dua data yang identik ke kelompok yang sama, *True Negative* (TN) adalah

keputusan menempatkan dua data yang tidak identik ke kelompok berbeda. Ada dua jenis kesalahan yang dapat terjadi pada proses pengklasteran yaitu *False positive* (FP) dan *False negative* (FN). FP adalah keputusan menempatkan dua data yang tidak identik ke kelompok yang sama. FN adalah keputusan menempatkan dua data yang identik ke kelompok yang berbeda [23].

2.5. Evaluasi Hasil

Selanjutnya dilakukan evaluasi untuk mengetahui metode mana yang menghasilkan pengklasifikasian terbaik dalam melakukan pengklasifikasian jenis mangga berdasarkan citra mangga. Evaluasi dilakukan dengan membandingkan nilai akurasi dari setiap metode, serta ditampilkan nilai *precision*, *recall*, serta *f1-score* untuk pengujian setiap jenis mangga.

3. Hasil dan Pembahasan

Pengujian yang dilakukan dalam penelitian ini menggunakan citra mangga dengan total 1.600 gambar, yang terdiri dari 8 kelas, dimana setiap kelas mewakili jenis mangga. Pada penelitian ini menggunakan metode *Random Forest*, *CNN*, dan *SVM* dimana sebelumnya data yang digunakan sudah dilakukan tahap pra-pemrosesan agar dapat diterapkan pada algoritma. Data yang sudah melalui tahap pra-pemrosesan dibagi menjadi data latih dan data uji dengan rasio perbandingan 70:30. Kemudian, dilakukan pelatihan pada data latih dan pengujian pada data uji dengan menggunakan metode *Random Forest*, *SVM*, dan *CNN* untuk dilakukan perbandingan.

Tabel 1. Tabel Hasil Random Forest

Jenis Mangga	Precision	Recall	F1-Score
anwar ratool	0.98	1	0.99
chaunsa - black	0.96	0.93	0.95
chaunsa - summer bahist	0.97	0.99	0.98
chaunsa - white	0.95	0.95	0.95
dosehri	0.97	1	0.98
fajri	0.97	0.89	0.93
langra	0.95	0.98	0.96
sindhri	0.97	0.98	0.98

Proses klasifikasi menggunakan metode *Random Forest* menghasilkan nilai *precision*, *recall* dan *f1-score* pada tiap jenis mangga yang dapat dilihat pada Tabel 1.

Tabel 2. Tabel Hasil SVM

Jenis Mangga	Precision	Recall	F1-Score
anwar ratool	1	0.97	0.98
chaunsa - black	0.95	0.95	0.95
chaunsa - summer bahist	0.94	0.96	0.95
chaunsa - white	0.95	0.97	0.96
dosehri	0.95	0.95	0.95
fajri	0.95	0.89	0.92
langra	0.95	1	0.97
sindhri	0.9	0.92	0.91

Proses klasifikasi menggunakan metode *SVM* menghasilkan nilai *precision*, *recall* dan *f1-score* pada tiap jenis mangga yang dapat dilihat pada Tabel 2.

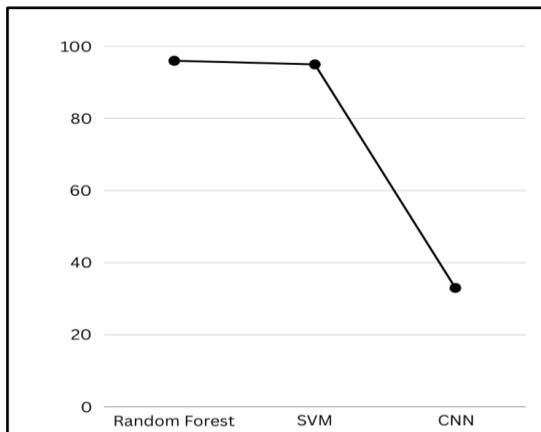
Proses klasifikasi menggunakan metode CNN menghasilkan nilai *precision*, *recall* dan *f1-score* pada tiap jenis mangga yang dapat dilihat pada Tabel 3.

Tabel 3. Tabel Hasil CNN

Jenis Mangga	Precision	Recall	F1-Score
anwar ratool	0.72	0.36	0.48
chaunsa - black	0.27	0.48	0.35
chaunsa – summer bahist	0.76	0.23	0.36
chaunsa - white	0.29	0.18	0.22
dosehri	0.5	0.07	0.12
fajri	0.25	0.62	0.36
langra	0.32	0.43	0.37
sindhri	0.31	0.28	0.29

Dari tabel-tabel yang berisi nilai hasil dari pemrosesan setiap metode, dapat dilihat bahwa nilai rata-rata dari *precision*, *recall*, serta *f1-score* dari setiap jenis mangga dari ketiga metode yang digunakan pada Tabel 4.

Dari Tabel 4 dapat dilihat bahwa klasifikasi menggunakan metode *Random Forest* menghasilkan nilai terbaik dengan *accuracy* 96%, rata-rata *precision* 96.5%, rata-rata *recall* 96.5%, dan rata-rata *f1-score* 96.5%. Klasifikasi dengan CNN menghasilkan nilai terendah dengan *accuracy* 33%, rata-rata *precision* 42.7%, rata-rata *recall* 33.1%, dan rata-rata *f1-score* 31.8%. Klasifikasi dengan SVM menghasilkan nilai terbaik kedua dengan *accuracy* 95%, rata-rata *precision* 94.8%, rata-rata *recall* 95.1%, dan rata-rata *f1-score* 94.8%.



Gambar 3. Grafik Titik Hasil Akurasi

Tabel 4. Tabel Hasil Akurasi

Metode	Akurasi	Precision	Recall	F1-Score
Random Forest	0.96	0.965	0.965	0.965
SVM	0.95	0.94875	0.95125	0.94875
CNN	0.33	0.4275	0.33125	0.31875

Pada Gambar 3 dapat dilihat gambaran grafik dari hasil yang didapat bahwa metode *Random Forest* menghasilkan nilai akurasi terbaik diantara ketiga metode yang digunakan. Hal ini disebabkan karena *Random Forest* cocok untuk analisis data dengan ukuran yang besar serta kualitas pemrosesan data yang digunakan sangat baik untuk metode *Random Forest*.

Dari hasil pengujian yang dapat dilihat pada Tabel 4, rata-rata nilai akurasi yang dihasilkan metode CNN paling rendah diantara ketiga metode yang digunakan. Berdasarkan hasil pengamatan, penyebab kecilnya nilai akurasi yang dihasilkan kemungkinan terletak pada data yang digunakan untuk klasifikasi tidak dinormalisasi sehingga perlu dilakukan normalisasi lebih dulu. Oleh karena itu, sebagai pembuktian hasil analisis mengenai rendahnya nilai akurasi CNN, maka dilakukan perbaikan pada metode CNN.

Dataset dilakukan perbaikan dengan menambahkan proses normalisasi sebelum dilakukan pengklasifikasian menggunakan metode CNN. Hasil pengujian dengan menggunakan data yang telah dinormalisasi untuk metode CNN pada tiap jenis mangga bisa dilihat pada Tabel 5.

Tabel 6 merupakan hasil nilai pengujian dari metode CNN yang dataset telah dilakukan normalisasi. Nilai akurasi yang dihasilkan dari metode CNN yang datanya telah dinormalisasi yaitu sebesar 83% dengan nilai rata-rata *precision* sebesar 84,1%, rata-rata *recall* sebesar 83,1%, dan rata-rata *f1-score* sebesar 83,2%. Hasil ini lebih besar dari hasil pengujian data sebelum di normalisasi yang bisa dilihat pada Tabel 4.

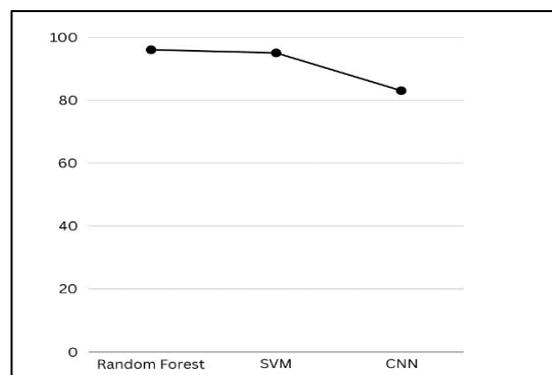
Tabel 5. Tabel Hasil CNN Dengan Normalisasi

Jenis Mangga	Precision	Recall	F1-Score
anwar ratool	0.98	0.83	0.90
chaunsa - black	0.79	0.95	0.86
chaunsa – summer bahist	0.71	0.86	0.78
chaunsa - white	0.77	0.74	0.75
dosehri	0.91	0.93	0.92
fajri	0.78	0.83	0.81
langra	0.98	0.87	0.92
sindhri	0.81	0.64	0.72

Tabel 6. Tabel Hasil Akurasi Perbaikan CNN

Metode	Akurasi	Precision	Recall	F1-Score
Random Forest	0.96	0.965	0.965	0.965
SVM	0.95	0.94875	0.95125	0.94875
CNN*	0.83	0.84125	0.83125	0.81875

*) CNN setelah menggunakan normalisasi



Gambar 4. Grafik Titik Hasil Akurasi Perbaikan CNN

Pengujian pada perbaikan normalisasi menunjukkan penyebab metode CNN menghasilkan nilai akurasi yang

rendah dikarenakan tidak dinormalisasi dengan baik. Adapun normalisasi yang dilakukan adalah mengubah rentang data dalam skala 0-1 agar data dapat diterapkan lebih baik oleh metode. Akan tetapi, dibandingkan dengan metode *machine learning* lain yang diujikan, ternyata nilai akurasi dari metode CNN rendah. Hal ini mungkin terjadi dikarenakan pengolahan fitur citra gambar yang digunakan tidak cukup baik dalam mewakili informasi yang penting dalam data dimana dalam penelitian ini menggunakan skala abu-abu dari citra.

Berdasarkan hasil penelitian tersebut, metode *Random Forest* mampu mengklasifikasikan jenis mangga dengan lebih baik daripada metode lain yang diuji. Selain itu, hasil penelitian ini juga menunjukkan bahwa pendekatan *machine learning* lebih efektif dibandingkan dengan pendekatan *deep learning* untuk kasus yang diuji dalam penelitian ini. Hal ini terlihat pada metode CNN yang memberikan hasil akurasi yang rendah dibandingkan dengan metode *Random Forest* meskipun sudah melakukan normalisasi data untuk metode CNN.

4. Kesimpulan

Penggunaan metode *Random Forest* sangat baik digunakan dalam pengklasifikasian jenis mangga berdasarkan citra dengan nilai akurasi sebesar 96% dengan nilai *precision*, *recall* dan *f1-score* dari algoritma *Random Forest* masing-masing sebesar 96.5%. Hasil dari penelitian ini menunjukkan bahwa metode *Random Forest* mampu mengklasifikasikan jenis mangga lebih baik jika dibandingkan dengan metode lain yang diujikan. Hasil penelitian ini menunjukkan bahwa metode *machine learning* lebih baik dibandingkan metode *deep learning* untuk kasus yang dilakukan pada penelitian ini. Penelitian lebih lanjut untuk klasifikasi jenis mangga berdasarkan citra perlu dilakukan dengan menambahkan fitur-fitur yang lebih signifikan untuk mengklasifikasikan jenis mangga tidak hanya berdasarkan skala abu-abu citra gambar.

Daftar Rujukan

[1] M. W. Khan, Z. Hussain, and I. Ghafoor, "Mango Fruit Orchards Infestation: A Case Study On Mango Fruits In Pakistan," 2023. [Online]. Available: <http://xisdxjxsu.asia>

[2] V. Farina, C. Gentile, G. Sortino, G. Gianguzzi, E. Palazzolo, and A. Mazzaglia, "Tree-Ripe Mango Fruit: Physicochemical Characterization, Antioxidant Properties And Sensory Profile Of Six Mediterranean Grown Cultivars," *Agronomy*, vol. 10, no. 6, 2020, doi: 10.3390/agronomy10060884.

[3] E. M. Yahia, J. de J. Ornelas-Paz, J. K. Brecht, P. García-Solís, and M. E. Maldonado Celis, "The Contribution of Mango Fruit (*Mangifera Indica* L.) To human nutrition and health," *Arabian Journal of Chemistry*, vol. 16, no. 7. 2023. doi: 10.1016/j.arabjc.2023.104860.

[4] H. M. Rizwan Iqbal and A. Hakim, "Classification and Grading of Harvested Mangoes Using Convolutional Neural Network," *Int. J. Fruit Sci.*, vol. 22, no. 1, 2022, doi: 10.1080/15538362.2021.2023069.

[5] H. Tang, Z. Rasool, A. I. Khan, A. A. Khan, M. A. Khan, and G. A. Azaz, "The Impact of Private Standards on Corporate Social Responsibility Compliance and Rural Workers'

Motivation in Developing Countries: A Study of Mango Farms in Pakistan," *Int. J. Food Sci.*, vol. 2021, 2021, doi: 10.1155/2021/9985784.

[6] I. A. Sabir, S. Ahmad, M. Nafees, A. S. Khan, Maryam, and I. Ahmad, "Trade Potential Evaluation of Indigenous and Exotic Mango Genotypes Through Physico-Chemical And Sensory Attributes," *Sarhad J. Agric.*, vol. 34, no. 3, 2018, doi: 10.17582/journal.sja/2018/34.3.680.689.

[7] S. Zahoor, F. Anwar, R. Qadir, W. Soufan, and M. Sakran, "Physicochemical Attributes and Antioxidant Potential of Kernel Oils from Selected Mango Varieties," *ACS Omega*, vol. 8, no. 25, 2023, doi: 10.1021/acsomega.3c01155.

[8] J. Jamaludin, C. Rozikin, and A. S. Y. Irawan, "Klasifikasi Jenis Buah Mangga dengan Metode Backpropagation," *Techné J. Ilm. Elektrotek.*, vol. 20, no. 1, 2021, doi: 10.31358/techne.v20i1.231.

[9] A. Raghavendra, D. S. Guru, M. K. Rao, and R. Sumithra, "Hierarchical Approach for Ripeness Grading of Mangoes," *Artif. Intell. Agric.*, vol. 4, 2020, doi: 10.1016/j.aiaa.2020.10.003.

[10] D. Worasawate, P. Sakunasinha, and S. Chiangga, "Automatic Classification of the Ripeness Stage of Mango Fruit Using a Machine Learning Approach," *AgriEngineering*, vol. 4, no. 1, pp. 32–47, Mar. 2022, doi: 10.3390/agriengineering4010003.

[11] D. Hidayat, "Klasifikasi Jenis Mangga Berdasarkan Bentuk dan Tekstur Daun Menggunakan Metode Convolutional Neural Network (CNN)," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, 2022, doi: 10.31539/intecom.v5i1.3401.

[12] S. Devella, Y. Yohannes, and F. N. Rahmawati, "Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 2, pp. 310–320, Aug. 2020, doi: 10.35957/jatisi.v7i2.289.

[13] M. Ichwan, I. A. Dewi, and Z. M. S., "Klasifikasi Support Vector Machine (SVM) Untuk Menentukan TingkatKemanisan Mangga Berdasarkan Fitur Warna," *MIND J.*, vol. 3, no. 2, 2019, doi: 10.26760/mindjournal.v3i2.16-23.

[14] Y. Amrozi, D. Yuliaty, A. Susilo, N. Novianto, and R. Ramadhan, "Klasifikasi Jenis Buah Pisang Berdasarkan Citra Warna dengan Metode SVM," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 3, pp. 394–399, Dec. 2022, doi: 10.32736/sisfokom.v11i3.1502.

[15] Shahane, Saurabh, "Mango Varieties Classification and Grading," <https://www.kaggle.com>. Accessed: Sep. 14, 2023. [Online]. Available: <https://www.kaggle.com/datasets/saurabhshahane/mango-varieties-classification>

[16] H. Nhat-Duc and T. Van-Duc, "Comparison of histogram-based gradient boosting classification machine, random Forest, and deep convolutional neural network for pavement raveling severity classification," *Autom. Constr.*, vol. 148, p. 104767, Apr. 2023, doi: 10.1016/J.AUTCON.2023.104767.

[17] F. Hamami and A. Dahlan, "Klasifikasi Cuaca Provinsi DKI Jakarta Menggunakan Algoritma Random Forest Dengan Teknik Oversampling," 2022.

[18] K. P. Siwilopo and H. Marcos, "Membandingkan Klasifikasi Pada Buah Jeruk Menggunakan Metode Convolutional Neural Network Dan K-Nearest Neighbor," *KOMPUTA J. Ilm. Komput. dan Inform.*, vol. 12, no. 1, 2023.

[19] J. Kusuma, Rubianto, R. Rosnelly, Hartono, and B. H. Hayadi, "Klasifikasi Penyakit Daun Pada Tanaman Jagung Menggunakan Algoritma Support Vector Machine, K-Nearest Neighbors dan Multilayer Perceptron," *J. Appl. Comput. Sci. Technol.*, vol. 4, no. 1, 2023, doi: 10.52158/jacost.v4i1.484.

[20] V. I. Yani, A. Aradea, and H. Mubarak, "Optimasi Prakiraan Cuaca Menggunakan Metode Ensemble pada Naïve Bayes dan C4.5," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 3, Dec. 2022, doi: 10.28932/jutisi.v8i3.5455.

[21] K. B. Dasari and N. Devarakonda, "Detection of different DDoS attacks using machine learning classification Algorithms," *Ing. des Syst. d'Information*, vol. 26, no. 5, 2021, doi: 10.18280/isi.260505.

[22] N. Arifin, U. Enri, and N. Sulistiyowati, "Penerapan Algoritma

- Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk [23] R. Siringoringo, "Peningkatan Performa Cluster Fuzzy C-Means Pada Pengklasteran Sentimen Menggunakan Particle Text Classification," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 6, no. 2, 2021, doi: 10.30998/string.v6i2.10133. Swarm Optimization," vol. 6, no. 4, pp. 349–354, 2018, doi: 10.25126/jtiik.2018561090.