

## BAB II

### LANDASAN TEORI

#### 2.1 Penelitian Terkait

Penelitian dalam bidang neural network untuk analisis pangan di Indonesia telah mulai berkembang dalam beberapa tahun terakhir. Penelitian tentang "Pendeteksi Visual Makanan dan Jumlah Kalorinya Menggunakan Algoritma Mask R-CNN Berbasis Bot Telegram" yang dipublikasikan dalam JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika) (Ash-shafa & Andono, 2025). Penelitian ini menggunakan metode Mask R-CNN (Region-based Convolutional Neural Network) untuk mendeteksi makanan secara visual dan menghitung kandungan kalorinya melalui platform Bot Telegram. Kelebihan dari penelitian ini adalah penggunaan teknologi *deep learning* yang canggih dengan *interface* yang mudah diakses melalui aplikasi populer seperti Telegram, sehingga memudahkan pengguna dalam menganalisis kandungan kalori makanan secara *real-time*. Namun demikian, penelitian ini memiliki keterbatasan karena hanya fokus pada prediksi kalori saja dan belum mencakup spektrum nutrisi yang lebih luas seperti protein, karbohidrat, lemak, vitamin, dan mineral. Hasil penelitian menunjukkan bahwa sistem mampu mendeteksi berbagai jenis makanan dan memberikan estimasi kalori dengan tingkat akurasi yang cukup baik untuk aplikasi praktis sehari-hari.

Dalam konteks yang lebih luas, penelitian tentang aplikasi neural network untuk analisis pangan juga didukung oleh penelitian Aditama tahun 2021 dari Universitas Jenderal Soedirman dengan judul "Klasifikasi Kualitas

Beras Dengan Citra Digital Menggunakan Metode Convolutional Neural Network (CNN) Berbasis Android". Meskipun fokus penelitian ini adalah klasifikasi kualitas beras, namun metodologi yang digunakan relevan dengan analisis pangan menggunakan neural network. Penelitian ini menggunakan CNN untuk mengklasifikasikan kualitas beras berdasarkan citra digital yang diambil melalui smartphone Android. Kelebihan penelitian ini adalah implementasi yang praktis pada platform mobile yang mudah diakses oleh masyarakat umum, serta penggunaan teknologi CNN yang terbukti efektif untuk analisis citra. Kelemahannya adalah terbatas pada satu jenis produk pangan (beras) dan belum mengintegrasikan analisis kandungan nutrisi. Hasil penelitian menunjukkan bahwa sistem dapat mengklasifikasikan kualitas beras dengan akurasi yang memadai untuk aplikasi komersial (Aditama, 2021).

Penelitian dalam bidang computer vision untuk quality assessment makanan di Indonesia telah dikembangkan oleh beberapa peneliti dengan pendekatan yang beragam. Penelitian dari Telkom University oleh tim peneliti (2020) dengan judul "Deteksi Makanan dengan Pengolahan Citra Menggunakan Metode Ekstraksi Ciri Histogram of Oriented Gradient (HOG) dan Klasifikasi K-Nearest Neighbors (K-NN) Berbasis Desktop" menggunakan metode ekstraksi ciri HOG untuk mendeteksi makanan secara otomatis pada sistem kasir restoran. Metode ini menggunakan pendekatan computer vision dengan ekstraksi fitur HOG yang kemudian diklasifikasikan menggunakan algoritma K-NN. Kelebihan penelitian ini adalah penerapan praktis untuk sistem otomatisasi restoran yang dapat meningkatkan efisiensi

pelayanan dan mengurangi kesalahan human error dalam proses identifikasi makanan. Pendekatan HOG terbukti efektif untuk mendeteksi bentuk dan tekstur makanan, sedangkan K-NN memberikan klasifikasi yang cepat dan akurat. Namun, kelemahan penelitian ini adalah keterbatasan pada identifikasi jenis makanan tanpa analisis kualitas atau kandungan nutrisi, serta implementasi yang masih berbasis desktop sehingga kurang fleksibel untuk penggunaan mobile. Hasil penelitian menunjukkan bahwa sistem mampu mendeteksi berbagai jenis makanan dengan tingkat akurasi yang baik untuk implementasi pada sistem point-of-sale restoran (Tim Peneliti Telkom University, 2020).

Penelitian lain yang relevan adalah tentang "Deteksi Kualitas Kemurnian Susu Sapi Melalui Pengolahan Citra Digital Menggunakan Metode Active Contour dengan Klasifikasi K-Nearest Neighbor". Penelitian ini menggunakan metode Active Contour untuk segmentasi citra susu sapi dan K-Nearest Neighbor untuk klasifikasi kualitas kemurnian. Pendekatan ini menggunakan analisis citra digital untuk mendeteksi adanya campuran atau adulteran dalam susu sapi yang dapat mempengaruhi kualitas dan nilai nutrisinya. Kelebihan penelitian ini adalah penerapan teknologi computer vision untuk quality control produk pangan yang sangat penting untuk kesehatan konsumen, serta penggunaan metode Active Contour yang efektif untuk segmentasi objek dengan bentuk yang tidak beraturan. Kelemahannya adalah terbatas pada satu jenis produk (susu sapi) dan belum mengintegrasikan analisis kandungan nutrisi secara komprehensif. Hasil penelitian menunjukkan

bahwa sistem dapat mendeteksi tingkat kemurnian susu dengan akurasi yang cukup tinggi untuk aplikasi quality control di industri dairy (Tim Peneliti Telkom University, 2021).

Penelitian tentang ekstraksi ciri warna untuk analisis kualitas pangan di Indonesia telah dikembangkan dalam berbagai konteks aplikasi. Jurnal Pangan Indonesia menerbitkan artikel tentang "Aplikasi Teknik Pengolahan Citra dalam Analisis Non-Destruktif Produk Pangan" yang memberikan tinjauan komprehensif tentang penggunaan teknologi pengolahan citra untuk analisis produk pangan tanpa merusak sampel. Penelitian ini menggunakan metode ekstraksi fitur warna, tekstur, dan bentuk untuk menganalisis kualitas berbagai produk pangan seperti buah-buahan, sayuran, dan biji-bijian. Kelebihan pendekatan ini adalah sifatnya yang non-destruktif sehingga sampel tidak rusak dan dapat langsung dikonsumsi atau dijual, efisiensi waktu dan biaya dibandingkan analisis laboratorium konvensional, serta kemampuan untuk memberikan hasil analisis secara real-time. Namun, kelemahan metode ini adalah sensitivitas terhadap kondisi pencahayaan dan lingkungan pengambilan gambar, serta keterbatasan dalam menganalisis kandungan nutrisi internal yang tidak tampak secara visual. Hasil penelitian menunjukkan bahwa analisis non-destruktif menggunakan pengolahan citra dapat menjadi alternatif yang efektif untuk quality control produk pangan dalam skala industri (Jurnal Pangan Indonesia, 2020).

Penelitian dari Universitas Tarumanagara yang dipublikasikan dalam Jurnal Ilmu Komputer dan Sistem Informasi (JIKSI) tentang "Klasifikasi Citra

Ikan Menggunakan Convolutional Neural Network" juga memberikan kontribusi dalam bidang ekstraksi ciri warna untuk analisis produk pangan. Meskipun fokus pada ikan sebagai produk perikanan, metodologi yang digunakan dapat diadaptasi untuk analisis produk pangan lainnya. Penelitian ini menggunakan CNN untuk mengekstrak fitur warna, tekstur, dan bentuk ikan untuk klasifikasi jenis dan kualitas. Kelebihan penelitian ini adalah penggunaan deep learning yang dapat mengekstrak fitur kompleks secara otomatis, kemampuan untuk menangani variasi dalam pencahayaan dan sudut pengambilan gambar, serta akurasi klasifikasi yang tinggi. Kelemahannya adalah memerlukan dataset training yang besar dan beragam, serta computational cost yang tinggi untuk training model. Hasil penelitian menunjukkan bahwa CNN efektif untuk klasifikasi produk perikanan dengan akurasi yang memadai untuk aplikasi komersial (Tim Peneliti Universitas Tarumanagara, 2022).

Meskipun penelitian spesifik tentang analisis telur menggunakan citra digital masih terbatas di Indonesia, namun prinsip-prinsip yang dikembangkan dalam penelitian-penelitian sebelumnya dapat diadaptasi untuk analisis telur. Penelitian dari Universitas Lampung dalam buku ajar "Pengolahan Citra: Dasar dan Contoh Penerapannya" memberikan landasan teoritis yang kuat untuk aplikasi pengolahan citra dalam berbagai bidang termasuk analisis produk pangan. Pendekatan yang dijelaskan meliputi preprocessing citra, ekstraksi fitur warna menggunakan berbagai color space (RGB, HSV, Lab), analisis tekstur menggunakan Gray Level Co-occurrence Matrix (GLCM), dan

klasifikasi menggunakan berbagai algoritma *machine learning*. Kelebihan pendekatan ini adalah fleksibilitas dalam penggunaan berbagai metode ekstraksi fitur sesuai dengan karakteristik produk yang dianalisis, kemampuan untuk menganalisis berbagai parameter kualitas secara simultan, serta implementasi yang relatif mudah dengan tools yang tersedia. Kelemahannya adalah memerlukan kalibrasi yang tepat untuk setiap jenis produk, sensitivitas terhadap variasi kondisi pengambilan gambar, serta keterbatasan dalam menganalisis kualitas internal produk. Hasil penelitian dan implementasi menunjukkan bahwa teknik pengolahan citra dapat diadaptasi untuk berbagai aplikasi analisis kualitas produk pangan termasuk telur.

## **2.2 Kandungan dan Protein dalam Kuning Telur**

### **2.2.1 Komposisi Nutrisi Kuning Telur**

#### **A. Struktur dan kandungan kuning telur**

Kuning telur merupakan bagian dari telur yang memiliki kandungan nutrisi yang sangat kompleks dan penting bagi kesehatan manusia. Menurut Kementerian Kesehatan RI (2021), kuning telur memiliki komposisi utama berupa protein, lemak, karbohidrat, vitamin, dan mineral. Komponen utama telur adalah kulit telur, putih telur (albumen) dan kuning telur. Komposisi ketiga komponen ini berbeda-beda tergantung pada jenis telur. Komposisi telur sebagian besar terdiri dari air, selain unsur-unsur penting lainnya (protein, lemak, karbohidrat, dan lain-lain) (RI, 2021).

Struktur kuning telur terdiri dari beberapa lapisan yang berbeda, yaitu vitellin membrane yang merupakan membran luar kuning telur, dan inti kuning telur yang mengandung berbagai nutrisi penting. Bagian kuning telur mengandung sekitar 32% dari total berat telur dan memiliki konsentrasi nutrisi yang lebih tinggi dibandingkan dengan putih telur (Indogen, 2021). Kandungan air dalam kuning telur berkisar antara 48-51%, sementara sisanya merupakan bahan kering yang terdiri dari protein, lemak, karbohidrat, vitamin, dan mineral.

#### B. Jenis-jenis protein dalam kuning telur

Protein dalam kuning telur memiliki karakteristik yang berbeda dengan protein pada putih telur. Berdasarkan penelitian yang dilakukan oleh Universitas Hasanuddin Makassar (2015), protein dalam putih telur terdiri dari : ovalbumin 54 %, conalbumin 13 %, ovomucoid 11 %, Lisozim 3,5 %, G2-globulin 4,0 %, G-3 Globulin 4,0 %, ovomucin 1,5 %, flavoprotein 0,8 %, ovoglikoprotein 0,5 %, ovoinhibitor 0,1 %. Namun, kuning telur memiliki komposisi protein yang berbeda dengan kandungan protein spesifik seperti phosvitin, lipovitellin, dan livetin.

Protein utama dalam kuning telur meliputi lipovitellin yang merupakan lipoprotein bermuatan tinggi dan berfungsi sebagai protein penyimpan, phosvitin yang merupakan protein dengan kandungan fosfor tinggi, serta livetin yang terdiri dari  $\alpha$ -livetin,  $\beta$ -

livetin, dan  $\gamma$ -livetin. Menurut Purwanto dan Marianti protein kuning telur memiliki karakteristik yang unik karena terikat dengan lemak dan fosfolipid, sehingga membentuk struktur lipoprotein kompleks yang memiliki nilai biologis tinggi (Purwanto, & Marianti, 2014).

#### C. Faktor yang mempengaruhi kandungan protein

Kandungan protein dalam kuning telur dipengaruhi oleh berbagai faktor, baik internal maupun eksternal. Faktor internal meliputi genetik ayam, umur ayam, dan kondisi fisiologis ayam. Ayam yang masih muda umumnya menghasilkan telur dengan kandungan protein yang lebih tinggi dibandingkan ayam yang sudah tua. Faktor genetik juga berperan penting dalam menentukan kualitas dan kuantitas protein yang dihasilkan.

Faktor eksternal yang mempengaruhi kandungan protein kuning telur antara lain pakan yang diberikan kepada ayam, kondisi lingkungan, dan sistem pemeliharaan. Pakan yang mengandung protein berkualitas tinggi akan menghasilkan telur dengan kandungan protein yang optimal. Kondisi lingkungan seperti suhu, kelembaban, dan stress juga dapat mempengaruhi kualitas protein telur. Menurut De Heus, faktor nutrisi dalam pakan ayam sangat berpengaruh terhadap komposisi nutrisi telur yang dihasilkan, termasuk kandungan protein di dalamnya (De Heus, 2024).

## 2.3 Metode Konvensional Analisis Protein

### A. Metode Kjeldahl

Metode Kjeldahl merupakan metode standar yang paling umum digunakan untuk analisis protein dalam berbagai sampel makanan, termasuk kuning telur. Metoda Kjeldahl dan metode Lowry. Mampu menetapkan kadar protein dari sampel berdasarkan metoda Kjeldahl dan metode Lowry. Prinsip dasar metode ini adalah penentuan kadar nitrogen total dalam sampel, yang kemudian dikonversi menjadi kadar protein dengan menggunakan faktor konversi tertentu.

Analisa protein cara Kjeldahl pada dasarnya dapat dibagi menjadi tiga tahapan yaitu proses destruksi, proses destilasi dan tahap titrasi. Tahap destruksi melibatkan pemanasan sampel dengan asam sulfat pekat dan katalisator untuk mengubah nitrogen organik menjadi amonium sulfat. Tahap destilasi mengubah amonium sulfat menjadi amonia yang kemudian ditangkap dalam larutan asam borat. Tahap titrasi melakukan titrasi amonia yang terlarut dengan larutan asam standar untuk menentukan kadar nitrogen total.

Metode Kjeldahl memiliki keunggulan dalam hal akurasi dan presisi yang tinggi, serta telah menjadi metode standar internasional untuk analisis protein. Namun, metode ini memiliki kelemahan dalam hal waktu analisis yang lama (6-8 jam), penggunaan bahan kimia berbahaya seperti asam sulfat pekat, dan membutuhkan peralatan khusus serta Metode Kjeldahl merupakan metode standar yang paling umum digunakan untuk

analisis protein dalam berbagai sampel makanan, termasuk kuning telur. Metoda Kjeldahl dan metode Lowry. Mampu menetapkan kadar protein dari sampel berdasarkan metoda Kjeldahl dan metode Lowry. Prinsip dasar metode ini adalah penentuan kadar nitrogen total dalam sampel, yang kemudian dikonversi menjadi kadar protein dengan menggunakan faktor konversi tertentu.

#### B. Metode Biuret

Metode Biuret merupakan metode spektrofotometri yang didasarkan pada reaksi antara protein dengan ion tembaga dalam suasana basa. Metode ini mengukur intensitas warna ungu yang terbentuk akibat reaksi antara ikatan peptida dalam protein dengan ion  $\text{Cu}^{2+}$  dalam reagent biuret. Intensitas warna yang terbentuk sebanding dengan konsentrasi protein dalam sampel.

Keunggulan metode Biuret adalah prosedur yang relatif sederhana, waktu analisis yang cepat (30-60 menit), dan tidak memerlukan destruksi sampel. Metode ini juga memberikan hasil yang cukup akurat untuk sampel dengan kandungan protein yang tinggi. Namun, metode ini memiliki kelemahan dalam hal sensitivitas yang rendah, sehingga tidak cocok untuk sampel dengan kandungan protein rendah. Selain itu, metode ini dapat terganggu oleh keberadaan senyawa lain yang dapat bereaksi dengan reagent biuret.

### C. Spektrofotometri

Metode spektrofotometri untuk analisis protein dapat dilakukan dengan berbagai pendekatan, termasuk pengukuran absorbansi pada panjang gelombang 280 nm yang merupakan panjang gelombang absorpsi maksimum protein. Perbandingan Analisa Kadar Protein Terlarut dengan berbagai Metode Spektroskopi UV-Visible menunjukkan bahwa metode spektrofotometri memiliki variasi yang cukup luas dalam aplikasinya. Metode spektrofotometri juga dapat menggunakan reagent khusus seperti dalam metode Lowry, Bradford, dan BCA (Bicinchoninic Acid). Masing-masing metode memiliki prinsip yang berbeda, dimana metode Lowry menggunakan kombinasi reaksi Biuret dengan reaksi reduksi reagent Folin-Ciocalteu, metode Bradford menggunakan pewarna Coomassie Brilliant Blue, dan metode BCA menggunakan reaksi reduksi ion tembaga oleh protein.

### D. Kelebihan dan kekurangan masing-masing metode

Setiap metode analisis protein memiliki karakteristik yang berbeda dalam hal akurasi, presisi, waktu analisis, dan biaya operasional. beberapa metode yang biasa digunakan dalam rangka penentuan konsentrasi preotein, yaitu metode Biuret, Lowry, dan lain sebagainya. Masing-masing metode mempunyai kekurangan dan kelebihan. Pemilihan metode yang terbaik dan tepat untuk suatu pengukuran bergantung pada beberapa faktor seperti misalnya, banyaknya material atau sampel yang tersedia, waktu

yang tersedia untuk melakukan pengukuran, alat spektrofotometri yang tersedia (VIS atau UV).

Metode Kjeldahl memiliki kelebihan dalam hal akurasi dan presisi yang tinggi, serta merupakan metode standar internasional. Namun, kekurangannya adalah waktu analisis yang lama, penggunaan bahan kimia berbahaya, dan biaya operasional yang tinggi. Metode ini ga bisa dipake pada produk fermentasi karena proteinnya sudah terhidrolisis. Kelemahan: senyawa lain dengan N juga bisa terukur.

Metode Biuret memiliki kelebihan dalam hal kesederhanaan prosedur dan waktu analisis yang cepat. Namun, kekurangannya adalah sensitivitas yang rendah dan dapat terganggu oleh senyawa pengganggu. Metode spektrofotometri memiliki kelebihan dalam hal kecepatan analisis dan kemudahan operasional, tetapi akurasinya dapat dipengaruhi oleh berbagai faktor seperti pH, suhu, dan keberadaan senyawa pengganggu lainnya.

## **2.4 Pengolahan Citra Digital**

### **2.4.1 Dasar-Dasar Pengolahan Citra**

#### **A. Definisi dan konsep citra digital**

Pengolahan citra digital merupakan bidang ilmu yang mempelajari tentang bagaimana suatu citra itu dibentuk, diolah, dan dianalisis sehingga menghasilkan informasi yang dapat dipahami oleh manusia. Menurut definisi yang lebih teknis, citra adalah sinyal dwimatra yang bersifat menerus (continue) yang dapat diamati oleh

sistem visual manusia. Secara matematis, citra adalah fungsi dwimatra yang menyatakan intensitas cahaya pada bidang dwimatra.

Citra digital dapat didefinisikan sebagai representasi visual dari dunia nyata dalam bentuk digital yang dapat dipahami dan diolah oleh komputer (ResearchGate, 2024). Berbeda dengan citra analog yang bersifat kontinu, citra digital terdiri dari elemen-elemen diskrit yang disebut piksel (picture element). Setiap piksel memiliki nilai intensitas tertentu yang merepresentasikan tingkat kecerahan atau warna pada posisi tertentu dalam citra.

Pengolahan citra digital memiliki tujuan utama untuk memperbaiki kualitas citra, mengekstrak informasi yang berguna dari citra, serta menganalisis dan menginterpretasi kandungan visual dalam citra. Dalam konteks penelitian ini, pengolahan citra digital digunakan untuk menganalisis karakteristik warna kuning telur yang berkorelasi dengan kandungan proteinnya.

#### B. Representasi citra dalam komputer

Representasi citra dalam komputer merupakan proses konversi informasi visual menjadi data numerik yang dapat diproses secara digital. Representasi gambar merupakan teknik yang dapat dianalisis dan dimanipulasi oleh komputer. Secara fundamental, citra digital direpresentasikan sebagai matriks dua dimensi yang terdiri dari baris dan kolom piksel.

Setiap piksel dalam citra memiliki koordinat (x,y) dan nilai intensitas tertentu. Untuk citra grayscale, setiap piksel memiliki satu nilai intensitas yang berkisar antara 0 (hitam) hingga 255 (putih) untuk citra 8-bit. Sedangkan untuk citra berwarna, setiap piksel memiliki beberapa nilai intensitas tergantung pada model warna yang digunakan. Misalnya, pada model warna RGB, setiap piksel memiliki tiga nilai intensitas yang masing-masing merepresentasikan komponen Red, Green, dan Blue.

Resolusi citra ditentukan oleh jumlah piksel dalam dimensi horizontal dan vertikal, yang biasanya dinyatakan dalam format lebar  $\times$  tinggi piksel. Semakin tinggi resolusi citra, semakin detail informasi yang dapat direpresentasikan, namun juga semakin besar ukuran file yang dihasilkan.

### C. Format file citra

Format file citra menentukan cara penyimpanan data citra dalam sistem komputer. Pemilihan format file yang tepat sangat penting dalam pengolahan citra digital karena mempengaruhi kualitas citra, ukuran file, dan kompatibilitas dengan berbagai aplikasi. Format file citra dapat dibagi menjadi dua kategori utama: format terkompresi dan format tidak terkompresi.

Format JPEG (Joint Photographic Experts Group) merupakan format terkompresi yang paling umum digunakan untuk citra fotografis. Format ini menggunakan kompresi lossy yang dapat

mengurangi ukuran file secara signifikan namun dengan sedikit kehilangan kualitas. Format PNG (Portable Network Graphics) menggunakan kompresi lossless yang mempertahankan kualitas citra asli namun dengan ukuran file yang lebih besar dibandingkan JPEG.

Format BMP (Bitmap) merupakan format tidak terkompresi yang menyimpan data citra secara mentah tanpa kompresi. Format ini memberikan kualitas citra terbaik namun menghasilkan ukuran file yang sangat besar. Format TIFF (Tagged Image File Format) merupakan format yang fleksibel dan dapat menyimpan citra dengan berbagai kedalaman bit serta mendukung berbagai skema kompresi.

#### **2.4.2 Model Warna**

##### **A. Model warna RGB (Red, Green, Blue)**

Model warna RGB merupakan sistem warna yang paling fundamental dalam pengolahan citra digital. Ruang warna RGB adalah sistem warna yang menggunakan tiga warna dasar: Merah (Red), Hijau (Green), dan Biru (Blue). Model ini berdasarkan pada teori bahwa semua warna dapat dibentuk melalui kombinasi tiga warna primer tersebut dengan intensitas yang berbeda-beda.

Setiap komponen warna dalam model RGB memiliki rentang nilai dari 0 hingga 255 untuk citra 8-bit, sehingga total kombinasi warna yang dapat direpresentasikan adalah  $256^3 = 16.777.216$  warna. Nilai (0,0,0) merepresentasikan warna hitam, sedangkan nilai

(255,255,255) merepresentasikan warna putih. Kombinasi yang sama dari ketiga komponen menghasilkan gradasi abu-abu.

Model RGB sangat sesuai untuk perangkat display seperti monitor dan televisi karena cara kerja perangkat tersebut yang memancarkan cahaya melalui fosfor merah, hijau, dan biru. Namun, model RGB memiliki keterbatasan dalam hal persepsi manusia terhadap warna, karena tidak semua perubahan nilai RGB menghasilkan perubahan visual yang seragam.

#### B. Model warna HSV (Hue, Saturation, Value)

Model warna HSV merupakan representasi warna yang lebih intuitif dibandingkan RGB karena lebih mendekati cara manusia mempersepsikan warna. HSV terdiri dari tiga komponen: Hue (warna dasar), Saturation (tingkat kejenuhan warna), dan Value (tingkat kecerahan). Model ini sering digunakan dalam aplikasi pengolahan citra yang memerlukan analisis warna yang lebih natural.

Komponen Hue direpresentasikan sebagai sudut dalam lingkaran warna dengan rentang  $0^\circ$  hingga  $360^\circ$ , dimana  $0^\circ$  merepresentasikan warna merah,  $120^\circ$  hijau, dan  $240^\circ$  biru. Saturation menunjukkan tingkat kemurnian warna dengan rentang 0 hingga 100%, dimana 0% menghasilkan warna abu-abu dan 100% menghasilkan warna murni. Value menunjukkan tingkat kecerahan dengan rentang 0 hingga 100%, dimana 0%

menghasilkan warna hitam dan 100% menghasilkan warna paling terang.

Keunggulan model HSV dalam pengolahan citra adalah kemampuannya untuk memisahkan informasi warna (Hue) dari informasi kecerahan (Value), sehingga lebih robust terhadap perubahan pencahayaan. Hal ini membuat model HSV sangat cocok untuk aplikasi deteksi objek berdasarkan warna, seperti dalam penelitian analisis kuning telur yang warnanya dapat bervariasi tergantung kondisi pencahayaan.

#### C. Model warna LAB

Model warna LAB (juga dikenal sebagai CIE Lab\*) merupakan model warna yang dikembangkan oleh Commission Internationale de l'Éclairage (CIE) sebagai standar internasional untuk representasi warna. Model ini terdiri dari tiga komponen:  $L^*$  (Lightness) yang merepresentasikan kecerahan,  $a^*$  yang merepresentasikan koordinat warna hijau-merah, dan  $b^*$  yang merepresentasikan koordinat warna biru-kuning.

Komponen  $L^*$  memiliki rentang nilai 0 hingga 100, dimana 0 merepresentasikan hitam sempurna dan 100 merepresentasikan putih sempurna. Komponen  $a^*$  dan  $b^*$  memiliki rentang teoritis -128 hingga +127, dimana nilai negatif  $a^*$  menunjukkan kecenderungan warna hijau dan nilai positif menunjukkan kecenderungan warna merah. Demikian pula, nilai negatif  $b^*$

menunjukkan kecenderungan warna biru dan nilai positif menunjukkan kecenderungan warna kuning.

Keunggulan utama model LAB adalah sifatnya yang device-independent dan perceptually uniform, artinya jarak euclidean antara dua titik dalam ruang warna LAB berkorelasi dengan perbedaan warna yang dipersepsikan oleh mata manusia. Hal ini membuat model LAB sangat berguna untuk aplikasi yang memerlukan pengukuran perbedaan warna yang akurat, seperti dalam quality control industri makanan.

#### D. Konversi antar model warna

Konversi antar model warna merupakan proses penting dalam pengolahan citra digital yang memungkinkan representasi citra dalam berbagai ruang warna sesuai kebutuhan aplikasi. Setiap model warna memiliki karakteristik dan keunggulan masing-masing, sehingga konversi diperlukan untuk mengoptimalkan proses analisis citra.

Konversi dari RGB ke HSV melibatkan perhitungan matematis yang mengubah representasi kartesian RGB menjadi representasi silindris HSV. Proses ini melibatkan penentuan nilai maksimum dan minimum dari komponen RGB, perhitungan selisih, dan aplikasi formula trigonometri untuk menentukan nilai Hue. Konversi ini sangat berguna dalam aplikasi yang memerlukan analisis warna yang robust terhadap perubahan pencahayaan.

Konversi dari RGB ke LAB merupakan proses yang lebih kompleks karena melibatkan konversi intermediate melalui ruang warna XYZ. Proses ini memerlukan koreksi gamma dan normalisasi terhadap white point standar. Meskipun lebih kompleks, konversi ke LAB memberikan keuntungan dalam hal akurasi pengukuran perbedaan warna dan konsistensi persepsi visual.

### **2.4.3 Ekstraksi Ciri Warna**

#### **A. Histogram warna**

Histogram warna merupakan representasi statistik distribusi intensitas warna dalam citra yang menunjukkan frekuensi kemunculan setiap nilai intensitas warna. Histogram memberikan informasi global tentang karakteristik warna citra dan sering digunakan sebagai ciri dasar dalam analisis citra. Untuk citra RGB, histogram dapat dibuat untuk setiap kanal warna (R, G, B) secara terpisah atau dikombinasikan.

Dalam konteks analisis kuning telur, histogram warna dapat memberikan informasi tentang distribusi warna kuning yang berkorelasi dengan kandungan nutrisi. Pola histogram yang berbeda dapat mengindikasikan perbedaan kualitas atau kandungan protein dalam kuning telur. Histogram yang memiliki puncak pada rentang warna kuning terang umumnya mengindikasikan kandungan nutrisi yang lebih tinggi.

Keunggulan histogram warna adalah kemudahan perhitungan dan interpretasinya, serta ketahanannya terhadap perubahan orientasi dan skala objek. Namun, histogram memiliki keterbatasan karena tidak menyimpan informasi spasial tentang distribusi warna dalam citra, sehingga dua citra dengan histogram yang sama belum tentu memiliki distribusi warna spasial yang sama.

#### B. Momen warna

Momen warna merupakan deskriptor statistik yang mengkarakterisasi distribusi warna dalam citra menggunakan momen statistik orde rendah. Momen warna terdiri dari mean (momen orde pertama), variance (momen orde kedua), skewness (momen orde ketiga), dan kurtosis (momen orde keempat). Setiap momen memberikan informasi yang berbeda tentang karakteristik distribusi warna.

Mean warna merepresentasikan nilai rata-rata intensitas warna dan memberikan informasi tentang kecenderungan warna dominan dalam citra. Variance menunjukkan tingkat penyebaran atau variabilitas intensitas warna terhadap nilai rata-rata. Skewness mengukur tingkat asimetri distribusi warna, dimana nilai positif menunjukkan ekor distribusi yang lebih panjang ke arah nilai intensitas tinggi. Kurtosis mengukur tingkat ketajaman atau kepilatan distribusi warna dibandingkan dengan distribusi normal.

Dalam penelitian analisis kuning telur, momen warna dapat memberikan karakterisasi yang lebih detail tentang pola distribusi warna dibandingkan histogram. Ciri statistik yang digunakan adalah nilai mean, max, dan min dari intensitas warna citra RGB dan HSV (Pemrograman Matlab, 2019). Kombinasi berbagai momen warna dapat memberikan signature yang unik untuk setiap sampel kuning telur dengan kandungan protein yang berbeda.

#### C. Statistik warna (mean, variance, skewness, kurtosis)

Statistik warna merupakan parameter numerik yang menggambarkan karakteristik distribusi warna dalam citra secara kuantitatif. Parameter-parameter ini dihitung untuk setiap kanal warna dan memberikan informasi yang komprehensif tentang sifat statistik warna dalam citra.

Mean ( $\mu$ ) dihitung sebagai rata-rata aritmatika dari semua nilai piksel dalam kanal warna tertentu, memberikan informasi tentang kecenderungan sentral distribusi warna. Variance ( $\sigma^2$ ) menunjukkan tingkat penyebaran nilai piksel terhadap mean, dimana variance tinggi mengindikasikan variabilitas warna yang besar dalam citra. Standard deviation ( $\sigma$ ) merupakan akar kuadrat dari variance dan memberikan ukuran penyebaran dalam satuan yang sama dengan data asli.

Skewness mengukur tingkat asimetri distribusi dengan formula yang melibatkan momen ketiga terhadap mean. Nilai

skewness positif menunjukkan distribusi dengan ekor yang lebih panjang ke arah nilai yang lebih tinggi, sedangkan nilai negatif menunjukkan sebaliknya. Kurtosis mengukur tingkat kepilatan distribusi dengan membandingkan konsentrasi data di sekitar mean dengan distribusi normal. Nilai kurtosis tinggi menunjukkan distribusi yang lebih runcing dengan ekor yang lebih tebal.

#### D. Teknik normalisasi ciri

Normalisasi ciri merupakan tahap preprocessing yang penting dalam ekstraksi ciri warna untuk memastikan bahwa semua ciri memiliki skala yang sebanding dan tidak ada ciri yang mendominasi proses pembelajaran atau klasifikasi karena perbedaan skala. Teknik normalisasi yang umum digunakan meliputi min-max normalization, z-score normalization, dan robust normalization.

Min-max normalization mengubah rentang nilai ciri menjadi interval  $[0,1]$  atau  $[-1,1]$  dengan formula linear yang melibatkan nilai minimum dan maksimum dari dataset. Teknik ini mempertahankan distribusi asli data namun dapat sensitif terhadap outlier. Z-score normalization (standardization) mengubah distribusi data menjadi distribusi dengan mean 0 dan standard deviation 1, sehingga lebih robust terhadap outlier dibandingkan min-max normalization.

Robust normalization menggunakan median dan interquartile range (IQR) sebagai parameter normalisasi, sehingga lebih tahan terhadap outlier ekstrem. Pemilihan teknik normalisasi yang tepat tergantung pada karakteristik data dan algoritma yang akan digunakan. Dalam konteks analisis kuning telur menggunakan neural network, normalisasi z-score umumnya memberikan performa yang baik karena membantu mempercepat konvergensi algoritma pembelajaran.

## **2.5 Neural Network dan Backpropagation**

### **2.5.1 Konsep Dasar Neural Network**

#### **A. Inspirasi dari sistem saraf biologis**

Neural Network atau Jaringan Syaraf Tiruan (JST) merupakan model komputasi yang terinspirasi dari cara kerja sistem syaraf biologis pada otak manusia. Konsep dasar neural network dikembangkan berdasarkan pemahaman tentang bagaimana neuron-neuron dalam otak saling terhubung dan berkomunikasi untuk memproses informasi. Neural network adalah sistem pemrosesan informasi yang memiliki karakteristik mirip dengan jaringan syaraf biologis, dimana informasi diproses melalui elemen-elemen sederhana yang disebut neuron buatan.

Sistem saraf biologis terdiri dari miliaran neuron yang saling terhubung melalui sinaps, membentuk jaringan kompleks yang mampu melakukan pembelajaran, pengenalan pola, dan

pengambilan keputusan. Setiap neuron menerima sinyal input melalui dendrit, memproses informasi di badan sel, dan mengirimkan output melalui akson. Proses ini melibatkan mekanisme aktivasi yang kompleks, dimana neuron akan "firing" ketika jumlah input yang diterima melampaui ambang batas tertentu.

Konsep biologis ini kemudian diadaptasi menjadi model matematis dalam neural network, dimana setiap neuron buatan menerima input dari neuron lain, mengolahnya menggunakan fungsi aktivasi, dan menghasilkan output yang diteruskan ke neuron berikutnya. Kemampuan pembelajaran neural network berasal dari penyesuaian kekuatan koneksi antar neuron, yang dalam konteks biologis setara dengan perubahan kekuatan sinaps.

#### B. Struktur neuron buatan

Neuron buatan merupakan unit dasar dalam neural network yang mensimulasikan fungsi neuron biologis dalam bentuk model matematis. Struktur neuron buatan terdiri dari beberapa komponen utama: input, bobot (weight), bias, fungsi penjumlahan (summation function), dan fungsi aktivasi. Setiap neuron menerima sejumlah input  $x_1, x_2, \dots, x_n$  yang masing-masing dikalikan dengan bobot  $w_1, w_2, \dots, w_n$ .

Proses perhitungan dalam neuron buatan dimulai dengan menghitung jumlah tertimbang (weighted sum) dari semua input

yang diterima, ditambah dengan nilai bias. Secara matematis, hal ini dapat ditulis sebagai  $net = \sum(x_i \times w_i) + b$ , dimana  $net$  adalah nilai input total,  $x_i$  adalah input ke- $i$ ,  $w_i$  adalah bobot untuk input ke- $i$ , dan  $b$  adalah bias. Hasil perhitungan ini kemudian diproses melalui fungsi aktivasi untuk menghasilkan output akhir neuron.

Bobot dalam neuron buatan memiliki peran yang sangat penting karena menentukan kekuatan pengaruh setiap input terhadap output neuron. Bobot positif memberikan pengaruh yang menguatkan, sedangkan bobot negatif memberikan pengaruh yang melemahkan. Bias berfungsi sebagai parameter tambahan yang memungkinkan neuron untuk mengaktivasi output bahkan ketika semua input bernilai nol, sehingga meningkatkan fleksibilitas model.

#### C. Fungsi aktivasi (sigmoid, ReLU, tanh)

Fungsi aktivasi merupakan komponen penting dalam neural network yang menentukan apakah suatu neuron akan aktif atau tidak berdasarkan input yang diterima. Fungsi aktivasi juga berperan dalam memberikan sifat non-linear pada neural network, yang memungkinkan model untuk mempelajari pola yang kompleks. Pemilihan fungsi aktivasi yang tepat sangat mempengaruhi performa dan kemampuan pembelajaran neural network.

Fungsi aktivasi sigmoid adalah salah satu fungsi yang paling sering digunakan dalam neural network klasik. Fungsi ini memiliki

bentuk S (sigmoid) dengan formula  $f(x) = 1/(1 + e^{(-x)})$ , yang menghasilkan output dalam rentang 0 hingga 1. Keunggulan fungsi sigmoid adalah outputnya yang dapat diinterpretasikan sebagai probabilitas, serta sifatnya yang smooth dan differentiable di semua titik. Namun, fungsi sigmoid memiliki kelemahan berupa vanishing gradient problem, dimana gradien menjadi sangat kecil untuk nilai input yang ekstrem.

Fungsi aktivasi ReLU (Rectified Linear Unit) telah menjadi pilihan populer dalam neural network modern karena kesederhanaannya dan efisiensi komputasinya. ReLU didefinisikan sebagai  $f(x) = \max(0, x)$ , yang menghasilkan output 0 untuk input negatif dan output yang sama dengan input untuk input positif. Keunggulan ReLU adalah kemampuannya mengatasi vanishing gradient problem dan mempercepat proses pembelajaran. Namun, ReLU memiliki kelemahan berupa dying ReLU problem, dimana neuron dapat menjadi tidak aktif secara permanen.

Fungsi aktivasi tanh (hyperbolic tangent) merupakan versi yang diperbaiki dari fungsi sigmoid dengan formula  $f(x) = (e^x - e^{(-x)})/(e^x + e^{(-x)})$ . Fungsi tanh menghasilkan output dalam rentang -1 hingga 1, yang membuatnya zero-centered dan memiliki gradien yang lebih kuat dibandingkan sigmoid. Fungsi tanh sering digunakan dalam hidden layer karena sifatnya yang zero-centered dapat membantu mempercepat konvergensi.

#### D. Arsitektur neural network (input layer, hidden layer, output layer)

Arsitektur neural network merujuk pada susunan dan organisasi neuron-neuron dalam jaringan, yang terdiri dari beberapa lapisan (layer) yang saling terhubung. Arsitektur dasar neural network terdiri dari tiga jenis lapisan utama: input layer, hidden layer, dan output layer. Setiap lapisan memiliki fungsi dan karakteristik yang berbeda dalam memproses informasi.

Input layer merupakan lapisan pertama yang menerima data input dari lingkungan eksternal. Jumlah neuron dalam input layer ditentukan oleh dimensi data input yang akan diproses. Dalam konteks penelitian analisis protein kuning telur, input layer akan menerima ciri-ciri warna yang telah diekstrak dari citra kuning telur. Input layer tidak melakukan pemrosesan data, melainkan hanya meneruskan data ke lapisan berikutnya.

Hidden layer merupakan lapisan yang terletak antara input layer dan output layer, dimana pemrosesan informasi yang sebenarnya terjadi. Hidden layer dapat terdiri dari satu atau lebih lapisan, tergantung pada kompleksitas masalah yang diselesaikan. Setiap neuron dalam hidden layer menerima input dari semua neuron di lapisan sebelumnya, memproses informasi menggunakan fungsi aktivasi, dan meneruskan output ke lapisan berikutnya. Jumlah neuron dan hidden layer mempengaruhi kapasitas pembelajaran dan kompleksitas model.

Output layer merupakan lapisan terakhir yang menghasilkan output akhir dari neural network. Jumlah neuron dalam output layer ditentukan oleh jenis masalah yang diselesaikan. Untuk masalah regresi seperti prediksi kandungan protein, output layer biasanya terdiri dari satu neuron yang menghasilkan nilai kontinyu. Untuk masalah klasifikasi, jumlah neuron output sesuai dengan jumlah kelas yang akan diprediksi.

## **2.5.2 Algoritma Backpropagation**

### **A. Prinsip kerja backpropagation**

Algoritma backpropagation merupakan metode pembelajaran supervised yang paling umum digunakan untuk melatih neural network. Prinsip kerja backpropagation didasarkan pada konsep meminimalkan error atau kesalahan prediksi melalui penyesuaian bobot secara iteratif. Nama "backpropagation" berasal dari proses propagasi error yang dilakukan secara mundur (backward) dari output layer menuju input layer.

Proses pembelajaran backpropagation terdiri dari dua fase utama: forward propagation dan backward propagation. Pada fase forward propagation, data input diproses melalui jaringan dari input layer hingga output layer untuk menghasilkan prediksi. Pada fase backward propagation, error antara prediksi dan target yang sebenarnya dihitung dan dipropagasi mundur untuk mengupdate bobot-bobot dalam jaringan.

Algoritma backpropagation menggunakan konsep gradient descent untuk mencari nilai bobot yang optimal. Gradient descent adalah teknik optimasi yang bergerak ke arah negatif dari gradien fungsi error untuk mencari titik minimum. Proses ini dilakukan secara iteratif hingga error mencapai nilai yang dapat diterima atau hingga mencapai jumlah iterasi maksimum yang telah ditentukan.

#### B. Forward propagation

Forward propagation adalah proses perhitungan output neural network berdasarkan input yang diberikan dan bobot yang ada saat ini. Proses ini dimulai dari input layer dan bergerak maju melalui hidden layer hingga mencapai output layer. Pada setiap lapisan, input dari lapisan sebelumnya dikalikan dengan bobot yang sesuai, dijumlahkan, ditambah bias, dan diproses melalui fungsi aktivasi.

Secara matematis, proses forward propagation dapat dinyatakan sebagai serangkaian operasi matriks. Untuk lapisan ke- $l$ , output  $h^{(l)}$  dihitung sebagai  $h^{(l)} = f(W^{(l)} \times h^{(l-1)} + b^{(l)})$ , dimana  $W^{(l)}$  adalah matriks bobot lapisan ke- $l$ ,  $h^{(l-1)}$  adalah output dari lapisan sebelumnya,  $b^{(l)}$  adalah vektor bias, dan  $f$  adalah fungsi aktivasi. Proses ini diulang untuk setiap lapisan hingga mencapai output layer.

Output dari forward propagation kemudian dibandingkan dengan target yang sebenarnya untuk menghitung error atau loss function. Error function yang umum digunakan adalah Mean

Squared Error (MSE) untuk masalah regresi dan Cross-Entropy untuk masalah klasifikasi. Nilai error ini menjadi dasar untuk proses backward propagation dan update bobot.

### C. Backward propagation

Backward propagation adalah proses perhitungan gradien error terhadap setiap bobot dalam neural network, dimulai dari output layer dan bergerak mundur hingga input layer. Tujuan utama backward propagation adalah menentukan seberapa besar kontribusi setiap bobot terhadap total error, sehingga bobot dapat diupdate dengan arah yang tepat untuk mengurangi error.

Proses backward propagation menggunakan chain rule dari kalkulus untuk menghitung gradien error terhadap setiap bobot. Chain rule memungkinkan perhitungan gradien fungsi komposit dengan mengalikan gradien-gradien fungsi yang lebih sederhana. Untuk neural network, chain rule diterapkan untuk menghitung gradien error terhadap bobot di setiap lapisan berdasarkan gradien error di lapisan berikutnya.

Gradien error terhadap bobot dihitung menggunakan formula  $\delta^{(l)} = f'(\text{net}^{(l)}) \times \delta^{(l+1)} \times W^{(l+1)}$ , dimana  $\delta^{(l)}$  adalah gradien error di lapisan ke- $l$ ,  $f'(\text{net}^{(l)})$  adalah turunan fungsi aktivasi, dan  $W^{(l+1)}$  adalah bobot lapisan berikutnya. Proses ini memungkinkan propagasi error dari output layer kembali ke semua lapisan dalam jaringan.

#### D. Perhitungan gradient dan update bobot

Perhitungan gradient merupakan inti dari algoritma backpropagation yang menentukan arah dan besarnya perubahan bobot untuk mengurangi error. Gradient menunjukkan arah perubahan yang akan memberikan peningkatan terbesar pada fungsi error, sehingga untuk meminimalkan error, bobot diupdate ke arah negatif dari gradient.

Setelah gradient dihitung melalui backward propagation, bobot diupdate menggunakan formula  $W^{(l)}_{\text{new}} = W^{(l)}_{\text{old}} - \alpha \times \partial E / \partial W^{(l)}$ , dimana  $\alpha$  adalah learning rate dan  $\partial E / \partial W^{(l)}$  adalah gradient error terhadap bobot di lapisan ke- $l$ . Learning rate menentukan seberapa besar langkah yang diambil dalam setiap update, mempengaruhi kecepatan konvergensi dan stabilitas pembelajaran.

Proses perhitungan gradient dan update bobot dilakukan untuk semua bobot dalam jaringan secara bersamaan. Hal ini memastikan bahwa seluruh jaringan berkembang secara konsisten menuju solusi yang optimal. Proses ini diulang untuk setiap batch data training hingga error mencapai nilai yang dapat diterima atau hingga mencapai jumlah epoch maksimum.

#### E. Learning rate dan momentum

Learning rate merupakan parameter krusial dalam algoritma backpropagation yang menentukan seberapa cepat neural network

belajar. Learning rate yang terlalu besar dapat menyebabkan osilasi atau bahkan divergensi dalam proses pembelajaran, sedangkan learning rate yang terlalu kecil dapat membuat proses pembelajaran menjadi sangat lambat. Pemilihan learning rate yang tepat sangat penting untuk mencapai konvergensi yang optimal.

Learning rate yang optimal biasanya ditentukan melalui eksperimen dengan berbagai nilai, mulai dari 0.001 hingga 0.1. Beberapa teknik adaptive learning rate seperti AdaGrad, RMSprop, dan Adam telah dikembangkan untuk menyesuaikan learning rate secara otomatis selama proses pembelajaran. Teknik-teknik ini dapat meningkatkan stabilitas dan kecepatan konvergensi dibandingkan dengan learning rate yang tetap.

Momentum adalah teknik yang digunakan untuk mempercepat konvergensi dan mengurangi osilasi dalam proses pembelajaran. Momentum bekerja dengan menambahkan fraksi dari update bobot sebelumnya ke update bobot saat ini, sehingga memberikan "inersia" pada proses pembelajaran. Formula update bobot dengan momentum adalah:  $v_t = \gamma \times v_{(t-1)} + \alpha \times \nabla E$ , dan  $W_{\text{new}} = W_{\text{old}} - v_t$ , dimana  $\gamma$  adalah koefisien momentum (biasanya 0.9),  $v_t$  adalah velocity saat ini, dan  $\nabla E$  adalah gradient.

### 2.5.3 Parameter dan Optimasi Neural Network

#### A. Inisialisasi bobot

Inisialisasi bobot merupakan langkah awal yang sangat penting dalam pelatihan neural network karena mempengaruhi kecepatan konvergensi dan kemampuan jaringan untuk mencapai solusi optimal. Inisialisasi yang buruk dapat menyebabkan vanishing gradient, exploding gradient, atau konvergensi ke local minimum yang tidak optimal. Terdapat beberapa metode inisialisasi bobot yang umum digunakan dalam praktik.

Inisialisasi random uniform atau gaussian dengan nilai kecil (biasanya antara -0.1 hingga 0.1) merupakan pendekatan yang sederhana namun sering efektif. Metode Xavier initialization (juga dikenal sebagai Glorot initialization) menggunakan distribusi uniform dengan rentang yang disesuaikan berdasarkan jumlah input dan output neuron, yaitu  $U(-\sqrt{6/\sqrt{n_{in} + n_{out}}}, \sqrt{6/\sqrt{n_{in} + n_{out}}})$ , dimana  $n_{in}$  dan  $n_{out}$  adalah jumlah neuron input dan output.

He initialization merupakan metode yang dikembangkan khusus untuk fungsi aktivasi ReLU. Metode ini menggunakan distribusi normal dengan standar deviasi  $\sqrt{2/n_{in}}$ , dimana  $n_{in}$  adalah jumlah neuron input. He initialization terbukti efektif untuk neural network yang menggunakan ReLU sebagai fungsi aktivasi karena memperhitungkan karakteristik spesifik fungsi ReLU.

## B. Overfitting dan underfitting

Overfitting dan underfitting merupakan dua masalah fundamental dalam machine learning yang juga mempengaruhi performa neural network. Overfitting terjadi ketika model terlalu kompleks dan mempelajari detail spesifik dari data training, termasuk noise, sehingga performa pada data testing menjadi buruk. Sebaliknya, underfitting terjadi ketika model terlalu sederhana dan tidak mampu menangkap pola yang mendasari dalam data.

Overfitting dapat diidentifikasi melalui monitoring training loss dan validation loss selama proses pelatihan. Jika training loss terus menurun sementara validation loss mulai meningkat, ini menandakan terjadinya overfitting. Fenomena ini sering terjadi pada neural network dengan banyak parameter yang dilatih pada dataset yang relatif kecil atau ketika proses pelatihan berlangsung terlalu lama.

Underfitting dapat diidentifikasi ketika baik training loss maupun validation loss tinggi dan tidak menunjukkan perbaikan yang signifikan selama pelatihan. Hal ini dapat disebabkan oleh arsitektur neural network yang terlalu sederhana, learning rate yang terlalu kecil, atau jumlah epoch yang tidak mencukupi. Solusi untuk underfitting meliputi penambahan kompleksitas model, penyesuaian hyperparameter, atau perpanjangan waktu pelatihan.

### C. Teknik regularisasi

Regularisasi merupakan teknik yang digunakan untuk mencegah overfitting dengan menambahkan penalty term pada loss function atau dengan memodifikasi arsitektur neural network. Tujuan utama regularisasi adalah mengurangi kompleksitas model agar dapat generalisasi dengan baik pada data baru. Terdapat beberapa teknik regularisasi yang umum digunakan dalam neural network.

L1 dan L2 regularization menambahkan penalty term pada loss function berdasarkan magnitude bobot. L1 regularization menambahkan penalty berupa jumlah absolute value dari bobot ( $\|W\|_1$ ), sedangkan L2 regularization menambahkan penalty berupa jumlah kuadrat dari bobot ( $\|W\|_2^2$ ). L2 regularization lebih umum digunakan karena menghasilkan solusi yang lebih stabil dan smooth.

Dropout merupakan teknik regularisasi yang bekerja dengan secara random "mematikan" sebagian neuron selama proses pelatihan. Setiap neuron memiliki probabilitas tertentu (biasanya 0.2-0.5) untuk tidak aktif dalam setiap forward pass. Teknik ini memaksa neural network untuk tidak terlalu bergantung pada neuron-neuron tertentu dan meningkatkan kemampuan generalisasi.

Early stopping adalah teknik regularisasi yang menghentikan proses pelatihan ketika validation loss mulai meningkat, meskipun training loss masih menurun. Teknik ini mencegah model untuk overfitting dengan menghentikan pelatihan pada titik optimal. Implementasi early stopping memerlukan monitoring validation loss dan penetapan patience parameter yang menentukan berapa epoch tanpa perbaikan sebelum pelatihan dihentikan.

#### D. Cross-validation

Cross-validation merupakan teknik evaluasi model yang membagi dataset menjadi beberapa fold untuk menguji performa model pada berbagai subset data. Teknik ini membantu menilai kemampuan generalisasi model dan mengurangi bias yang dapat terjadi karena pembagian data training dan testing yang tidak representatif. K-fold cross-validation adalah metode yang paling umum digunakan.

Dalam k-fold cross-validation, dataset dibagi menjadi k bagian yang sama besar. Model dilatih pada k-1 bagian dan diuji pada 1 bagian yang tersisa. Proses ini diulang k kali dengan bagian yang berbeda sebagai data testing, sehingga setiap bagian pernah menjadi data testing. Performa akhir model dihitung sebagai rata-rata dari k hasil evaluasi.

Stratified cross-validation adalah variasi dari k-fold cross-validation yang mempertahankan proporsi kelas dalam setiap fold.

Teknik ini penting untuk dataset yang tidak seimbang, dimana beberapa kelas memiliki jumlah sampel yang jauh lebih sedikit daripada kelas lainnya. Cross-validation sangat berguna untuk hyperparameter tuning dan model selection dalam neural network.

#### E. Evaluasi performa (MSE, MAE, $R^2$ )

Evaluasi performa neural network memerlukan metrik yang sesuai dengan jenis masalah yang diselesaikan. Untuk masalah regresi seperti prediksi kandungan protein kuning telur, terdapat beberapa metrik evaluasi yang umum digunakan: Mean Squared Error (MSE), Mean Absolute Error (MAE), dan coefficient of determination ( $R^2$ ).

Mean Squared Error (MSE) menghitung rata-rata kuadrat perbedaan antara nilai prediksi dan nilai aktual. Formula MSE adalah  $MSE = (1/n) \times \sum (y_{\text{actual}} - y_{\text{predicted}})^2$ , dimana  $n$  adalah jumlah sampel. MSE memberikan penalty yang lebih besar untuk error yang besar karena adanya operasi kuadrat, sehingga sensitif terhadap outlier. MSE sering digunakan sebagai loss function dalam pelatihan neural network untuk masalah regresi.

Mean Absolute Error (MAE) menghitung rata-rata absolute perbedaan antara nilai prediksi dan nilai aktual. Formula MAE adalah  $MAE = (1/n) \times \sum |y_{\text{actual}} - y_{\text{predicted}}|$ . MAE lebih robust terhadap outlier dibandingkan MSE karena tidak menggunakan

operasi kuadrat. MAE memberikan interpretasi yang lebih intuitif karena nilainya dalam satuan yang sama dengan variabel target.

Coefficient of determination ( $R^2$ ) mengukur proporsi variasi dalam variabel target yang dapat dijelaskan oleh model.  $R^2$  berkisar antara 0 hingga 1, dimana nilai 1 menunjukkan model sempurna dan nilai 0 menunjukkan model tidak lebih baik dari prediksi rata-rata. Formula  $R^2$  adalah  $R^2 = 1 - SS_{\text{res}}/SS_{\text{tot}}$ , dimana  $SS_{\text{res}}$  adalah sum of squares of residuals dan  $SS_{\text{tot}}$  adalah total sum of squares.  $R^2$  memberikan gambaran tentang seberapa baik model menjelaskan variabilitas data.