

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Penelitian oleh Wahyuningsih, Nugraha, dan DwiYansaputra (2024) dalam *Jurnal Teknik Informatika (Jutif)* Universitas Mataram mengkaji klasifikasi penyakit gigi berupa karies melalui citra digital dengan membandingkan arsitektur *Efficientnet-B0* dan beberapa arsitektur Convolutional Neural Network (CNN) lainnya. Penelitian ini menggunakan *dataset* awal sebanyak 1.554 citra yang kemudian diaugmentasi menjadi 6.348 citra, dengan pembagian data *training*, *validation*, dan *testing* menggunakan rasio 70:15:15. Hasil evaluasi menunjukkan bahwa *Efficientnet-B0* mampu menghasilkan akurasi yang lebih baik dibandingkan arsitektur CNN lainnya, sehingga dinilai efisien dalam mendeteksi penyakit gigi berbasis citra. Meskipun demikian, penelitian ini masih menggunakan arsitektur CNN tunggal yang berfokus pada ekstraksi fitur lokal, sehingga kemampuan *model* dalam menangkap hubungan global antarbagian citra penyakit gigi belum dieksplorasi lebih lanjut[5].

Penelitian lain yang dilakukan oleh Lavenia, Ramdani, dan Hoeronis berfokus pada klasifikasi penyakit pulpitis menggunakan citra radiografi periapikal dengan metode Convolutional Neural Network (CNN). *Dataset* yang digunakan terdiri dari 1.000 citra yang terbagi ke dalam dua kelas, yaitu pulpitis dan normal. Hasil penelitian menunjukkan bahwa pemilihan hyperparameter, khususnya jumlah *epoch* dan optimizer, memiliki pengaruh signifikan terhadap performa *model*. Dengan menggunakan optimizer RMSPROP dan 50 *epoch*, penelitian ini berhasil mencapai akurasi sebesar 98,75%. Temuan tersebut membuktikan bahwa CNN dapat menjadi pendekatan yang efektif dalam membantu diagnosis pulpitis secara otomatis melalui citra radiografi. Namun, penelitian ini masih terbatas pada penggunaan CNN murni dan citra radiografi, serta belum mempertimbangkan

pendekatan arsitektur lanjutan yang mampu *memodel* kan hubungan global pada citra medis secara lebih komprehensif[6].

Sementara itu, penelitian yang dilakukan oleh Andika Nur Pratama mengenai klasifikasi penyakit kulit menggunakan arsitektur *Efficientnet*-B2 juga menunjukkan hasil yang menjanjikan dalam bidang citra medis. Penelitian ini bertujuan untuk mengatasi keterbatasan diagnosis manual yang memerlukan waktu lama dan berpotensi menimbulkan kesalahan dengan memanfaatkan *deep learning*. Hasil evaluasi menunjukkan bahwa *Efficientnet*-B2 mampu mencapai akurasi sebesar 84,0%, precision 85,0%, recall 83,0%, dan F1-score 84,0%. Hasil tersebut menunjukkan bahwa *Efficientnet*-B2 memiliki kinerja yang cukup baik dalam klasifikasi citra medis. Namun, penelitian ini masih menggunakan *model* CNN tunggal dan belum mengombinasikannya dengan arsitektur lain maupun menerapkan teknik augmentasi lanjutan untuk meningkatkan kemampuan generalisasi *model* [7].

Berdasarkan penelitian-penelitian tersebut, dapat disimpulkan bahwa arsitektur Convolutional Neural Network, khususnya *Efficientnet*, telah banyak digunakan dan terbukti efektif dalam klasifikasi citra medis, termasuk pada kasus penyakit gigi dan mulut. CNN unggul dalam mengekstraksi fitur lokal seperti tekstur dan pola visual, namun memiliki keterbatasan dalam menangkap hubungan global antarbagian citra. Keterbatasan ini menjadi tantangan tersendiri pada citra penyakit gigi dan mulut yang memiliki variasi visual dan pola penyakit yang kompleks.

Seiring dengan perkembangan *deep learning*, *Vision Transformer* (ViT) diperkenalkan sebagai pendekatan yang mampu *memodel* kan hubungan global pada citra melalui mekanisme self-attention. *Vision Transformer* menunjukkan potensi yang menjanjikan dalam berbagai tugas klasifikasi citra medis. Namun, penerapan ViT pada *dataset* medis yang terbatas masih menghadapi tantangan, terutama terkait kebutuhan data pelatihan yang relatif besar agar *model* dapat mencapai performa optimal.

Selain pendekatan CNN, Penelitian berjudul “Pendeteksian Dini Stunting pada Balita Menggunakan Vision Transformer (ViT) Berbasis Citra Tubuh” yang

dilakukan oleh Reyvan Revolusioner Ar, Agusriyati, dan Sumiarni Moka membahas penerapan *Vision Transformer* untuk membantu proses skrining dini stunting berdasarkan citra tubuh balita. *Model Vision Transformer* dimanfaatkan untuk menangkap hubungan global antarbagian citra tubuh sehingga pola visual yang berkaitan dengan kondisi stunting dapat dianalisis secara menyeluruh. Hasil pengujian menunjukkan bahwa model ViT mampu mencapai akurasi sebesar 98%, yang menandakan bahwa pendekatan ini memiliki performa yang sangat baik dalam tugas pendeteksian dini stunting berbasis citra. Meskipun demikian, penelitian ini masih memiliki keterbatasan, terutama pada ketergantungan terhadap kualitas dan resolusi citra yang baik agar klasifikasi dapat dilakukan secara akurat [8].

Penelitian lain berjudul “Penerapan Computer Vision untuk Klasifikasi Penyakit Mata Menggunakan Arsitektur Vision Transformers pada Citra Fundus” yang disusun oleh Irvan Yudistiansyah mengkaji penggunaan *Vision Transformer* untuk klasifikasi penyakit mata berdasarkan citra fundus. Penelitian ini menunjukkan bahwa *Vision Transformer* mampu memodelkan struktur global dan pola visual kompleks pada citra fundus mata. Berdasarkan hasil pengujian, model yang dikembangkan mencapai akurasi sebesar 91,38%, yang menunjukkan efektivitas *Vision Transformer* dalam klasifikasi penyakit mata berbasis citra fundus. Namun demikian, penelitian ini juga memiliki keterbatasan, yaitu sensitivitas model terhadap variasi kualitas citra fundus serta kebutuhan data latih yang relatif besar, sehingga pendekatan ini masih berpotensi mengalami penurunan performa pada dataset yang terbatas atau tidak seimbang[9].

Untuk mengatasi keterbatasan masing-masing pendekatan, pengembangan *model hybrid* yang mengombinasikan *Convolutional Neural Network* dan *Vision Transformer* mulai banyak diteliti. Pendekatan *hybrid* bertujuan memanfaatkan keunggulan CNN dalam ekstraksi fitur lokal dan kemampuan *Vision Transformer* dalam memodelkan konteks global, sehingga diharapkan mampu menghasilkan representasi fitur yang lebih komprehensif dan meningkatkan akurasi klasifikasi citra medis, termasuk citra penyakit gigi dan mulut.

Selain arsitektur *model*, keterbatasan jumlah dan distribusi data menjadi permasalahan utama dalam pengembangan sistem klasifikasi citra medis. *Dataset*

citra penyakit gigi dan mulut umumnya memiliki jumlah data yang terbatas dan distribusi kelas yang tidak seimbang. Oleh karena itu, diperlukan teknik data augmentation yang efektif untuk meningkatkan keragaman data pelatihan. Salah satu teknik augmentasi yang relevan adalah *CutMix*, yaitu metode yang menggabungkan dua citra dengan memotong dan menempelkan sebagian area citra serta mengombinasikan labelnya. *CutMix* terbukti mampu meningkatkan kemampuan generalisasi *model* dan mengurangi *overfitting*, sehingga sesuai untuk diterapkan pada klasifikasi citra penyakit gigi dan mulut berbasis *deep learning*.

2.2. Celah Penelitian

Berbagai penelitian telah membuktikan bahwa arsitektur *Convolutional Neural Network* (CNN), termasuk *Efficientnet*, mampu memberikan performa yang baik dalam klasifikasi citra medis seperti penyakit gigi, pulpitis, dan penyakit kulit. CNN efektif dalam mengekstraksi fitur lokal pada citra, sehingga banyak digunakan dalam berbagai studi diagnosis berbasis citra.

Di sisi lain, perkembangan arsitektur *Vision Transformer* (ViT) memperkenalkan pendekatan baru dalam pengolahan citra dengan kemampuan memodelkan hubungan global antarbagian citra melalui mekanisme self-attention. Beberapa penelitian menunjukkan bahwa *Vision Transformer* memiliki potensi yang menjanjikan dalam bidang *medical imaging*. Namun, penerapannya masih menghadapi tantangan, terutama pada *dataset* medis yang terbatas, serta belum banyak diintegrasikan dengan arsitektur CNN secara optimal pada kasus gigi dan mulut.

Meski demikian, masih terdapat beberapa kesenjangan (research gap) yang belum sepenuhnya dijawab oleh penelitian terdahulu, yaitu:

- 1) Terbatasnya penelitian yang mengombinasikan arsitektur *Efficientnet* dan *Vision Transformer* pada klasifikasi citra penyakit gigi dan mulut. Sebagian besar penelitian sebelumnya masih menggunakan *model* tunggal berbasis CNN atau *Vision Transformer* secara terpisah. Penelitian yang secara khusus mengembangkan dan mengevaluasi *model hybrid EfficientNet–Vision Transformer* pada domain penyakit

gigi dan mulut masih sangat terbatas..

- 2) Minimnya eksplorasi teknik data augmentation lanjutan seperti *CutMix Augmentation* pada citra penyakit gigi dan mulut. Penelitian terdahulu umumnya hanya menerapkan augmentasi dasar seperti rotasi, flipping, dan scaling. Pemanfaatan *CutMix Augmentation* sebagai teknik augmentasi untuk meningkatkan kemampuan generalisasi *model* pada klasifikasi citra penyakit gigi dan mulut belum banyak dikaji secara mendalam.
- 3) Kurangnya penelitian yang mengintegrasikan *model* klasifikasi citra khususnya penyakit gigi dan mulut ke dalam sistem aplikasi yang dapat digunakan secara praktis. Sebagian besar penelitian masih berfokus pada evaluasi performa *model* di lingkungan eksperimen, tanpa membahas implementasi *model* ke dalam sistem aplikasi yang dapat dimanfaatkan secara langsung sebagai alat bantu diagnosis..

Oleh karena itu, penelitian ini diarahkan untuk mengisi celah tersebut dengan mengembangkan *model hybrid EfficientNet–Vision Transformer* yang dikombinasikan dengan teknik *CutMix Augmentation* untuk klasifikasi citra penyakit gigi dan mulut. Selain itu, *model* yang dikembangkan diimplementasikan ke dalam aplikasi berbasis Streamlit sebagai sistem diagnosis penyakit gigi dan mulut berbasis citra, sehingga diharapkan dapat memberikan kontribusi baik secara metodologis maupun praktis.

Tabel 2. 1. Tabel Perbandingan Penelitian

Penulis	Objek yang diteliti	<i>Model</i> / Arsitektur	Teknik Augmentasi	Keterbatasan Penelitian
Wahyuningsih, Nugraha, dan Dwiyanaputra [5]	Penyakit gigi (karies) berbasis citra digital	<i>Efficientnet-B0 (CNN)</i>	Augmentasi dasar (rotasi, flipping)	Menggunakan <i>model</i> CNN tunggal, belum mempertimbangkan pemodelan konteks global pada citra

Penulis	Objek yang diteliti	Model / Arsitektur	Teknik Augmentasi	Keterbatasan Penelitian
Lavenia, Ramdani, dan Hoeronis [6]	Pulpitis (citra radiografi periapikal)	<i>Efficientnet-B2 (CNN)</i>	Tidak dijelaskan secara spesifik	Terbatas pada CNN murni dan citra radiografi, tanpa eksplorasi arsitektur lanjutan
Andika Nur Pratama [7]	Penyakit kulit berbasis citra medis	<i>Efficientnet-B2 (CNN)</i>	Augmentasi dasar	Belum mengombinasikan <i>Efficientnet</i> dengan arsitektur lain dan belum membahas augmentasi lanjutan
Reyvan Revolusioner Ar et al. [8]	Stuntin g pada balita berbasis citra tubuh	Vision Transformer (ViT)	<i>Tidak dijelaskan secara spesifik</i>	Bergantung pada kualitas citra dan jumlah data pelatihan yang memadai
Irvan Yudistiansyah [9]	Penyakit mata berbasis citra fundus	Vision Transformer (ViT)	<i>Tidak dijelaskan secara spesifik</i>	Sensitif terhadap variasi kualitas citra dan membutuhkan data latih yang relatif besar
Penulis	Penyakit gigi dan mulut berbasis citra	<i>Hybrid Efficientnet–Vision Transformer</i>	<i>CutMix Augmentatio n</i>	

2.3. Landasan Teori

2.3.1. Penyakit Gigi dan Mulut

Penyakit gigi merupakan bagian dari penyakit mulut yang menyerang jaringan keras dan jaringan pendukung gigi, serta dapat memengaruhi fungsi mastikasi, estetika, dan kualitas hidup penderitanya. Penyakit gigi dapat disebabkan oleh faktor bakteri, kebersihan mulut yang buruk, pola makan, maupun faktor genetik. World Health Organization (WHO) menyatakan bahwa penyakit gigi dan mulut merupakan salah satu masalah kesehatan paling umum di dunia dan berdampak signifikan terhadap kesehatan masyarakat[1]. Beberapa jenis penyakit gigi dan mulut yang umum ditemukan dan menjadi fokus dalam penelitian ini dijelaskan sebagai berikut.

2.3.1.1. Karies (Gigi Berlubang)

Karies atau gigi berlubang merupakan penyakit yang paling umum diderita. Karies gigi terjadi karena proses demineralisasi pada jaringan keras gigi yang disebabkan oleh mikroorganisme seperti *Streptococcus mutans* dan *Lactobacillus*. Keberadaan mikroorganisme ini memfermentasi karbohidrat menjadi asam. Asam inilah yang menyebabkan hancurnya mineral pada gigi sehingga akan muncul lubang[10].



Gambar 2. 1 Karies

2.3.1.2. Calculus (Karang Gigi)

Calculus atau yang biasa dikenal dengan karang gigi didefinisikan sebagai

plak gigi termineralisasi yang terbentuk melalui interaksi kompleks antara saliva dan bakteri yang melekat di permukaan gigi[11]. Ketika saliva, sisa makanan, dan protein berinteraksi dengan bakteri *Streptococcus mutans*, maka akan terbentuk lapisan lendir di gigi bagian dalam. Lama-kelamaan, lapisan lendir itu akan mengeras dan membentuk karang gigi.



Gambar 2. 2 Calculus

2.3.1.3. *Gingivitis* (Radang Gusi)

Gingivitis atau radang gusi merupakan suatu inflamasi yang melibatkan jaringan lunak di sekitar gigi yaitu jaringan gingiva. Penyebab *gingivitis* dibagi menjadi 2, yaitu penyebab utama dan penyebab predisposisi. Penyebab utama dari *gingivitis* adalah plak gigi yang dibentuk akibat mikroorganisme yang melekat pada tepi gingiva. Sedangkan untuk penyebab predisposisi terjadi karena restorasi gigi yang gagal, yang salah seperti faktor obat-obatan[12].



Gambar 2. 3 *Gingivitis*

penggunaan gigi palsu hingga faktor sistemik hormonal, nutrisi, dan

2.3.1.4. *Tooth Discolouration*

Tooth Discolouration atau perubahan warna gigi merupakan kondisi perubahan warna alami gigi yang dapat disebabkan oleh faktor intrinsik maupun ekstrinsik. Faktor ekstrinsik meliputi konsumsi makanan dan minuman berwarna, kebiasaan merokok, serta akumulasi plak dan karang gigi, sedangkan faktor intrinsik dapat disebabkan oleh gangguan pembentukan enamel dan dentin, penggunaan obat-obatan tertentu, serta faktor genetik. Perubahan warna gigi tidak selalu menimbulkan rasa sakit, namun dapat memengaruhi estetika dan kepercayaan diri seseorang. Kementerian Kesehatan Republik Indonesia menyatakan bahwa perubahan warna gigi sering ditemukan pada masyarakat dengan kebersihan gigi dan mulut yang kurang baik serta pola hidup yang tidak sehat [13].



Gambar 2. 4 Tooh Discolouration

2.3.1.5. *Mouth ulcer*

Mouth ulcers atau sariawan merupakan luka terbuka pada jaringan lunak rongga mulut yang biasanya berbentuk bulat atau oval, berwarna putih atau kekuningan, dan dikelilingi oleh jaringan kemerahan. Sariawan dapat disebabkan oleh berbagai faktor, antara lain trauma mekanis, kekurangan nutrisi, infeksi, stres, serta gangguan sistem imun. Meskipun sebagian besar bersifat ringan dan dapat sembuh dengan sendirinya, sariawan dapat menimbulkan rasa nyeri yang mengganggu aktivitas makan dan berbicara. Menurut Kementerian Kesehatan Republik Indonesia, sariawan termasuk salah satu keluhan kesehatan mulut yang paling sering dialami oleh masyarakat Indonesia[14].

2.3.1.6. *Hypodontia*

Hypodontia merupakan kelainan perkembangan gigi yang ditandai dengan tidak tumbuhnya satu atau lebih gigi permanen akibat faktor genetik atau gangguan selama proses pembentukan gigi. Kondisi ini dapat menyebabkan gangguan fungsi pengunyahan, estetika wajah, serta susunan gigi secara keseluruhan. *Hypodontia* umumnya terdeteksi melalui pemeriksaan klinis dan radiografi gigi. Penelitian di Indonesia menunjukkan bahwa *hypodontia* merupakan salah satu kelainan gigi kongenital yang cukup sering ditemukan, meskipun sering tidak terdiagnosis sejak dini [15].



Gambar 2. 5 Hypodontia

2.3.2. *Deep learning*

Deep learning adalah cabang dari *machine learning*. Algoritma dari *deep*

learning terinspirasi dari struktur otak manusia. Struktur tiruan tersebut kemudian dinamakan sebagai jaringan saraf tiruan atau *artificial network neuron* (ANN). *Deep learning* juga memiliki beberapa algoritma yang masing-masing dari algoritma tersebut memiliki fungsi tersendiri. Contoh dari algoritma *deep learning* yaitu : *Convolutional Neural Network* (CNN), *Long Short Term Memory Network* (LSTM), (RNN), *Self Organizing Maps* (SOM)[16].

2.3.3. *Convolutional Neural Network* (CNN)

Convolutional Neural Network (CNN) adalah salah satu arsitektur *deep learning* yang paling banyak digunakan untuk pemrosesan citra digital dan pengenalan pola visual. CNN mampu belajar otomatis representasi fitur dari data berstruktur *grid* (mis. *pixel* citra) melalui operasi *convolution*, fungsi aktivasi non-linier, serta lapisan *pooling* dan *fully connected*. Arsitektur ini memiliki keunggulan dibanding pendekatan tradisional karena tidak memerlukan ekstraksi fitur manual dan mampu mengekstrak fitur dari tingkat rendah ke tingkat tinggi secara hierarkis[17].

Cara kerja dari CNN adalah CNN memiliki beberapa lapisan (*layer*) yang mengekstrak informasi dari gambar. Lapisan awal mengidentifikasi fitur sederhana seperti garis dan tepi, sedangkan lapisan yang lebih dalam mengenali pola, bentuk, dan akhirnya, seluruh objek yang lebih kompleks. Setelah itu maka gambar akan diklasifikasi sesuai skor klasifikasi[18].

Meskipun memiliki keunggulan yang mampu mengekstraksi fitur secara langsung, *model* CNN juga memiliki kelemahan, yaitu CNN memerlukan sumber daya komputasi yang tinggi, waktu pelatihan yang lama, serta cenderung sulit melakukan generalisasi pada data baru[19]. Oleh karena itu CNN telah ditingkatkan dan dimodifikasi ke dalam berbagai arsitektur modern yang lebih efisien, seperti *MobileNetV2* untuk perangkat dengan sumber daya terbatas dan *Efficientnet* yang menyeimbangkan kedalaman dan resolusi *model* untuk akurasi yang lebih tinggi.

2.3.4. *Efficientnet*

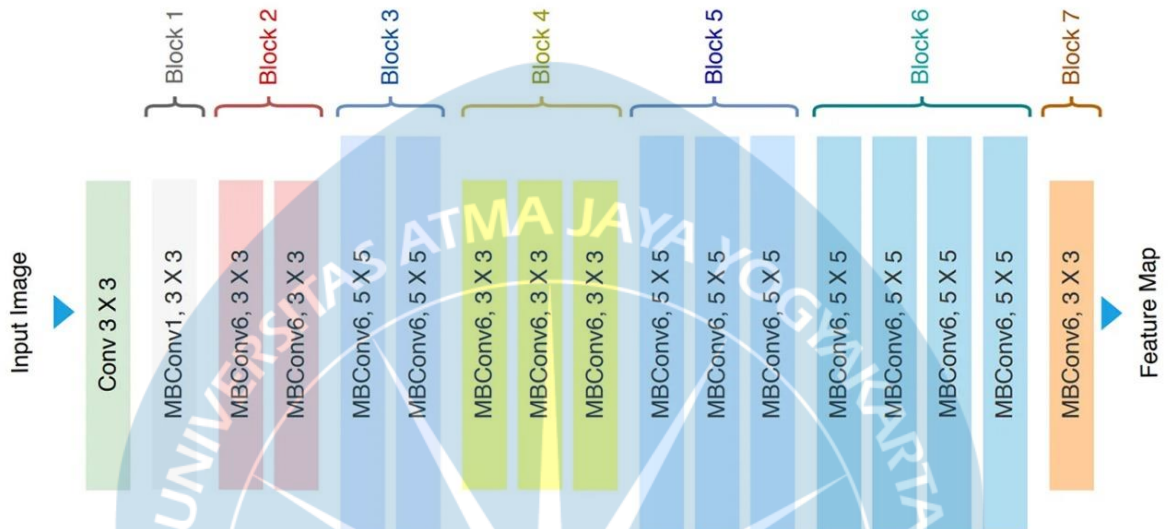
Efficientnet adalah arsitektur Convolutional Neural Network (CNN) yang dirancang untuk menggabungkan akurasi dan efisiensi komputasi secara optimal. Arsitektur ini diusulkan oleh Tan dan Le sebagai solusi terhadap keterbatasan pendekatan *model* konvensional yang biasanya hanya meningkatkan satu elemen jaringan, seperti kedalaman atau lebar, tanpa mempertimbangkan keseimbangan jaringan secara keseluruhan.

Berbeda dengan CNN, *Efficientnet* menggunakan metode penskalaan campuran. Metode ini dapat menyesuaikan tiga dimensi utama jaringan. Tiga dimensi jaringan itu adalah kedalaman (depth), lebar (width), dan resolusi (resolution) secara bersamaan dan seimbang. Metode ini memungkinkan *Efficientnet* untuk mencapai efisiensi tanpa mengorbankan akurasi. Selain itu, penggunaan *Efficientnet* juga dinilai mampu untuk beradaptasi sesuai dengan kemampuan perangkat keras dan juga berbagai anggaran komputasi[20].

Arsitektur *Efficientnet* dibangun berdasarkan blok dasar yang disebut *Mobile Inverted Bottleneck Convolution (MBConv)*. Blok dasar ini terdiri dari lapisan *depthwise separable convolution*, *batch normalization*, serta fungsi aktivasi *Swish*, yang dirancang untuk meningkatkan efisiensi komputasi tanpa mengorbankan akurasi. Selain itu, *Efficientnet* juga menggunakan mekanisme *Squeeze-and-Excitation (SE)* yang bertujuan untuk meningkatkan kemampuan jaringan dalam menekankan fitur-fitur penting pada setiap channel[21].

Efficientnet juga memiliki beberapa varian *model*, mulai dari *Efficientnet-B0* hingga *Efficientnet-B7*, yang dihasilkan melalui penerapan *compound scaling*. Varian-varian ini memungkinkan peneliti untuk dapat bereksperimen menggunakan *Efficientnet* sesuai dengan kebutuhan dan kompatibilitas perangkat keras masing-masing.

Pada masa sekarang, *Efficientnet* telah banyak digunakan untuk berbagai penelitian. Beberapa peneliti telah menggunakan *model* ini untuk membantu mengklasifikasi beerbagai penyakit. *Efficientnet* mampu memberikan kinerja yang



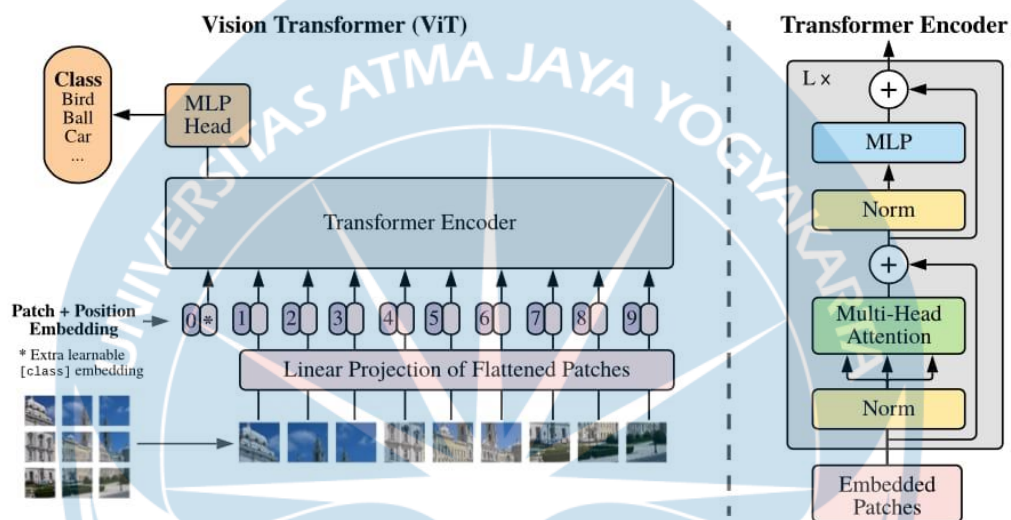
Gambar 2. 6 Compound Scaling pada EfficientNet, Source: ultralytics.com

baik pada *dataset* dengan jumlah data terbatas karena efisiensi parameter dan kemampuannya dalam mengekstraksi fitur lokal secara efektif.

2.3.5. Vision Transformer

Vision Transformer atau yang biasa disingkat ViT adalah evolusi dari arsitektur *Transformer*, yang awalnya arsitektur *Transformer* dikembangkan hanya untuk memproses bahasa alami. Berbeda dengan CNN yang hanya mengekstraksi fitur lokal, cara kerja dari ViT adalah dengan mempelajari hubungan global antar citra dengan memanfaatkan fitur *self-attention*. Pendekatan ini mampu membuat ViT untuk menangkap konteks global dengan lebih efektif terutama pada citra yang kompleks.

Vision Transformer memulai dengan membagi gambar yang dimasukkan menjadi beberapa bagian kecil berukuran tetap yang disebut *patch*. Setiap *patch* kemudian melalui proses *flatten* atau diratakan yang kemudian diproyeksikan melalui lapisan *linear embedding* ke dalam vektor berdimensi tertentu. Setelah itu, vektor-vektor ini diperkaya dengan *positional embedding* untuk mempertahankan informasi posisi spasial sebelum diproses oleh encoder *Transformer*[22].



Gambar 2. 7 Arsitektur Pada Vision Transformer. Source:Albutsimple.com

Komponen utama Dalam arsitektur *Vision Transformer* adalah *patch embedding*, *Transformer encoder*, dan *classification head*. Tahap pertama yaitu *patch embedding* berfungsi mengubah citra menjadi urutan vektor (*sequence of tokens*) yang dapat diproses oleh Transformer. Kemudian citra yang telah dibagi direpresentasikan oleh sebuah token. Selain itu, maka ditambahkanlah *Positional embedding* untuk mempertahankan urutan spasial antarpatch.

Selanjutnya adalah bagian *Transformer encoder*. *Transformer Encoder* terdiri dari beberapa lapisan *multi-head self-attention* dan *feed-forward network*. Mekanisme *self-attention* memungkinkan setiap patch untuk berinteraksi dengan seluruh patch lainnya dalam citra, sehingga *model* dapat mempelajari hubungan global secara menyeluruh. Pada tahap akhir, representasi fitur yang dihasilkan digunakan oleh *classification head* untuk menentukan kelas keluaran citra[23].

Sama seperti *Efficientnet*, penggunaan *model* arsitektur *Vision Transformer*

juga sudah banyak digunakan dalam penelitian, tidak terkecuali citra medis. Beberapa penelitian melaporkan bahwa ViT mampu memberikan hasil yang baik dalam klasifikasi citra penyakit ketika digunakan pada *dataset* berskala besar atau dikombinasikan dengan arsitektur CNN. Pendekatan hybrid CNN–*Transformer* terbukti mampu meningkatkan akurasi dengan memanfaatkan ekstraksi fitur lokal dari CNN dan pemodelan konteks global dari *Transformer*.

2.3.6. *CutMix Augmentation*

CutMix Augmentation adalah salah satu teknik augmentasi data yang digunakan untuk meningkatkan performa dari *model deep learning* pada tugas klasifikasi citra supaya tidak *overfit*. *CutMix Augmentation* bekerja dengan memotong sebagian area dari sebuah citra dan menempelkannya pada citra lain. Kemudian potongan dari citra-citra tersebut disatukan sehingga bisa membentuk citra yang baru. Pendekatan ini bertujuan untuk mendorong *model* agar mempelajari fitur visual yang lebih beragam dan tidak bergantung pada satu area tertentu dalam citra[24].

CutMix Augmentation banyak digunakan pada *model* CNN dan arsitektur *deep learning* lainnya karena bisa mengurangi *overfitting*, khususnya penelitian yang memiliki *dataset* terbatas. Penggunaan *CutMix Augmentation* bisa membantu meningkatkan ketahanan *model* terhadap gangguan visual dan variasi data. Penggunaan *CutMix Augmentation* sangat relevan dalam citra medis karena pada umumnya *dataset* citra medis masih sangat terbatas. Penggunaan *CutMix Augmentation* dapat memperbanyak variasi data penelitian. Teknik ini membantu *model* untuk lebih fokus pada fitur-fitur penting yang tersebar di berbagai area citra[25].

2.3.7. *Evaluation Metrics*

Evaluation metrics merupakan komponen penting dalam penelitian *machine learning* dan *deep learning* untuk mengukur kinerja *model* klasifikasi secara objektif dan kuantitatif. Metrik evaluasi digunakan untuk menilai sejauh mana *model* mampu melakukan prediksi yang benar berdasarkan data uji yang tersedia.

Pemilihan metrik yang tepat sangat penting, terutama pada kasus klasifikasi multikelas dan *dataset* yang tidak seimbang, karena satu metrik saja (seperti akurasi) sering kali tidak cukup untuk merepresentasikan performa *model* secara menyeluruh. Berikut adalah metrik-metrik utama dan rumusnya:

2.3.7.1. Accuracy

Accuracy merupakan metrik evaluasi yang paling umum digunakan untuk mengukur kinerja *model* klasifikasi. *Accuracy* menunjukkan proporsi prediksi yang benar terhadap seluruh jumlah data yang diuji. Metrik ini memberikan gambaran umum tentang tingkat ketepatan *model*, namun dapat menjadi kurang representatif pada kondisi *dataset* yang tidak seimbang antar kelas. Secara matematis, *accuracy* dirumuskan sebagai berikut [26]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

di mana:

- **TP (True Positive)** adalah jumlah data positif yang diprediksi benar,
- **TN (True Negative)** adalah jumlah data negatif yang diprediksi benar,
- **FP (False Positive)** adalah jumlah data negatif yang diprediksi sebagai positif,
- **FN (False Negative)** adalah jumlah data positif yang diprediksi sebagai negatif.

2.3.7.2. F1-Score

F1-score merupakan metrik evaluasi yang menggabungkan *precision* dan *recall* dalam satu nilai harmonik. Metrik ini sangat berguna pada kasus klasifikasi dengan distribusi kelas yang tidak seimbang, karena mampu merepresentasikan keseimbangan antara ketepatan prediksi dan kemampuan *model* dalam menemukan seluruh data relevan.

F1-score didefinisikan sebagai berikut [27]:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Dengan :

$$Precision = \frac{TP}{TP + FN}$$

$$Recall = \frac{TP}{TP + FN}$$

Nilai *F1-score* berada pada rentang 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan performa klasifikasi yang lebih baik.

2.3.7.3. Confusion Matrix

Confusion matrix merupakan tabel evaluasi yang digunakan untuk menggambarkan kinerja *model* klasifikasi secara rinci dengan membandingkan hasil prediksi *model* terhadap label sebenarnya. *Confusion matrix* menampilkan jumlah prediksi benar dan salah untuk setiap kelas, sehingga memudahkan analisis kesalahan klasifikasi yang dilakukan oleh *model*.

Pada kasus klasifikasi multikelas, *confusion matrix* berbentuk matriks persegi dengan ukuran sesuai jumlah kelas. Setiap baris merepresentasikan kelas aktual, sedangkan setiap kolom merepresentasikan kelas hasil prediksi. *Confusion matrix* menjadi dasar perhitungan metrik lain seperti *accuracy*, *precision*, *recall*, dan *F1-score* [28].

2.3.7.4. Classification Report

Classification report merupakan ringkasan evaluasi kinerja *model* klasifikasi yang menyajikan nilai *precision*, *recall*, *F1-score*, dan *support* untuk setiap kelas. *Support* menunjukkan jumlah sampel aktual pada masing-masing kelas. *Classification report* memberikan gambaran yang lebih lengkap dibandingkan penggunaan satu metrik tunggal, terutama untuk menganalisis performa *model* pada setiap kelas secara individual.

Penggunaan *classification report* sangat direkomendasikan dalam evaluasi *model* klasifikasi multikelas karena mampu menunjukkan kelas mana yang telah diklasifikasikan dengan baik dan kelas mana yang masih mengalami kesalahan prediksi [29].

