Terakreditasi SINTA Peringkat 2

Surat Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti No. 10/E/KPT/2019

masa berlaku mulai Vol. 1 No. 1 tahun 2017 s.d Vol. 5 No. 3 tahun 2021

Terbit online pada laman web jurnal: **http://jurnal.iaii.or.id**

|  |  |  |
| --- | --- | --- |
| logok.png | JURNAL RESTI  (Rekayasa Sistem dan Teknologi Informasi) | |
| Vol. 5 No. x (2021) x - x | ISSN Media Elektronik: 2580-0760 |

Penanganan Masalah Overfitting dan Underfitting pada Klasifikasi Jenis Bunga Menggunakan CNN

Yulisha Dian Cahyaningrum1, Eka Yunia Primastuti2 , Yufis Azhar3

1,2,3 Teknik Informatika, Universitas Muhammadyah Malang

[1yulishacahya@gmail.com](mailto:1yulishadian@gmail.comorg) , [2ekayunia45@gmail.com](mailto:2ekayunia45@gmail.com), [2yufisazhar@gmail.com](mailto:2yufisazhar@gmail.com)

# *Abstract*

*Classification of an object or image on another object is an activity that has very significant and challenging problems in the field of image selection. There are many types of flowers with various shapes and colors, so image processing technology is needed that can recognize the characteristics of flowers by classifying the flowers in terms of shape and color. Many methods have been used by previous researchers such as DNN, SVM, KNN, etc. However, in this research, image processing activities still pay less attention to problems that occur such as overfitting and underfitting. Therefore, in this paper we propose the use of the convolutional neural network (CNN) method and carry out data imbalance activities in the preprocessing section for classification activities. The data is divided into three parts, namely, validation, train, and test. The data will be added and entered into a predetermined model and level of accuracy in the program to classify an image. The classification got a precision result of 50% daisy, 52% dandelion, 88% rose, 1.00% sunflower, 1.00% tulip and got a recall value of 88% daisy, 92% dandelion, 78% rose, 13% sunflower, 98% tulip with a value of 62% . And got a total accuracy of 88%.*

# *Keywords: classification, interest, CNN*

# Abstrak

Pengklasifikasian suatu objek bunga maupun citra pada objek lainnya merupakan kegiatan yang memiliki permasalahan yang sangat sigifikan dan menantang dalam bidang pemilihan citra. Banyaknya jenis bunga dengan bentuk dan warna yang bermacam-macam, sehingga diperlukannya teknologi pengolahan citra yang dapat mengenali karakteristik bunga dengan cara pengklasifikasian pada bunga tersebut baik dari segi bentuk maupun warna. Banyak metode yang telah digunakan oleh peneliti-peneliti sebelumnya seperti metode DNN, SVM, KNN, dll. akan tetapi, dalam penelitian tersebut kegiatan pengolahan citra masih kurang memperhatikan dalam mengatasi permasalahan yang terjadi seperti overfitting dan underfitting. Oleh karena itu, dalam makalah ini kami mengusulkan menggunakan metode convolutional neural network (CNN) dan melakukan kegiatan imbalance data dalam bagian preprocessing untuk kegiatan pengklasifikasian. Data terbagi menjadi tiga bagian yaitu, validasi, train, dan tes. Data tersebut akan di augmentasi dan dimasukkan kedalam model yang sudah ditentukan dan dievaluasi tingkat akurasi dalam ketepatan program untuk mengklasifikasikan suatu citra. Klasifikasi mendapat hasil precission 50% daisy, 52% dandelion, 88% rose, 1.00% sunflowers, 1.00% tulip dan mendapat nilai recall 88% daisy, 92% dandelion, 78% rose, 13% sunflowers, 98% tulip dengan nilai 62%. Dan mendapat akurasi keseluruhan 88%.

Kata kunci: klasifikasi, bunga, CNN

1. **Pendahuluan**

Tumbuhan merupakan salah satu makhluk hidup yang terdapat di dalam alam semesta [1]. Tumbuhan sendiri memiliki banyak jenis yang tak terbatas dengan klasifikasi yang berbeda-beda. Selain itu, tumbuhan memiliki organ seperti makhluk hidup lainnya, organ pada tumbuhan dibagi menjadi 2 yaitu, organ bagian pokok tumbuhan dan organ sekunder [2]. Organ bagian pokok tumbuhan merupakan bagian yang selalu dimiliki oleh setiap tumbuhan yang terdiri dari akar, batang, dan daun. Sedangkan bagian organ sekunder merupakan organ yang terbentuk dari modifikasi bagian pokok atau kombinasi bagian-bagian pokok, antara lain seperti kuncup (*gemma*) yaitu modifikasi dari batang dan daun, bunga (*flos*) modifikasi dari batang dan daun, maupun duri (*spina*) modifikasi dari dahan dan daun, serta organ-organ kombinasi lainnya.

Bunga merupakan salah satu organ sekunder dari tumbuhan dan merupakan suatu kombinasi dari organ pokok tumbuhan yang berkemungkinan besar memiliki jenis yang selalu dapat diperbarui melakukan persilangan atau kegiatan pengkombinasian yang berbeda dan hasil dari kombinasi tersebut dapat memberikan jenis bunga yang baru. Dengan banyaknya jenis bunga dengan bentuk dan warna yang bermacam-macam sehingga diperlukannya teknologi pengolahan citra yang dapat mengenali karakteristik bunga dengan cara pengklasifikasian pada bunga tersebut baik dari segi bentuk maupun warna [3].

Pengklasifikasian suatu objek maupun citra merupakan kegiatan yang cukup mudah dilakukan oleh manusia, saat ini pengklasifikasian juga sangat digunakan terutama dengan adanya pertumbuhan data dalam berbagai bidang antara lain kesehatan, e-commerce, game, dan bidang-bidang lainnya [4]. Karena ketersediaan data yang besar, kegiatan pengklasifikasian diberikan pengembangan dalam bidang teknologi agar permasalahan dapat diselesaikan dengan cara yang cepat dan akurat. Akan tetapi, pengklasifikasian bagi komputer maupun mesin merupakan tugas yang sangat kompleks. Sehingga, klasifikasi objek menjadi masalah utama pada bidang *Computer Vision* yang sejak dahulu dicari pemecahannya [5].

Pemecahan pada kegiatan pengklasifikasian citra pada era saat ini, sudah berkembang semakin cepat dan menjadi tren di kalangan pengembang teknologi. Pengklasifikasian citra sendiri memiliki komponen dan prosedur yang kompleks dan bermacam-macam seperti metode *machine learning* maupun *deep learning*. Metode *machine larning* yang saat ini sering digunakan adalah *support vector machine (SVM), neural network*, maupun metode *k-nearest network.* Untuk metode *deep learning* sediri yang sangat dominan untuk digunakan karena memiliki model, dan struktur yang kompleks yaitu metode *convolutional neural network* (CNN) [6]. CNN juga memiliki lebih banyak keunggulan, di satu sisi memiliki lebih banyak lapisan tersembunyi dan disisi lain memiliki kemampuan pembelajaran fitur dan ekspresi fitur yang lebih kuat. Sehingga, kegiatan pengklasifikasian citra dengan menggunakan metode CNN memiliki banyak varian, tentu dengan kelebihan dan kekurangan masing-masing.

Penelitian mengenai klasifikasi citra pada jenis bunga menggunakan metode machine learning maupun deep learning sudah banyak dilakukan sebelumnya. [7] Pada penelitian yang berjudul *“Research on Flower Image Classification Algorithm Based on Convolutional Neural Network”* peneliti menggunakan metode CNN untuk membandingkan tingkat akurasi dengan membedakan ukuran *batch size* dan *learning rate* pada empat model. Model pertama *learning rate* : 0.0001 dan *batchsize* : 20, model kedua *learning rate* : 0.001 dan *batchsize* : 20, model ketiga *learning rate* : 0.0001 dan *batchsize* : 60, model keempat *learning rate* : 0.001 dan *batchsize* : 60. Nilai akurasi yang dihasilkan oleh keempat model tersebut secara urut adalah 25%, 25%, 85%, 98%. Sehingga, dapat disimpulkan bahwa jika nilai learning rate ditingkatkan dan memiliki data yang cukup banyak maka hasil nilai akurasi akan optimal.

Pada penelitian *“An Image Classification Approach Based on Deep Learning and Transfer Learning”* peneliti menggunakan metode *Convolutional Neural Network* (CNN) dengan model Arsitektur Inception V3 [8]. Penelitian tersebut memiliki dua model yang berbeda, model pertama peneliti menggunakan metode *ensemble* CNN dan *transfer learning*, untuk model kedua peneliti hanya menggunakan metode *ensemble* CNN dimana masing-masing model memiliki nilai akurasi sebesar 88.3% dan 60.2%.

Pada penelitian yang berjudul *“A Full Stage Data Augmentation Method in Deep Convolutional Neural Network for Natural Image Classification”* peneliti menambahkan proses data augmentasi tahap penuh untuk meningkatkan nilai akurasi pada metode CNN [9]. Peneliti membagi dengan membedakan dataset antara lain CIFAR-10 dengan dua model yang berbeda, model pertama CIFAR-10 dengan melakukan proses augmentasi data, model kedua dataset CIFAR-10 menggunakan metode *dropout*, untuk dataset kedua yaitu CIFAR-100, memiliki tiga model yang berbeda, model pertama menggunakan metode CNN pada umumnya, model kedua dataset melakukan proses data augmentasi, model ketiga dataset menggunakan metode *dropout*. Dari kedua model dataset CIFAR-10 model tersebut memiliki nilai rata-rata akurasi adalah 89.74% dan 88.27%. sedangkan pada ketiga model dataset CIFAR-100 menghasilkan nilai akurasi 65.6%, 63.06%, dan 63.01%.

Pada penelitian yang berjudul *“A Study on Image Classification Based on Deep Learning and Tensorflow”* peneliti menggunakan metode *Deep Convolutional Neural Network* [4]*.* Dataset dari penelitian tersebut menggunakan dataset yang diambil peneliti dari *ImageNet*. Data tersebut merupakan data dari 5 jenis bunga yang terdiri dari *daisy*, *dandelion*, *rose*, *sunflower*, *tulip* dengan total jumlah dataset dari kelima jenis bunga tersebut adalah 3670 gambar. Pada jurnal tersebut, peneliti menerapkan dua model yang berbeda, pada model pertama peneliti menggunakan metode sesi training *MobileNet* 0.50 dan 1.00. Rata-rata nilai akurasi yang dihasilkan pada model adalah 90%. Pada kedua model tersebut peneliti menyimpulkan bahwa ukuran model yang lebih besar akan membutuhkan lebih banyak waktu untuk menyelesaikan sesi pelatihan, tetapi meskipun memerlukan waktu yang lama, hasil nilai akurasi akan tetap lebih tinggi dibandingkan model *MobileNet* yang lebih kecil. Metode yang digunakan oleh peneliti memang sudah memiliki nilai akurasi yang cukup tinggi, tetapi dalam implementasinya jika model ukuran diubah menjadi lebih kecil, maka sesi training juga akan lebih singkat dan nilai akurasi juga akan rendah.

Pada hasil yang didapatkan pada penelitian [4], permasalahan umum yang terjadi pada metode tersebut adalah rendahnya tingkat akurasi yang disebut juga sebagai *underfitting*, salah satunya dikarenakan adanya model ukuran diubah menjadi lebih kecil sehingga sesi pelatihan atau kegiatan training data menjadi singkat dan nilai akurasi rendah.permasalahan kedua adalah terjadinya *overfitting* saat pengimplementasian suatu model dari metode CNN.[8], *overfitting* merupakan kondisi dimana model yang dibuat terlalu *rigid*/ketat dalam klasifikasi data *train*.

Sehingga, dengan adanya permasalahan yang telah dialami pada penelitian sebelumnya, penelitian ini mencoba untuk mencari solusi agar model pada metode CNN tidak mengalami *underfitting* maupun *overfitting*. Oleh karena itu, penelitian ini akan merubah arsitektur yang terdapat pada model dengan menambahkan kegaiatan oversamplingsesuai dengan kebutuhan dataset dan modeluntuk menangani permasalahan diatas. Jumlah data yang akan digunakan akan lebih besar yaitu dengan jumlah 4317 gambar

# Metode Penelitian

* 1. Pengumpulan Data

Data yang digunakan pada penelitian ini berupa flowers yang berasal dari paper rujukan dengan judul “A study on image classification based on deep learning and tensorflow” [4]. Dataset flowers berjumlah 4242 image yang terbagi menjadi beberapa kelas jenis bunga yaitu 746 image daisy, 1052 image dandelion, 784 image rose, 733 image sunflower dan 984 image tulip, sebagaimana pada Gambar 1.

(a) (b) (c) (d) (e)

Gambar 1. (a) Daisy (b) Dandelion (c) Rose (d) Sunflower (e) Tulip

* 1. Prepocessing Data

Prepocessing data merupakan Teknik awal data mining untuk mengubah data mentah menjadi informasi data lebih bersih untuk pengolahan selanjutnya. Pada prepocessing data terdiri dari proses splitting data yang terbagi menjadi tiga bagian yaitu data train, data validation dan data test dan augmentasi untuk proses resize dan normalisasi data, juga menambahkan kegiatan *oversampling* untuk mem-balance kan data, teknik oversampling yang kami gunakan menggunakan teknik manual dengan menggabungkan dataset yang terdapat pada sumber (<https://www.kaggle.com/bogdancretu/flower299>).

* 1. **Modelling**

CNN yaitu jaringan saraf multi layer yang digunakan untuk mengidentifikasi dua dimensi fitur. CNN terdiri dari lima bagian yaitu Input layer, Convolution layer, Pooling layer, Full connection layer dan Outpit layer [10].

Convolutional + Relu

Feature learning

Input

Convolutional + Relu

Pooling

(a)

Fully Connected Layer

Softmax

Flatten

(b)

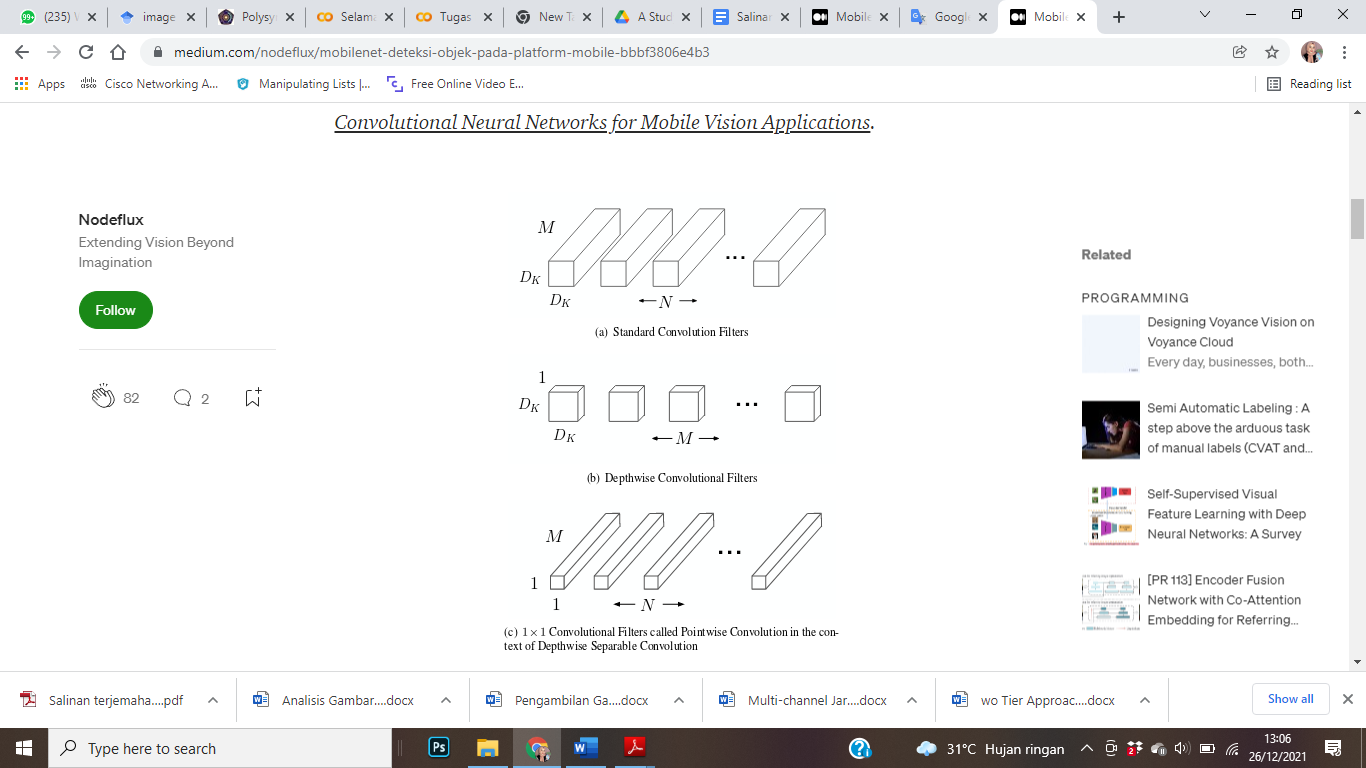
Gambar 2. Skema CNN (a) Feature Learning (b) Convolutional

Pada convolution layer terdiri dari neuron yang membentuk filter Panjang dan tinggi (pixel). Contoh terapat layer pertama feature extraction atau conv berukuran 5x5x3. Panjang 2 pixels, tinggi 5 pixels dan jumlah5 kelas sesuai channel dari image.

Pooling layer terdiri dari sebuah filter dengan ukuran stride yang akan bergeser pada seluruh area feature map. Terdapat beberapa macam Poooling layer yaitu Max Pooling, Average Pooling, Globalmaxpooling, Averagemaxpooling, dsb. Sebuah contoh jika menggunakan Maxpooling 2x2 dengan stride 2, maka setiap pergeseran filter, nilai maximum area 2x2 pixels akan dipilih dan Average memilih nilai rata-rata.

Fully Connected Layers merupakan lapisan dimana neuron aktivitas sebelumnya terhubung dengan neuron lapisan selanjutnya seperti jaringan srayaf tiruan. Setiap aktivitas dari lapisan sebelumnya dapat dihubungkan ke semua neuron di lapisan fully connected

MobileNet merupakan salah satu arsoitektur CNN untuk mengatasi kebutuhan dan computing resource berlebihan. MobileNet memiliki 2 cara mengkonfigurasi MobileNet yaitu memasukkan resolusi citra dan ukuran model dalam MobileNet [4]. MobileNet membagi konvolusi menjadi depthwise concolution dan pointwise concolution.



Gambar 3. Struktur konvolusi standart 2 lapisan depthwise dan

Dropout merupakan proses mencegah terjadinya overfiting danmempercepat proses learning. Dropout menghilangkan neuron berupa hidden layer yang visible dalam jaringan.

# 3. Hasil dan Pembahasan

Tujuan utama dari peneliti untuk mengklasifikasi citra flowers yang terdiri dari 5 kelas yaitu tulip, sunflowers, rose, dandelion, daisy . Terdapat empat fase selama proses yaitu menggunakan Tensorflow sebagai perangkat lunak sumber terbuka & Python sebagai Bahasa pemograman, mengumpulkan beberapa citra, menerapkan metode CNN dan pengklasifikasian citra kedalam kelompoknya.

3.1 Split Data

Pembagian data terbagi menjadi 80% data training, 19% data validation, 1% data testing

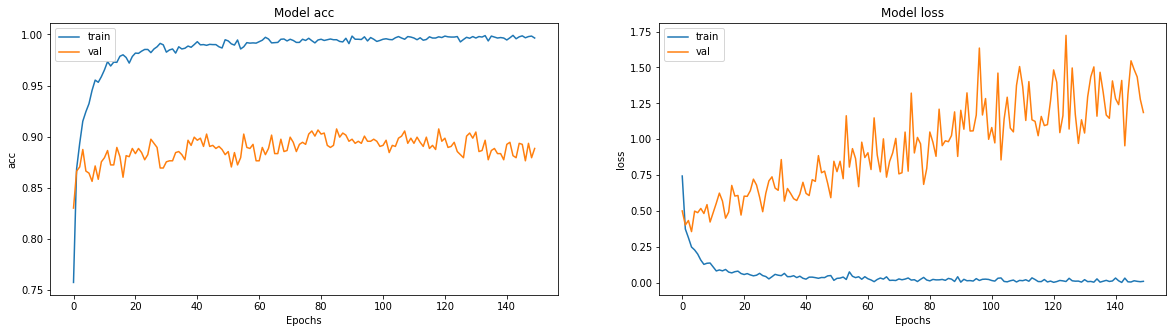
3.2 Modeling

Pelatihan menggunakan model MobileNet dengan jumlah shape 150, 150, 3 dan layer trainable false.

3.4 Hasil Ploting

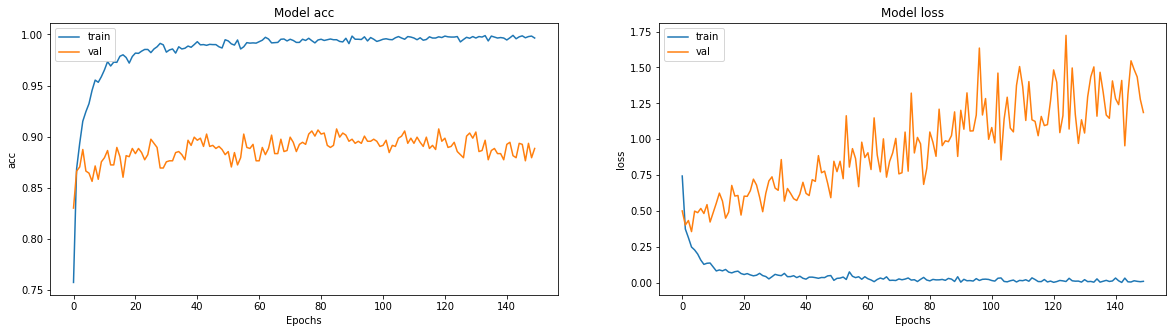
3.4.1 Accurasy Plot

Setelah peneliti menambahkan kegaiatan oversampling pada preprocessing, hasil grafik pada nilai akurasi dan loss masih menunjukkan overfitting, tetapi untuk penanganan underfitting peneliti berhasil menangani dengan menambahkan kegiatan data augmentasi pada bagian preprocessing, agar gambar atau dataset memiliki lebih banyak variasi.



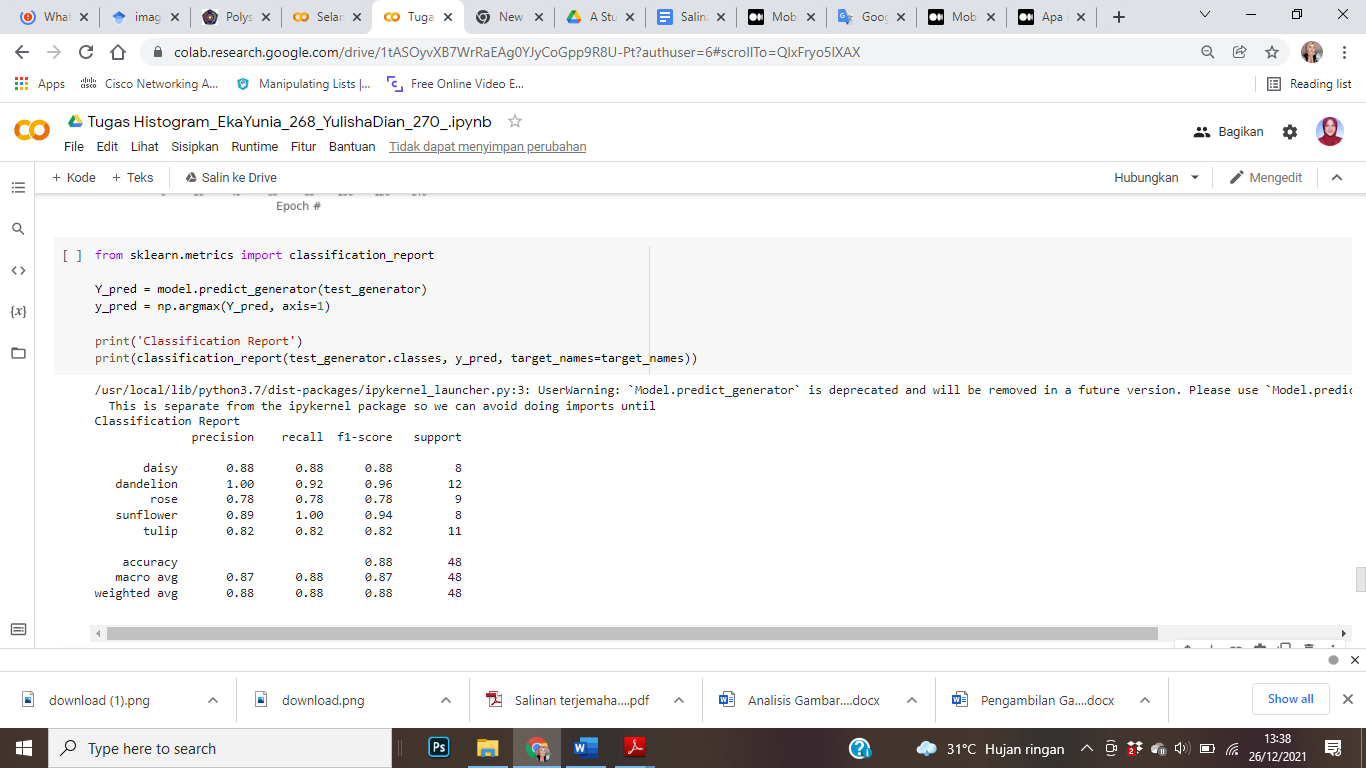
Gambar 5. Hasil accuracy

3.4.2 Loss Plot



Gambar 6. Hasil loss

3.5 Classification Report



Gambar 7. Hasil classification report

Hasil classification report menunjukkan bahwa hasil akurasi terbaik menggunakan data flowers berjumlah 5 kelas jenis daisy, dandelion, rose, sunflowers, tulip secara keseluruhan mendapatkan nilai 88%. Dan berikut hasil dari Precission, Recall, dan F1-score.

|  |  |  |
| --- | --- | --- |
| **NO** | **Flowers** | **Precission** |
| 1. | DAISY | 0.88% |
| 2. | DANDELION | 1.00% |
| 3. | ROSE | 0.78% |
| 4. | SUNFLOWER | 0.89% |
| 5. | TULIP | 0.82% |

Tabel 8. Hasil Precission

|  |  |  |
| --- | --- | --- |
| **NO** | **Flowers** | **Recall** |
| 1. | DAISY | 0.88% |
| 2. | DANDELION | 0.92% |
| 3. | ROSE | 0.78% |
| 4. | SUNFLOWER | 1.00% |
| 5. | TULIP | 0.82% |

Tabel 9. Hasil Recall

|  |  |  |
| --- | --- | --- |
| **NO** | **Flowers** | **F1-Score** |
| 1. | DAISY | 0.88% |
| 2. | DANDELION | 0.96% |
| 3. | ROSE | 0.78% |
| 4. | SUNFLOWER | 0.94% |
| 5. | TULIP | 0.82% |

Tabel 10. Hasil F1-Score

# 4. Kesimpulan

Penelitian ini tentang klasifikasi citra dengan menggunakan deep learning melalui framework Tensorflow. Tujuan penelitian ini untuk memperoleh hasil terbaik menggunakan model MobileNet metode CNN. MobileNet merupakan model transfer learning yang sering dipakai untuk meningkatkan akurasi. Teknik CNN juga digunakan lebih detail mulai assembling, training model, dan mengklasifikasikan citra. Peran epoch sangat penting untuk mengontrol akurasi dan mencegah masalah overfitting sehingga bisa mendapatkan akurasi terbaik sebesar 88% terhadap lima jenis bunga. Dan mendapat hasil precission 50% daisy, 52% dandelion, 88% rose, 1.00% sunflowers, 1.00% tulip dan mendapat nilai recall 88% daisy, 92% dandelion, 78% rose, 13% sunflowers, 98% tulip dengan nilai 62%. Dengan akurasi keseluruhan 88%. Dapat dilihat pada pembahasan hasil diatas, pada penelitian kali ini, peneliti masih belum bisa menangani permasalahan overfitting hanya dengan penambahan kegiatan oversampling pada bagian preprocessing, kegiatan oversampling juga masih dilakukan secara manual. Sehingga, untuk kedepannya peneliti mencoba untuk menambahkan layer Batchnormalization dan layer Dropout untuk mengatasi permasalahan overfitting diatas serta mengganti teknik oversampling yang manual menjadi otomatis langsung melalui program.

# Daftar Rujukan

[1] E. Design, “模具制造中的电极设计与加工 2 2 2,” pp. 86–89, 2009.

[2] T. P. M. F. Kelly, “Struktur Tumbuhan,” *Angew. Chemie Int. Ed. 6(11), 951–952.*, 2020.

[3] M. Zainuri and D. P. Pamungkas, “Implementasi Metode Convolutional Neural Network (CNN) Untuk Klasifikasi Jenis Bunga Anggrek,” *Seminar Nasional Inovasi Teknologi*. pp. 87–92, 2020.

[4] M. A. Abu, N. H. Indra, A. H. A. Rahman, N. A. Sapiee, and I. Ahmad, “A study on image classification based on deep learning and tensorflow,” *Int. J. Eng. Res. Technol.*, vol. 12, no. 4, pp. 563–569, 2019.

[5] A. Peryanto, A. Yudhana, and R. Umar, “Klasifikasi Citra Menggunakan Convolutional Neural Network dan K Fold Cross Validation,” *J. Appl. Informatics Comput.*, vol. 4, no. 1, pp. 45–51, 2020, doi: 10.30871/jaic.v4i1.2017.

[6] A. Perbandingan, A. Svm, and D. A. N. C. N. N. Untuk, “Comparative Analysis of Image Classification Algorithm for,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 2, pp. 311–318, 2021, doi: 10.25126/jtiik.202184553.

[7] Y. Niu, X. Hu, S. Chen, and C. Yang, “Research on Flower Retrieval Based on Deep Learning,” pp. 135–140, 2019, doi: 10.12792/iciae2019.027.

[8] K. Han, J. He, Y. Wang, Y. Xiong, and C. Zhang, “An Image Classification Approach based on Deep Learning and Transfer Learning,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 768, no. 7, 2020, doi: 10.1088/1757-899X/768/7/072055.

[9] Q. Zheng, M. Yang, X. Tian, N. Jiang, and D. Wang, “A full stage data augmentation method in deep convolutional neural network for natural image classification,” *Discret. Dyn. Nat. Soc.*, vol. 2020, 2020, doi: 10.1155/2020/4706576.

[10] J. Feng, Z. Wang, M. Zha, and X. Cao, “Flower recognition based on transfer learning and ADam deep learning optimization algorithm,” *ACM Int. Conf. Proceeding Ser.*, pp. 598–604, 2019, doi: 10.1145/3366194.3366301.