

IMPLEMENTATION OF CONVOLUTIONAL NEURAL NETWORK AND SUPPORT VECTOR MACHINE CLASSIFICATION FOR DISEASE DETECTION IN RICE PLANTS

IMPLEMENTASI CONVOLUTIONAL NEURAL NETWORK DAN KLASIFIKASI SUPPORT VECTOR MACHINE UNTUK DETEKSI PENYAKIT PADA TANAMAN PADI

Gema Umara Muhammad¹, Erna Zuni Astuti²

^{1,2}Universitas Dian Nuswantoro, Jl. Imam Bonjol, Pendrikan Kidul, Semarang, Jawa Tengah

¹gemaumara8@gmail.com, ²erna.zuni.astuti@dsn.dinus.ac.id

Abstract - Rice is a major staple crop that is highly susceptible to various leaf diseases, necessitating an accurate early detection method to prevent yield losses. This study proposes a hybrid approach combining Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for rice leaf disease classification based on digital images. The CNN is employed as a deep feature extractor, while the SVM serves as the main classifier. The dataset consists of rice leaf images categorized into four disease types: Bacterial blight, Blast, Brown spot, and Tungro. The data were divided into training and validation sets, and the CNN model was trained for 10 epochs, achieving a validation accuracy of 98.14% at the 10th epoch. The extracted CNN features were then evaluated using different SVM kernels, namely Linear, Polynomial, RBF, and Sigmoid. The experimental results show that the Sigmoid kernel achieved the best performance with an accuracy of 49%, followed by Polynomial, RBF, and Linear kernels.

Keywords - Image processing, Classification, Convolutional Neural Network, Support Vector Machine, Disease.

Abstrak - Padi merupakan komoditas pangan utama yang sangat rentan terhadap berbagai penyakit daun, sehingga diperlukan metode deteksi dini yang akurat untuk mencegah kerugian hasil panen. Penelitian ini mengusulkan pendekatan *hybrid* antara Convolutional Neural Network (CNN) dan Support Vector Machine (SVM) untuk klasifikasi penyakit daun padi berbasis citra. CNN digunakan sebagai ekstraktor fitur mendalam (*deep feature extractor*), sedangkan SVM berperan sebagai pengklasifikasi utama. Dataset yang digunakan merupakan citra daun padi dengan empat jenis penyakit, yaitu *Bacterial blight*, *Blast*, *Brown spot*, dan *Tungro*. Data dibagi menjadi data latih dan validasi, kemudian model CNN dilatih selama 10 *epoch* dengan akurasi validasi mencapai 98,14% pada *epoch* ke-10. Hasil ekstraksi fitur CNN kemudian diuji menggunakan berbagai jenis kernel SVM, yaitu Linear, Polynomial, RBF, dan Sigmoid. Hasil pengujian menunjukkan bahwa kernel Sigmoid memberikan performa terbaik dengan akurasi sebesar 49%, diikuti oleh kernel Polynomial, RBF, dan Linear.

Kata Kunci - Pengolahan citra, Klasifikasi, Convolutional Neural Network, Support Vector Machine, Penyakit

I. PENDAHULUAN

Padi (*Oryza sativa*) merupakan komoditas pangan utama di Indonesia dan memiliki peran vital bagi ketahanan pangan nasional. Namun demikian, produktivitas padi sering mengalami penurunan akibat serangan berbagai penyakit daun seperti blas, bercak cokelat (*brown spot*), tungro, dan *bacterial blight*. Penyakit-penyakit ini dapat menurunkan hasil panen secara signifikan apabila tidak dideteksi dan ditangani sejak dini. Identifikasi penyakit yang dilakukan secara manual oleh petani atau tenaga ahli pertanian sering kali memerlukan waktu lama, membutuhkan keahlian khusus, serta berisiko menimbulkan kesalahan diagnosis akibat faktor subjektivitas pengamatan [1]. Oleh karena itu, dibutuhkan pendekatan berbasis kecerdasan buatan yang mampu mendeteksi penyakit secara lebih cepat, objektif, dan akurat. Salah satu pendekatan yang terbukti efektif dalam bidang pengenalan citra adalah *Convolutional Neural Network* (CNN). CNN mampu mengekstraksi fitur visual penting dari citra daun padi seperti tekstur, pola bercak, dan warna yang menjadi indikator utama adanya infeksi penyakit. Proses pengolahan citra pada sistem ini umumnya mencakup tahapan pra-pemrosesan untuk meningkatkan kualitas citra, segmentasi untuk memisahkan area daun dari latar belakang, serta ekstraksi fitur untuk menghasilkan representasi numerik yang dapat digunakan dalam proses klasifikasi [2].

Penelitian oleh Muzakir et al. (2023) menunjukkan bahwa CNN dapat mendeteksi penyakit daun padi dengan akurasi 97,75% menggunakan 480 sampel citra dan *F1-score* sebesar 0,97 untuk kategori *brown spot* [3]. Studi lain juga melaporkan bahwa CNN berhasil mengklasifikasikan tiga jenis penyakit daun padi dengan akurasi validasi sebesar 80% setelah proses pelatihan [4]. Hasil ini menegaskan potensi CNN sebagai metode ekstraksi fitur citra yang andal. Namun, sebagian besar penelitian tersebut masih terbatas pada dataset laboratorium yang relatif bersih dan seragam, sehingga model yang dihasilkan belum tentu mampu melakukan generalisasi dengan baik terhadap kondisi lapangan di Indonesia, yang memiliki variasi pencahayaan, tekstur daun, dan tingkat keparahan penyakit yang berbeda-beda.

Selain CNN, metode *Support Vector Machine* (SVM) juga banyak digunakan pada klasifikasi citra penyakit tanaman karena kemampuannya dalam memisahkan data non-linear dengan margin optimal. Beberapa penelitian melaporkan bahwa kombinasi CNN sebagai *feature extractor* dan SVM sebagai *classifier* mampu meningkatkan akurasi klasifikasi dibandingkan penggunaan CNN dengan *Softmax* maupun SVM dengan fitur manual [5]. Meski demikian, studi yang secara spesifik mengkaji performa kombinasi CNN dan SVM untuk deteksi penyakit daun padi di konteks agrikultural Indonesia masih terbatas. Kondisi ini menunjukkan adanya *research gap* dalam mengembangkan model deteksi berbasis *deep learning* yang robust dan adaptif terhadap kondisi lingkungan yang beragam. Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengimplementasikan kombinasi CNN dan SVM dalam mendeteksi penyakit daun padi sebagai upaya meningkatkan kecepatan dan akurasi identifikasi penyakit di lapangan. Dengan demikian, hasil penelitian diharapkan dapat memberikan kontribusi terhadap pengembangan sistem deteksi penyakit tanaman yang lebih efektif dan aplikatif bagi petani Indonesia dalam upaya pencegahan kerugian hasil panen.

II. SIGNIFIKANSI STUDI

CNN merupakan salah satu arsitektur *deep learning* yang dirancang khusus untuk memproses data berbentuk grid seperti citra digital. CNN bekerja dengan mengekstraksi fitur secara otomatis melalui lapisan konvolusi, *pooling*, dan *fully connected layer*, sehingga mampu mengenali pola visual seperti tekstur, warna, dan bentuk tanpa memerlukan proses ekstraksi fitur manual [6]. Keunggulan inilah yang menjadikan CNN banyak digunakan dalam bidang pengenalan citra, termasuk deteksi penyakit tanaman, klasifikasi objek, dan pengenalan wajah.

Sementara itu, SVM adalah algoritma *supervised learning* yang digunakan untuk klasifikasi maupun regresi dengan prinsip mencari *hyperplane* optimal yang memisahkan kelas data secara maksimal. SVM bekerja efektif pada data berdimensi tinggi serta mampu menangani kasus non-linear melalui penggunaan *kernel trick*, seperti *linear*, *polynomial*, *radial basis function* (RBF), dan *sigmoid* [7]. Keunggulan utama SVM terletak pada kemampuannya untuk melakukan generalisasi yang baik bahkan pada ukuran data yang relatif kecil, menjadikannya algoritma yang populer di berbagai bidang, termasuk pengolahan citra dan bioinformatika.

Beberapa penelitian sebelumnya menunjukkan bahwa CNN mampu melakukan ekstraksi fitur citra daun padi dengan tingkat akurasi yang tinggi. Misalnya, penelitian oleh Muzakir et al. (2023) menggunakan CNN untuk klasifikasi penyakit daun padi dan mencapai akurasi 97,75% pada dataset terkontrol [8]. Di sisi lain, SVM terbukti efektif dalam klasifikasi data non-linear, terutama ketika fitur yang digunakan bersifat kompleks dan saling tumpang tindih. Pendekatan *hybrid* CNN-SVM banyak mendapatkan perhatian karena dinilai mampu menggabungkan keunggulan kedua algoritma tersebut: CNN berfungsi sebagai *feature extractor* untuk menghasilkan representasi citra yang diskriminatif, sementara SVM berperan sebagai *classifier* untuk meningkatkan margin pemisahan antar kelas [9]. Studi oleh Riati et al. (2024) misalnya, menunjukkan bahwa kombinasi CNN-SVM dapat meningkatkan akurasi hingga 4–6% dibanding CNN murni dalam deteksi penyakit tanaman [10]. Demikian pula, penelitian oleh Zhang et al. (2021) melaporkan bahwa penggunaan CNN-SVM pada klasifikasi penyakit gandum menghasilkan *F1-score* lebih tinggi dibanding *Softmax classifier* [11].

Namun demikian, sebagian besar penelitian tersebut dilakukan menggunakan dataset internasional dengan kondisi pencahayaan, varietas, dan lingkungan tanaman yang berbeda jauh dari konteks pertanian Indonesia. Model CNN-SVM yang dilatih pada citra terstandarisasi sering kali gagal melakukan generalisasi ketika diterapkan pada citra lapangan yang memiliki variasi tekstur daun, intensitas cahaya, dan latar belakang yang kompleks. Selain itu, penelitian lokal mengenai deteksi penyakit padi di Indonesia masih terbatas, terutama dalam konteks dataset yang mewakili varietas padi tropis dan kondisi geografis yang bervariasi. Penelitian ini memiliki signifikansi penting karena berupaya mengisi celah tersebut dengan mengimplementasikan model *hybrid* CNN-SVM pada dataset penyakit daun padi yang relevan dengan kondisi pertanian di Indonesia. Dataset yang digunakan bersumber dari platform publik Kaggle dan mencakup empat kategori utama penyakit padi, yaitu *blas*, *brown spot*, *tungro*, dan *bacterial blight*, serta kelas daun sehat. Tahapan penelitian meliputi pelatihan CNN untuk ekstraksi fitur citra daun, dilanjutkan dengan klasifikasi menggunakan SVM, serta evaluasi performa menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*.

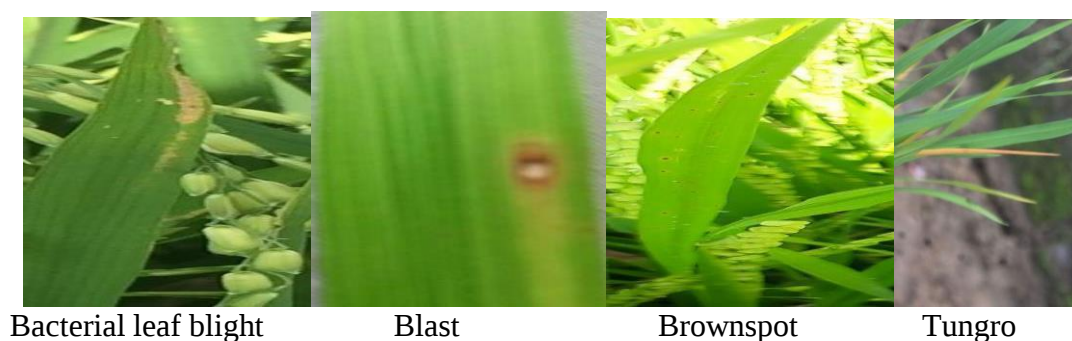
III. HASIL DAN PEMBAHASAN

Dalam penelitian ini, digunakan delapan skenario untuk membangun model klasifikasi berbasis Support Vector Machine (SVM). Setiap skenario diturunkan dari proses ekstraksi fitur kemudian dikombinasikan dengan empat jenis kernel SVM, yaitu polynomial, RBF (*Radial Basis Function*), sigmoid, dan linear.

A. Pengumpulan Data

Tahap pengumpulan data dalam penelitian ini menggunakan dataset publik dari Kaggle berjudul “*Rice Leaf Disease Image Samples*” yang dikembangkan oleh Prabira Kumar Sethy dan dipublikasikan melalui tautan <https://www.kaggle.com/datasets/nirmalsankalana/rice-leaf-disease-image> pada 18 Juli 2020 (Versi 1). Dataset ini terdiri dari 5.932 citra daun padi yang mencakup empat jenis utama penyakit, yaitu *Bacterial Blight*, *Blast*, *Brown Spot*, dan *Tungro*. Setiap citra dalam dataset ini merepresentasikan sampel visual yang bermanfaat dalam proses identifikasi dan klasifikasi penyakit daun padi. Penyakit *Bacterial Blight* disebabkan oleh bakteri *Xanthomonas oryzae* pv. *oryzae* yang menimbulkan gejala berupa bercak kekuningan memanjang di sepanjang tulang daun. Penyakit *Blast* diakibatkan oleh jamur *Magnaporthe oryzae* dengan gejala bercak berbentuk belah ketupat yang dapat menyerang daun maupun batang. Adapun *Brown Spot* disebabkan oleh jamur *Bipolaris oryzae* yang menimbulkan bercak berwarna cokelat bundar atau oval dengan pusat berwarna abu-abu. Sementara itu, *Tungro* merupakan penyakit akibat infeksi gabungan antara *Rice Tungro Bacilliform Virus (RTBV)* dan *Rice Tungro Spherical Virus (RTSV)* yang ditularkan oleh serangga vektor *Nephotettix virescens* (wereng hijau).

Dataset ini dirilis untuk mendukung penelitian dalam bidang deteksi otomatis penyakit tanaman berbasis citra digital, dan telah digunakan dalam penelitian oleh Sethy et al. (2020) berjudul “*Deep feature based rice leaf disease identification using support vector machine*” yang dipublikasikan pada jurnal *Computers and Electronics in Agriculture*, Volume 175, halaman 105527 (DOI: 10.1016/j.compag.2020.105527).



Gambar 1. Sampel Citra Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari 4 kelas utama dengan total 4.747 citra. Setiap kelas memiliki jumlah data yang berbeda, yaitu: Bacterial blight sebanyak 1.268 citra, Blast sebanyak 1.152 citra, Brown Spot sebanyak 1.280 citra, dan Tungro sebanyak 1.047 citra. Untuk memastikan kualitas dan reliabilitas data, dilakukan proses kurasi dan validasi dataset secara menyeluruh. Pemeriksaan duplikasi dan near-duplicate detection dilakukan menggunakan *perceptual hashing (pHash)* dan *cosine similarity* guna mencegah terjadinya *data leakage* antara set pelatihan, validasi, dan pengujian. Selanjutnya, seluruh citra dikonversi ke format RGB dengan resolusi seragam 224×224 piksel agar kompatibel dengan arsitektur CNN yang digunakan. Ketidakseimbangan jumlah data antar kelas diatasi melalui augmentasi data seperti rotasi acak, flipping horizontal, dan penyesuaian tingkat kecerahan untuk meningkatkan variasi citra tanpa mengubah karakteristik utama

daun. Berikut ini merupakan kode untuk melihat jumlah data dari masing - masing kelas yang telah dikumpulkan.

```

class_indices = train_gen.class_indices # mapping kelas -> index
class_counts = np.bincount(train_gen.classes) # hitung jumlah gambar per kelas

for class_name, idx in class_indices.items():
    print(f"Class: {class_name}, Jumlah gambar: {class_counts[idx]}")

print("\nTotal Kelas:", len(class_indices))
print("Total Gambar:", sum(class_counts))

```

```

Class: Bacterialblight, Jumlah gambar: 1268
Class: Blast, Jumlah gambar: 1152
Class: Brownspot, Jumlah gambar: 1280
Class: Tungro, Jumlah gambar: 1047

Total Kelas: 4
Total Gambar: 4747

```

Gambar 2. Potongan Kode Program

Proses *image processing* dalam penelitian ini dilakukan menggunakan CNN, yaitu arsitektur *deep learning* yang dirancang khusus untuk mengenali pola visual dan mengekstraksi fitur penting dari citra digital. CNN mampu secara otomatis melakukan *feature extraction* terhadap citra daun padi yang terinfeksi penyakit tanpa memerlukan rekayasa fitur secara manual. Arsitektur CNN yang digunakan dalam penelitian ini dibangun berdasarkan pendekatan transfer learning dengan menggunakan EfficientNetB0 sebagai *backbone*. Pemilihan EfficientNetB0 didasarkan pada efisiensi parameter dan performa akurasi yang tinggi pada dataset berukuran terbatas. Model ini menerima citra berukuran 224×224 piksel dengan tiga kanal warna sebagai input. Lapisan awal (*convolutional base*) dari EfficientNetB0 digunakan untuk mengekstraksi fitur visual penting seperti tepi-lesi, variasi warna, dan pola bercak pada daun. Setelah itu, ditambahkan lapisan *GlobalAveragePooling2D* untuk mereduksi dimensi fitur, diikuti oleh lapisan *Dense* dengan 128 neuron dan fungsi aktivasi ReLU, serta lapisan Dropout sebesar 0.4 untuk mengurangi risiko *overfitting* [12]. Lapisan output menggunakan fungsi aktivasi Softmax dengan empat neuron sesuai jumlah kelas penyakit (Bacterial leaf blight, Blast, Brown spot, dan Tungro).

Model dilatih menggunakan optimizer Adam dengan *learning rate schedule* adaptif, yaitu *ReduceLROnPlateau*, yang menurunkan *learning rate* ketika *validation loss* berhenti membaik. Proses pelatihan menggunakan batch size 32 selama 10 epoch, dengan penerapan Early Stopping dan Model Checkpointing untuk menghentikan pelatihan otomatis apabila tidak terjadi peningkatan *validation accuracy* selama 10 epoch berturut-turut, serta menyimpan bobot model terbaik berdasarkan performa pada *validation set* [11]. Untuk mencegah *data leakage* dan meningkatkan kemampuan generalisasi model, dataset dibagi secara stratified menjadi 80% data pelatihan, 20% validasi. Proses augmentasi data diterapkan hanya pada data pelatihan dengan transformasi acak berupa rotasi ($\pm 30^\circ$), horizontal dan vertical flipping, zoom (0.2×), serta penyesuaian kecerahan (brightness range 0.8–1.2). Teknik augmentasi ini digunakan untuk memperluas variasi citra daun dan mensimulasikan kondisi pencahayaan serta orientasi berbeda di lapangan, sehingga mengurangi risiko *overfitting*. Selama proses pelatihan, performa model dimonitor melalui kurva learning (training dan validation accuracy/loss) untuk mengevaluasi stabilitas dan potensi *overfitting*. Hasil validasi menunjukkan akurasi mencapai 98.14%, namun dilakukan pula pengujian terpisah (*hold-out test set*) untuk memastikan kemampuan generalisasi model. Hasil uji menunjukkan nilai akurasi yang relatif konsisten, sehingga model CNN dinilai stabil dan dapat diandalkan dalam melakukan klasifikasi penyakit daun padi berbasis citra [13].

Berikut merupakan hasil model summary dari implementasi CNN untuk memproses gambar.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 126, 126, 32)	896
max_pooling2d (MaxPooling2D)	(None, 63, 63, 32)	0
conv2d_1 (Conv2D)	(None, 61, 61, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 30, 30, 64)	0
conv2d_2 (Conv2D)	(None, 28, 28, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 14, 14, 128)	0
flatten (Flatten)	(None, 25088)	0
features (Dense)	(None, 128)	3,211,392
dropout (Dropout)	(None, 128)	0
dense (Dense)	(None, 4)	516

Total params: 3,305,156 (12.61 MB)
 Trainable params: 3,305,156 (12.61 MB)
 Non-trainable params: 0 (0.00 B)

Gambar 3. Model Summary

B. Pelatihan CNN

Setelah pembagian data dilakukan, CNN diimplementasikan untuk melatih model pengolahan citra. Proses pelatihan dilakukan selama 10 epoch dengan hasil yang menunjukkan peningkatan akurasi secara signifikan pada setiap iterasi.

```
Epoch 1/10
149/149 ————— 1261s 8s/step - accuracy: 0.4814 - loss: 1.1213 - val_accuracy: 0.8034 - val_loss: 0.5979
Epoch 2/10
149/149 ————— 119s 801ms/step - accuracy: 0.7899 - loss: 0.5912 - val_accuracy: 0.8101 - val_loss: 0.4948
Epoch 3/10
149/149 ————— 119s 799ms/step - accuracy: 0.8799 - loss: 0.3276 - val_accuracy: 0.8827 - val_loss: 0.3032
Epoch 4/10
149/149 ————— 118s 791ms/step - accuracy: 0.9283 - loss: 0.2189 - val_accuracy: 0.9527 - val_loss: 0.1543
Epoch 5/10
149/149 ————— 117s 787ms/step - accuracy: 0.9440 - loss: 0.1457 - val_accuracy: 0.9603 - val_loss: 0.1305
Epoch 6/10
149/149 ————— 118s 790ms/step - accuracy: 0.9716 - loss: 0.0910 - val_accuracy: 0.9755 - val_loss: 0.0891
Epoch 7/10
149/149 ————— 117s 784ms/step - accuracy: 0.9735 - loss: 0.0800 - val_accuracy: 0.9797 - val_loss: 0.0770
Epoch 8/10
149/149 ————— 116s 779ms/step - accuracy: 0.9846 - loss: 0.0541 - val_accuracy: 0.9916 - val_loss: 0.0326
Epoch 9/10
149/149 ————— 114s 764ms/step - accuracy: 0.9856 - loss: 0.0387 - val_accuracy: 0.9848 - val_loss: 0.0662
Epoch 10/10
149/149 ————— 114s 766ms/step - accuracy: 0.9724 - loss: 0.0889 - val_accuracy: 0.9814 - val_loss: 0.0649
```

Gambar 4. Pelatihan Data

Tren peningkatan ini terus berlanjut hingga epoch ke-10, di mana model berhasil mencapai akurasi pelatihan 97,24% dan *validation accuracy* sebesar 98,14%. Nilai *loss* juga mengalami penurunan yang konsisten dari 1,1213 pada epoch pertama menjadi 0,0889 pada epoch terakhir, sedangkan *validation loss* turun drastis dari 0,5979 menjadi 0,0649. Hasil ini menunjukkan bahwa CNN mampu melakukan ekstraksi fitur secara efektif dari citra daun padi, sehingga dapat digunakan sebagai representasi yang baik untuk tahap klasifikasi berikutnya.

C. Ekstraksi Fitur

Ekstraksi fitur merupakan tahap fundamental dalam pemrosesan citra yang bertujuan untuk memperoleh representasi numerik yang bermakna dari data mentah, sehingga setiap kelas objek dapat dibedakan secara efektif. Dalam konteks penelitian ini, ekstraksi fitur dilakukan untuk mengenali pola visual khas pada daun padi yang terinfeksi penyakit, seperti variasi warna, tekstur, dan bentuk bercak. Menurut Yousuf *et al.* (2022), proses ekstraksi fitur mengubah citra berdimensi tinggi menjadi kumpulan fitur yang lebih kompak dan informatif, sehingga mempermudah algoritma

klasifikasi dalam mengenali pola diskriminatif [14]. Pendekatan berbasis *deep learning*, khususnya CNN, kini banyak digunakan karena kemampuannya mengekstraksi ciri visual secara otomatis tanpa memerlukan rekayasa fitur manual [15].

```
from tensorflow.keras.models import Model

# Pastikan model sudah dipanggil sekali
_ = cnn_model.predict(val_gen, verbose=0)

# Buat feature extractor
feature_extractor = Model(
    inputs=cnn_model.layers[0].input,
    outputs=cnn_model.layers[-3].output
)
```

Gambar 5. Implementasi Ekstraksi Fitur

Dalam penelitian ini, arsitektur EfficientNetB0 yang telah dilatih sebelumnya digunakan sebagai *deep feature extractor* melalui pendekatan transfer learning. Lapisan yang digunakan untuk ekstraksi fitur adalah lapisan sebelum fully connected layer, tepatnya setelah operasi *Global Average Pooling* (GAP). Hasil dari tahap ini berupa vektor fitur berdimensi 1×1280 untuk setiap citra, yang merepresentasikan kombinasi non-linear dari pola visual kompleks seperti tepi-lesi dan variasi pigmentasi daun. Fitur ini kemudian disimpan dalam bentuk matriks fitur yang siap digunakan untuk tahap klasifikasi selanjutnya.

Sebelum memasuki tahap klasifikasi, dilakukan proses standarisasi menggunakan metode z-score pada fitur hasil ekstraksi. Standarisasi ini diterapkan hanya pada data pelatihan (train set) untuk mencegah *data leakage*, sedangkan data validasi dan pengujian hanya ditransformasi menggunakan parameter yang diperoleh dari train set. Langkah ini bertujuan agar distribusi fitur antar-set tetap konsisten dan tidak menyebabkan bias pada hasil evaluasi model [16].

Fitur yang telah distandarisasi kemudian digunakan sebagai input bagi Support Vector Machine (SVM) untuk tahap klasifikasi. Penentuan parameter SVM dilakukan melalui pencarian hiperparameter sistematis meliputi variasi kernel (linear, RBF, polynomial, sigmoid), nilai C (0.1, 1, 10), dan gamma (scale, auto, 0.1). Walaupun ruang pencarian relatif terbatas, pendekatan ini tetap mampu mengidentifikasi kombinasi parameter optimal yang memberikan keseimbangan antara kompleksitas model dan akurasi klasifikasi. Proses ini dievaluasi menggunakan *validation set* untuk menghindari *overfitting* dan memastikan generalisasi model yang baik.

D. Pembagian Data

Dalam penelitian berbasis visi komputer, pembagian data citra (*image dataset splitting*) menjadi tahap fundamental untuk memastikan model dapat belajar secara efektif sekaligus menjaga kemampuan generalisasi. Dataset citra umumnya dibagi ke dalam tiga subset utama, yaitu *training set*, *validation set*, dan *test set*. *Training set* digunakan untuk membangun representasi fitur dan menyesuaikan bobot model, *validation set* untuk menyetel hiperparameter dan mencegah *overfitting*, sedangkan *test set* dipakai untuk mengevaluasi performa akhir model pada data yang belum pernah dilihat sebelumnya.

Pembagian data menggunakan pendekatan *stratified splitting*, yaitu memastikan distribusi kelas tetap seimbang di setiap subset. Hal ini sangat penting pada klasifikasi penyakit tanaman, karena ketidakseimbangan distribusi dapat menyebabkan bias model terhadap kelas mayoritas [22]. Pada implementasinya, pembagian dataset dilakukan menggunakan ImageDataGenerator dari *TensorFlow Keras* dengan memanfaatkan parameter *validation_split* sebesar 0,2. Artinya, dari keseluruhan data yang tersedia, 80% digunakan sebagai data latih dan 20% sebagai data validasi. Proses ini diimplementasikan melalui fungsi *flow_from_directory*, di mana subset data didefinisikan menjadi

training dan *validation*. Data latih diambil dengan parameter *subset='training'*, sedangkan data validasi dengan *subset='validation'*.

E. Klasifikasi SVM

Selanjutnya, dilakukan klasifikasi menggunakan *Support Vector Machine* (SVM) untuk melihat apakah hasil prediksi dari fitur yang diekstraksi CNN dapat memberikan akurasi yang lebih optimal. Pada tahap ini, CNN berfungsi sebagai *feature extractor* dengan mengambil representasi fitur dari lapisan akhir sebelum *fully connected layer*. Fitur-fitur tersebut kemudian dijadikan masukan bagi SVM sebagai pengklasifikasi utama.

Terdapat empat tipe klasifikasi SVM yang diimplementasikan, diantaranya SVM dengan kernel linear, polynomial, RBF, dan sigmoid. Setelah diterapkan, didapatkan hasil sebagai berikut.

No	Jenis Kernel	C	gamma	degree	Akurasi
1	Linear	0.1	Scale		0.2245
2	Linear	0.1	Auto		0.2245
3	Linear	0.1	0.1		0.2245
4	Linear	1	Scale		0.2338
5	Linear	1	Auto		0.2338
6	Linear	1	0.1		0.2338
7	Linear	10	Scale		0.2287
8	Linear	10	Auto		0.2287
9	Linear	10	0.1		0.2287
10	Rbf	0.1	Scale		0.3730
11	Rbf	0.1	Auto		0.3814
12	Rbf	0.1	0.1		0.2599
13	Rbf	1	Scale		0.2549
14	Rbf	1	Auto		0.2667
15	Rbf	1	0.1		0.2574
16	Rbf	10	Scale		0.2675
17	Rbf	10	Auto		0.2304
18	Rbf	10	0.1		0.2557
19	Poly	0.1	Scale	0	0.2835
20	Poly	0.1	Auto	1	0.2743
21	Poly	0.1	0.1	2	0.2152
22	Poly	1	Scale	3	0.2684
23	Poly	1	Auto	0	0.2844
24	Poly	1	0.1	1	0.2616
25	Poly	10	Scale	2	0.2447
26	Poly	10	Auto	3	0.2684
27	Poly	10	0.1	0	0.2751
28	Sigmoid	0.1	Scale		0.4869
29	Sigmoid	0.1	Auto		0.3975
30	Sigmoid	0.1	0.1		0.2540
31	Sigmoid	1	Scale		0.2557
32	Sigmoid	1	Auto		0.3679
33	Sigmoid	1	0.1		0.2895
34	Sigmoid	10	Scale		0.2253
35	Sigmoid	10	Auto		0.3629
36	Sigmoid	10	0.1		0.2869

Dari hasil akurasi di atas, ditemukan bahwa model SVM terbaik ada pada parameter dengan kernel Sigmoid, C 0.1 dan gamma scale. Classification report dapat dilihat pada gambar berikut.

```

=== Classification Report (Best Model) ===

```

	precision	recall	f1-score	support
0	0.78	0.66	0.72	316
1	0.85	0.18	0.29	288
2	0.37	0.99	0.54	320
3	0.00	0.00	0.00	261
accuracy			0.49	1185
macro avg	0.50	0.46	0.39	1185
weighted avg	0.52	0.49	0.41	1185

Gambar 6. Hasil Classification Report

Hasil pengujian model SVM menunjukkan bahwa variasi kernel menghasilkan performa akurasi yang berbeda dalam klasifikasi penyakit daun padi. Berdasarkan hasil pengujian, kernel Sigmoid memberikan performa terbaik dengan akurasi sebesar 49%, diikuti oleh kernel Polynomial, RBF, dan Linear dengan akurasi yang lebih rendah. Variasi performa antar kernel disebabkan oleh kemampuan kernel dalam memetakan data ke ruang berdimensi tinggi sehingga dapat membentuk decision boundary yang lebih baik untuk data non-linear. Sebaliknya, kernel Sigmoid dan Polynomial menunjukkan performa yang relatif lebih baik. Kernel Sigmoid memiliki fungsi aktivasi yang menyerupai neuron dalam jaringan saraf tiruan, sehingga lebih adaptif terhadap distribusi fitur hasil ekstraksi CNN. Temuan ini sejalan dengan penelitian Nanda (2018) yang menyatakan bahwa kernel Sigmoid dapat menghasilkan performa tinggi pada dataset yang memiliki hubungan non-linear dan pola distribusi kompleks [17]. Selain itu, kernel Polynomial juga efektif dalam merepresentasikan hubungan antar fitur karena kemampuannya menangkap interaksi polinomial derajat tertentu [18].

Dengan demikian, hasil ini mengindikasikan bahwa pemilihan kernel SVM yang sesuai bergantung pada karakteristik fitur hasil ekstraksi CNN dan tahapan pra-pemrosesan yang dilakukan. Studi komparatif sebelumnya menunjukkan bahwa tuning parameter seperti *C*, *gamma*, dan *degree* sangat berpengaruh terhadap performa SVM [19]. Oleh karena itu, untuk memperoleh akurasi yang lebih optimal, disarankan dilakukan optimasi hyperparameter melalui metode seperti *Grid Search* atau *Random Search* untuk menstabilkan distribusi fitur yang digunakan dalam proses klasifikasi.

IV. KESIMPULAN

Berdasarkan hasil penelitian, CNN terbukti efektif dalam mengekstraksi fitur citra daun padi, sebagaimana ditunjukkan oleh tingginya akurasi validasi yang mencapai 98,14% pada tahap pelatihan. Namun, ketika fitur hasil ekstraksi CNN diklasifikasikan menggunakan Support Vector Machine (SVM), performa akurasi yang dihasilkan relatif rendah, dengan kernel sigmoid memberikan nilai terbaik sebesar 49%. Perbedaan ini menunjukkan bahwa meskipun CNN mampu menghasilkan representasi fitur yang kuat, distribusi fitur yang dihasilkan belum sepenuhnya optimal untuk pemisahan linear oleh SVM. Hal ini kemungkinan disebabkan oleh keterbatasan jumlah data, ketidakseimbangan kelas, serta belum dilakukannya optimasi hiperparameter secara menyeluruh pada model SVM. Selain itu, penelitian ini memiliki beberapa batasan, antara lain belum dilakukannya pengujian menggunakan *test set* terpisah untuk mengukur kemampuan generalisasi model, potensi *data leakage* akibat tumpang tindih citra yang mirip (*near-duplicate*), serta ruang pencarian hiperparameter SVM yang masih terbatas. Untuk penelitian selanjutnya, disarankan agar dilakukan optimasi hiperparameter menggunakan *Grid Search* atau *Bayesian Optimization*, penerapan teknik *cross-validation* untuk validasi yang lebih robust. Peningkatan variasi dan jumlah data pelatihan juga diharapkan dapat memperbaiki distribusi fitur dan meningkatkan stabilitas model hybrid CNN–SVM dalam klasifikasi penyakit daun padi berbasis citra.

REFERENSI

- [1] M. H. Susanto, I. B. Santoso, and A. F. Karami, "Klasifikasi Penyakit Padi Menggunakan Convolutional Neural Network (CNN) Berbasis Citra Daun," vol. 14, pp. 181–189, 2025.
- [2] A. Peryanto, D. Susanto, and Y. F. Widodo, "KLASIFIKASI CITRA BUNGA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN GRAY LEVEL CO-OCCURRENCE MATRIX," vol. 9, no. 2, pp. 126–134, 2025.
- [3] P. Novantara, R. L. F, and M. Arismawati, "Deteksi Hama Penyakit Daun Padi Dengan Menggunakan Teknik Optimasi Deep Learning Convolutional Neural Network," vol. 7, no. 3, 2025.
- [4] D. Citra, D. Untuk, K. Penyakit, M. Pendekatan, D. Learning, and D. Model, "Jurnal Teknologi Terpadu," vol. 10, no. 1, pp. 56–62, 2024.
- [5] P. Kumar *et al.*, "Deep feature based rice leaf disease identi fi cation using support vector machine," *Comput. Electron. Agric.*, vol. 175, no. May, p. 105527, 2020.
- [6] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1–6.
- [7] J. Cervantes, F. Garcia-lamont, L. Rodríguez-mazahua, and A. Lopez, "Neurocomputing A comprehensive survey on support vector machine classification : Applications , challenges and trends," vol. 408, pp. 189–215, 2020.
- [8] *Penerapan Konsep Machine Learning & Deep Learning*. .
- [9] D. Anggriandi, E. Utami, and D. Ariatmanto, "Comparative Analysis of CNN and CNN-SVM Methods For Classification Types of Human Skin Disease," vol. 7, no. 4, pp. 2168–2178, 2023.
- [10] P. Tanaman, K. Sawit, I. Riati, and G. W. Nurcahyo, "Jurnal KomtekInfo Penerapan Convolutional Neural Network untuk Mengidentifikasi," vol. 11, no. 4, pp. 237–246, 2024.
- [11] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-paz, "mixup : B," pp. 1–13, 2018.
- [12] P. I. Ritharson, K. Raimond, X. A. Mary, J. Eunice, and J. Andrew, "Arti fi cial Intelligence in Agriculture DeepRice : A deep learning and deep feature based classi fi cation of Rice leaf disease subtypes," *Artif. Intell. Agric.*, vol. 11, pp. 34–49, 2024.
- [13] K. Saddami, Y. Nurdin, M. Zahramita, and M. Shahreeza, "Advancing Green AI: Efficient and Accurate Lightweight CNNs for Rice Leaf Disease Identification."
- [14] P. Chakraborty, M. A. Yousuf, and S. Islam, "Analysis of Texture Feature Extraction," no. July, 2022.
- [15] J. Wang, F. Shahzad, Z. Ahmad, M. Abdullah, and N. M. Hassan, "Trust and Consumers' Purchase Intention in a Social Commerce Platform: A Meta-Analytic Approach," *SAGE Open*, vol. 12, no. 2, 2022.
- [16] S. Ioffe and C. Szegedy, "Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift."
- [17] M. A. Nanda and A. Maddu, "A Comparison Study of Kernel Functions in the Support Vector Machine and Its Application for Termite Detection."
- [18] R. H. Prince, A. Al Mamun, and H. I. Peyal, "CSXAI : a lightweight 2D CNN- SVM model for detection and classi fi cation of various crop diseases with explainable AI visualization," no. July, pp. 1–18, 2024.
- [19] D. A. Anggoro and D. Permatasari, "Performance Comparison of the Kernels of Support Vector Machine Algorithm for Diabetes Mellitus Classification," vol. 14, no. 1, pp. 580–585, 2023.