

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1. Tinjauan Pustaka

Penelitian terdahulu yang menjadi tinjauan pustaka dalam penelitian ini berfokuskan pada pemanfaatan data multimodal khususnya X-ray dan CT-Scan dalam membangun model klasifikasi citra yang digunakan dalam melakukan proses diagnosa penyakit. Selain itu juga, penelitian pada klasifikasi citra yang memanfaatkan Dempster-Shafer Theory juga digunakan sebagai tinjauan pustaka pada penelitian ini.

Pemanfaatan citra CT-Scan dan CXR sebagai modalitas dalam melakukan klasifikasi penyakit paru-paru berbasis *deep learning* telah dilakukan dalam berbagai penelitian. Al-Sheikh et al. [23] mengusulkan sistem deteksi untuk melakukan klasifikasi berbagai penyakit paru-paru menggunakan dataset CT-Scan dan CXR secara terpisah. Pemanfaatan kedua modalitas didasarkan pada banyaknya penelitian lain yang menunjukkan kemampuan keduanya dalam melakukan diagnosa dini penyakit paru-paru, sekaligus ketersediaannya kedua modalitas secara publik. Proses klasifikasi dilakukan setelah menggabungkan arsitektur CNN kustom dengan dua model *pre-trained deep learning*, yaitu AlexNet dan VGG16Net, dilengkapi dengan algoritma *image enhancement* berbasis fungsi *k-symbol Lerch transcendent*. Hasil eksperimen menunjukkan akurasi klasifikasi sebesar 98,60%, 98,40%, and 98,50% pada dataset citra CXR dan 98,80%, 98,50%, 98,40% pada dataset citra CT-Scan.

Selain itu, pemanfaatan citra CXR dan CT-Scan paru-paru dalam pengembangan model klasifikasi pneumonia, dilakukan oleh Moussaid et al. [24] yaitu dengan memanfaatkan arsitektur *Deep Learning* berbasis *EfficientNetB7* untuk mengklasifikasi citra kedua modalitas menjadi tiga kelas berbeda, yaitu pneumonia umum, pneumonia COVID-19, dan normal. Kedua modalitas yang digunakan memiliki alasan tersendiri mengapa keduanya terpilih. Dengan X-ray digunakan dengan alasan ketersediaan data yang luas serta banyak penelitian terdahulu yang membuktikan efektivitas modalitas ini dalam melakukan diagnosa penyakit pneumonia. Penggunaan CT-Scan didasarkan pada kemampuannya memberikan gambaran lebih detail serta menjadi metode yang direkomendasikan oleh WHO dalam melakukan diagnosa penyakit pneumonia [24]. Model menunjukkan akurasi yang tinggi dengan akurasi 99,81% pada citra CXR dan 99,88% pada citra CT-Scan.

Pemanfaatan pendekatan multimodal CT-Scan dan CXR untuk melakukan klasifikasi pneumonia juga didemonstrasikan pada penelitian yang dilakukan oleh Hilmizen et al. [25]. Penelitian ini mengusulkan model *deep learning* multimodal yang menggabungkan citra CT-Scan dan CXR secara terintegrasi dalam melakukan diagnosis pneumonia COVID-19. Penelitian ini dilatarbelakangi oleh keterbatasan pendekatan single-modality yang hanya memanfaatkan satu jenis modalitas pencitraan, sehingga menghasilkan akurasi klasifikasi yang tidak optimal. Pemilihan CT-Scan dan CXR dikarenakan keefektifitas keduanya dalam melakukan klasifikasi penyakit pneumonia dengan memanfaatkan keduanya dalam multimodal tiap modalitas akan memberikan informasi komplementer yang melengkapi satu dengan yang lainnya. Dengan memanfaatkan *Transfer Learning* pada kedua modalitas dan menggabungkan prediksinya melalui mekanisme fusi, model yang diusulkan mampu menghasilkan performa klasifikasi dengan nilai akurasi 99,87% lebih tinggi dari pendekatan berbasis modalitas tunggal yang berakurasi 98,00% pada citra CT-Scan dan 98,93% pada citra X-Ray

Penelitian lain menunjukkan pemanfaatan multimodal telah dilakukan oleh Chen et al. [26]. Penelitian ini didasarkan akan kurangnya efektivitas dari model berbasis modalitas tunggal dalam melakukan tugasnya. Penelitian ini berfokus dalam melakukan diagnosa penyakit kanker payudara memanfaatkan modalitas mamografi digital dan citra ultrasonik. Kedua modalitas saling melengkapi dalam proses pendeteksian pada payudara, mamografi digital dapat mendeteksi mikrokalsifikasi yang terlewat oleh citra ultrasonik, sementara citra ultrasonik dapat memberikan tambahan informasi pada area kelenjar yang sulit dibedakan pada mamografi digital. Dalam penentuan model dasar yang digunakan, penelitian ini mencoba enam model lain sebelum akhirnya menggunakan ResNet-18 dikarenakan performanya pada data mamografi digital dan citra ultrasonik dibandingkan model lain yang dicoba. Hasil penelitian menunjukkan model multimodal mengungguli model modalitas tunggal dalam hampir semua metrik evaluasi, dengan skor AUC sebesar 0,968, akurasi 93,78%, spesifisitas 96,41, dan presisi 83,66%. Meskipun model modalitas tunggal lebih unggul dalam hal sensitivitas, model multimodal secara signifikan mampu mengurangi angka positif palsu.

Kemudian penelitian pada pemanfaatan *Evidential Neural Network* dalam melakukan klasifikasi citra dilakukan oleh Huang et al. [21]. Pada penelitian ini pendekatan *Evidential Neural Network* digunakan dalam melakukan klasifikasi kasus COVID-19 dari citra CT-Scan. Metode yang diusulkan mengekstraksi fitur mendalam dari citra CT-Scan, memetakannya ke dalam *belief degree maps*, kemudian menghasilkan keputusan klasifikasi

akhir berdasarkan representasi *belief functions* tersebut. Hasil penelitian menunjukkan bahwa pendekatan berbasis *belief functions* mampu menghasilkan performa yang kompetitif dibandingkan model deep learning konvensional, dengan akurasi 0,81, F1-score 0,812, dan AUC 0,875. Lebih lanjut, model berbasis *belief functions* menunjukkan keunggulan yang lebih signifikan dibandingkan model konvensional pada skenario semi-supervised, di mana ketersediaan data berlabel sangat terbatas. Selain itu, representasi *belief functions* yang dihasilkan memberikan tingkat interpretabilitas yang lebih tinggi dibandingkan model deep learning konvensional, karena setiap keputusan klasifikasi disertai dengan nilai keyakinan dan ketidakpastian yang dapat dijelaskan secara kuantitatif.

Evidential Neural Network juga telah digunakan dalam melakukan tugas segmentasi pada citra medis *multimodal*, yang didemonstrasikan pada penelitian Huang et al. [10]. Penelitian ini menggunakan kerangka *deep evidential fusion* untuk segmentasi citra medis *multimodal* yang menggabungkan DST dengan *deep neural network*. Penelitian ini dilakukan pada dua macam penyakit yang berbeda, penyakit tumor otak yang dilakukan menggunakan *multimodal Multi-MRI* dan penyakit limfoma yang dilakukan menggunakan *multimodal PET-CT*. Dalam prosesnya, fitur diekstraksi dari setiap modalitas pencitraan, dipetakan ke dalam mass functions, kemudian dikoreksi menggunakan operasi *contextual discounting* dengan koefisien yang dipelajari untuk mengkuantifikasi tingkat keandalan setiap modalitas terhadap setiap kelas, sebelum akhirnya digabungkan menggunakan *Dempster's rule of combination*. Dengan metode yang digunakan dalam penelitian ini model memberikan hasil yang lebih baik secara akurasi serta keandalan dalam melakukan segmentasi.

Tabel 2.1 Ringkasan Tinjauan Pustaka

Peneliti (tahun)	Topik	Metode	Objek	Hasil
Mona Hmoud Al-Sheikh et al. (2023) [23]	Klasifikasi penyakit paru-paru berbasis <i>deep learning</i>	AlexNet, VGG16Net, dan algoritma <i>image enhancement</i> berbasis k-symbol Lerch transcendent	Citra CXR dan CT-Scan	Model menghasilkan akurasi klasifikasi sebesar 98,60%, 98.40%, and 98.50% pada dataset citra

				CXR dan 98,80%, 98.50%, 98.40% pada dataset citra CT-Scan.
Moussaid et al. (2023) [24]	Klasifikasi pneumonia berbasis <i>deep learning</i>	<i>Deep learning</i> berbasis EfficientNetB7	Citra CXR dan CT-Scan	Model berakurasi 99,81% pada citra CXR dan 99,88% pada citra CT-Scan.
Hilmizen et al. (2020) [25]	Pemanfaatan multimodal CT-Scan dan CXR dalam klasifikasi pneumonia	DenseNet121, MobileNet, Xception, InceptionV3, ResNet50, dan VGG16	Multimodal CT-Scan dan CXR	Model multimodal menghasilkan performa klasifikasi lebih tinggi dari pendekatan berbasis modalitas tunggal
Chen et al. (2025) [26]	Pemanfaatan multimodal mamografi digital dan citra ultrasonik dalam klasifikasi penyakit kanker payudara	ResNet-18	Multimodal mamografi digital dan citra ultrasonik	Model multimodal menghasilkan performa klasifikasi lebih tinggi daripada pendekatan berbasis modalitas tunggal, serta model secara signifikan

				mampu mengurangi angka positif palsu.
Huang et al. (2021) [21]	Pemanfaatan <i>Evidential Neural Network</i> dalam melakukan klasifikasi penyakit COVID-19 pneumonia	Demspter-Shafer Theory, <i>Deep Learning</i>	CT-Scan	Menghasilkan model yang kompetif dengan model konvensional lainnya serta pemanfaatan belief function menjadikan hasil dari model mudah dijelaskan.
Huang et al. (2025) [10]	Segmentasi citra medis multimodal yang menggabungkan DST dengan <i>deep neural network</i>	Deep Neural Network, <i>Evidential Neural Network</i>	<i>Multi-MRI</i> pada penyakit tumor otak dan PET-CT pada penyakit limfoma	Menghasilkan model yang kuantifikasi ketidakpastian dan segmentasi gambar yang lebih baik dibandingkan model lainnya.

2.2. Landasan Teori

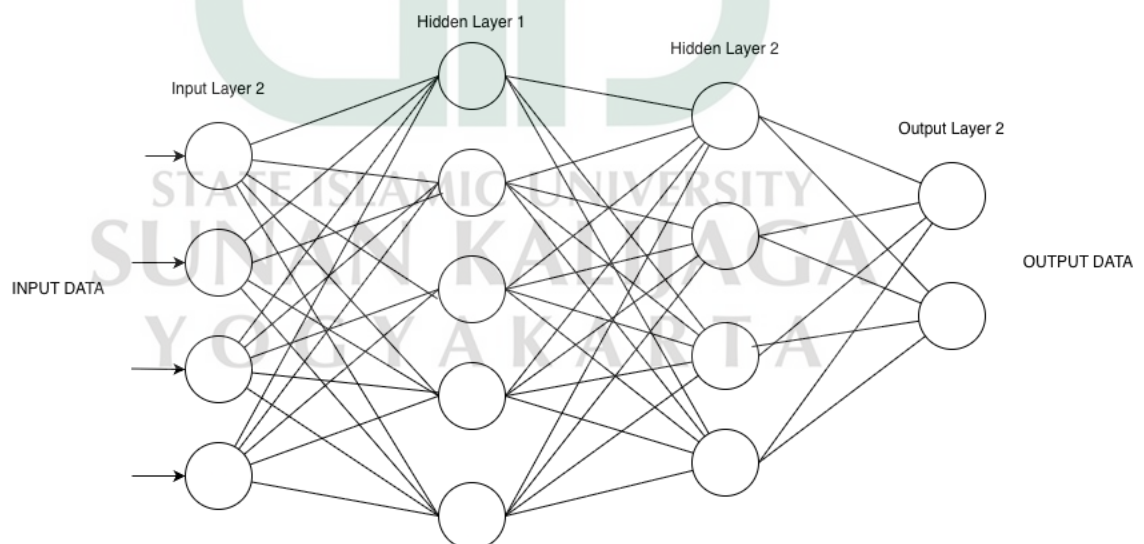
2.2.1. Machine Learning

Machine Learning (ML) merupakan bagian dari kecerdasan buatan atau *Artificial Intelligences* yang memungkinkan komputer untuk belajar dan meningkatkan kinerjanya tanpa diperlukannya pemrograman secara eksplisit [27]. *Machine Learning* meniru bagaimana manusia belajar dari pengalaman, model *machine learning* akan menggunakan hasil pelatihan sebelumnya untuk mengenali pola dan menghasilkan prediksi terhadap data yang baru [28].

Machine learning terbagi menjadi tiga tipe *supervised learning*, *unsupervised learning*, and *reinforcement learning*, pembagian ini didasarkan masalah yang diberikan dan kesediaan data [29]. Dalam pendekatan *supervised learning*, komputer dilatih menggunakan data masukan (input) yang disertai dengan label keluaran (output) yang telah ditentukan. Melalui proses pelatihan ini, model belajar mengenai pola dan hubungan antara data masukan dan keluaran [28]. Proses inilah yang memungkinkan untuk model memprediksikan hasil pada data baru yang belum pernah dilihat. Pendekatan ini dapat digunakan dalam berbagai macam hal, salah satunya dalam bidang medis, terlebih pada pencitraan medis. Model klasifikasi citra medis terbukti efektif dalam melakukan diagnosis pada berbagai macam penyakit [30].

2.2.2. Deep Learning

Deep Learning merupakan bagian dari *machine learning* yang memanfaatkan jaringan saraf (*neural network*) dengan banyak lapisan dalam melakukan penyelesaian masalah [27]. Neural network pada *deep learning* berkerja sama hal nya seperti neuron di otak seorang manusia. Berbeda dari metode *machine learning*, *deep learning* mampu mempelajari data tanpa perancangan fitur manual [31]. Pemanfaatan *neural network* menjadikan *deep learning* efektif dalam melakukan permasalahan kompleks seperti, rekognisi citra dan suara [27].



Gambar 2.1 Jaringan Deep Learning

Tiap lapisan pada neural network memiliki fungsinya masing-masing. Pada lapisan pertama atau biasa disebut input layer bertugas untuk menerima data sebagai masukan. Lapisan kedua biasa tersusun dari satu atau lebih *hidden layer* yang bertugas

dalam mengekstraksi dan mentransformasi citra. Semakin banyak hidden layer yang digunakan semakin kompleks juga representasi fitur yang dipelajari. Lapisan terakhir atau output layer menghasilkan keputusan berdasarkan fitur yang telah diekstraksi oleh lapisan-lapisan sebelumnya.

Dalam konteks pengolahan citra, arsitektur *deep learning* yang paling umum digunakan adalah *Convolutional Neural Network* (CNN). CNN terbukti sangat sukses dalam melakukan tugas rekognisi citra ditunjukkan dengan menghasilkan nilai terbaik di beberapa tes yang dilakukan [32]. Semua ini dikarenakan kemampuan CNN dalam menangkap fitur dan pola spasial yang ada pada gambar [32]. Reliabilitas inilah yang menjadikan CNN banyak digunakan dalam tugas pencitraan medis, seperti klasifikasi citra, deteksi objek, dan segmentasi citra.

2.2.3. Pembelajaran Multimodal pada Citra

Multimodal learning atau pembelajaran multimodal merupakan pendekatan pembelajaran mesin yang memanfaatkan lebih dari satu jenis sumber data atau modalitas untuk meningkatkan kualitas dan keandalan model dalam melakukan pengambilan keputusan. Setiap modalitas membawa informasi yang berbeda dan saling melengkapi, sehingga integrasi beberapa modalitas dapat menghasilkan representasi yang lebih kaya dibandingkan penggunaan satu modalitas saja. Dalam konteks pencitraan, multimodal learning memungkinkan penggabungan informasi dari berbagai jenis citra yang memiliki keunggulan yang berbeda-beda [33]. Pemanfaatan citra dalam melakukan *multimodal learning* biasa disebut sebagai *multimodal imaging*. *Multimodal Imaging* pada konteks medis merujuk pada pemanfaatan dua atau lebih modalitas pencitraan medis untuk mendapatkan gambaran yang lebih lengkap dan komprehensif tentang kondisi penyakit yang sedang didiagnosis. Setiap modalitas memiliki karakteristiknya masing-masing, pemanfaatan multimodal imaging memungkinkan modalitas saling melengkapi informasi yang mungkin tidak bisa ditangkap. Berdasarkan tahapan dilakukannya penggabungan informasi, pembelajaran multimodal dibagi menjadi *Data-Level Fusion*, *Feature-Level Fusion*, dan *Decision-Level Fusion*. Pada *Data-Level Fusion* penggabungan informasi dilakukan sebelum proses ekstraksi fitur atau pembelajaran, data mentah dari tiap modalitas dijadikan satu representasi sebelum nantinya digunakan. Dengan *Feature-Level Fusion* proses penggabungan dilakukan setelah dilakukan proses *feature extraction*, menjadikan fitur hasil ekstraksi menjadi lebih kaya akan informasi. Dan pada *Decision-Level Fusion* penggabungan

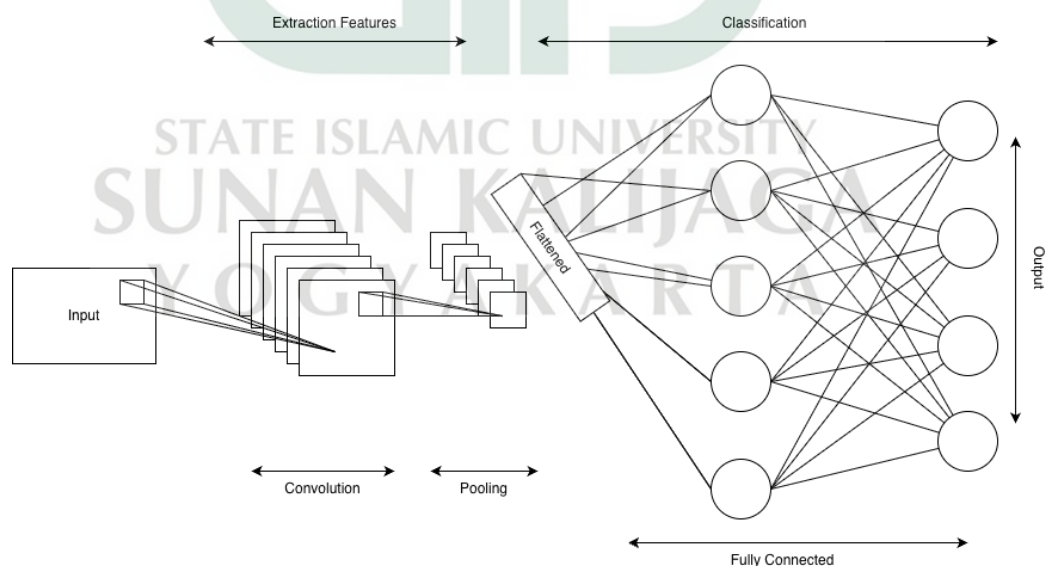
informasi dilakukan ketika masing-masing modalitas telah menghasilkan hasil prediksinya masing-masing, yang nantinya hasil prediksi digabungkan untuk menghasilkan keputusan akhir.

2.2.4. Feature Extraction

Ekstraksi fitur atau *feature extraction* adalah proses untuk mengekstraksi informasi penting dari data mentah sehingga dapat merepresentasikan karakteristik utama dari objek yang dianalisis. Metode ini meminimalisasi penggunaan daya komputasi tanpa menghilangkan fitur yang relevan [34]. Dalam konteks pengolahan citra, *feature extraction* bertujuan untuk menangkap pola visual, tekstur, bentuk, dan struktur spasial yang terdapat pada citra. Representasi fitur yang didapatkan akan digunakan sebagai masukan dalam proses pembelajaran model, seperti klasifikasi dan segmentasi. Biasanya proses feature extraction dilakukan secara manual namun dengan berkembangnya deep learning proses ini bisa dilakukan secara otomatis dengan memanfaatkan arsitektur *Convolutional Neural Network* [31].

2.2.5. Convolutional Neural Network

Convolutional Neural Network (CNN) adalah arsitektur deep learning yang pertama kali diperkenalkan oleh LeCun et al. [35] dirancang untuk memproses data yang memiliki struktur grid, menjadikannya cocok dalam memproses data seperti citra.



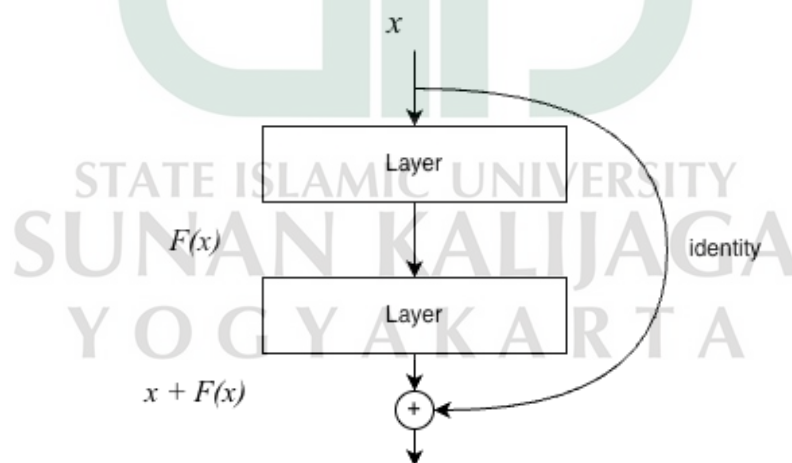
Gambar 2.2 Jaringan Convolutional Neural Network

Secara arsitektural, CNN tersusun atas lapisan konvolusi, lapisan *pooling*, dan lapisan *fully connected*. Proses ini dilakukan secara berurutan melalui lapisan konvolusi yang

bertugas dalam proses mengekstraksi fitur dari citra untuk menghasilkan *feature map* yang merepresentasikan pola tertentu pada citra. Pada lapisan *pooling* dimensi spasial dari *feature map* yang dihasilkan oleh lapisan konvolusi akan direduksi, sehingga mengurangi kompleksitas komputasi sekaligus membuat representasi fitur lebih tidak sensitif terhadap pergeseran kecil pada citra (*translation invariance*). Pada lapisan *fully connected* seluruh fitur yang telah diekstraksi oleh lapisan-lapisan sebelumnya diintegrasikan untuk menghasilkan keputusan klasifikasi akhir.

2.2.6. Residual Neural Network

Residual Neural Network (ResNet) merupakan salah satu arsitektur *Convolutional Neural Network* (CNN) yang dirancang untuk mengatasi permasalahan *vanishing gradient* pada *deep learning*. Permasalahan ini sering menyebabkan penurunan performa seiring bertambahnya jumlah lapisan pada jaringan *deep learning*. Pada arsitektur ResNet terdapat *residual block* yang mengimplementasikan konsep *residual learning*. Pada setiap *residual block*, input akan diproses melalui beberapa lapisan konvolusi untuk menghasilkan fungsi residual, kemudian hasil tersebut dijumlahkan dengan input awal melalui mekanisme *skip connection*. Hal ini mengubah proses pembelajaran dari pemetaan langsung menjadi pembelajaran residual, membantu model untuk dilatih tanpa mengalami penurunan performa.



Gambar 2.3 *Residual Block* pada *Residual Neural Network*

2.2.7. Dempster Shafer Theory

Dempster-Shafer Theory (DST) atau dikenal dengan Teori *Belief Functions* merupakan kerangka matematis yang digunakan untuk merepresentasikan dan menggabungkan ketidakpastian dari berbagai sumber informasi [36]. Berbeda dengan

teori probabilitas klasik, DST memungkinkan representasi ketidakpastian tanpa harus menetapkan probabilitas pasti untuk setiap hipotesis, sehingga dapat memodelkan ketidaktahuan (*ignorance*) secara eksplisit. Konsep dasar dalam DST adalah *frame of discernment* (Ω), yaitu himpunan lengkap dari semua kemungkinan hipotesis yang saling eksklusif dalam suatu permasalahan. Dari *frame of discernment*, dibentuk power set 2^Ω yang mencakup seluruh kemungkinan himpunan bagian, termasuk himpunan kosong ($\emptyset$) dan himpunan penuh (Ω) [15]. *Frame of discernment* ini bisa digambarkan sebagai berikut:

$$\Omega = \{\omega_1, \dots, \omega_M\} \quad (1)$$

Setiap sumber informasi atau modalitas dalam DST direpresentasikan menggunakan *basic belief assignment* (BBA) atau *mass function* ($m(A)$), yaitu sebuah fungsi yang mengalokasikan nilai kepercayaan pada setiap elemen dalam *frame of discernment* (1). Nilai *mass function* untuk setiap himpunan merepresentasikan seberapa besar kepercayaan yang secara langsung diberikan kepada himpunan berdasarkan bukti yang tersedia, dengan syarat bahwa jumlah seluruh nilai *mass function* harus sama dengan satu dan himpunan kosong sama dengan nol [15, 37]. Dengan tiap subset dari $A \subseteq \Omega$ yang memiliki nilai *mass function* $m(A) > 0$ disebut sebagai *focal element*, dan nilai *mass function* yang dimiliki oleh $m(\Omega)$ menggambarkan *degree of ignorance* atau ketidaktahuan, kondisi dimana bukti belum mencukupi [16].

$$\sum_{A \subseteq \Omega} m(\emptyset) = 0 \quad (2)$$

$$\sum_{A \subseteq \Omega} m(A) = 1$$

Terasosiasi dengan *mass function* terdapat fungsi *belief* dan *plausibility* yang keduanya digunakan dalam mengukur kepercayaan ini, kedua fungsi ini ditunjukkan pada Persamaan (3) dan (4) secara berurutan. Pada suatu *frame of discernment* dimana A merupakan *focal element* ($A \subseteq \Omega$), ukuran *belief* (lihat Persamaan 3) merupakan ukuran total keyakinan terhadap suatu hipotesis A , dengan *plausibility* (lihat Persamaan 4) sebagai ukuran kurangnya kepercayaan yang diberikan pada komplemen dari hipotesis A [16], [37]. Ada pula batasan yang diberikan pada nilai *plausibility* yang hanya diterapkan pada *singleton* ($\omega \in \Omega$) yang disebut sebagai fungsi kontur, disajikan pada Persamaan (5).

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (3)$$

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) = 1 - Bel(\bar{A}) \quad (4)$$

$$pl(\omega) = Pl(\{\omega\}) \quad (5)$$

Salah satu keunggulan utama teori ini adalah kemampuannya menggabungkan informasi dari beberapa sumber melalui aturan kombinasi Dempster (*Dempster's rule of combination*). Aturan ini menggabungkan mass functions dari m_1 dan m_2 yang berasal dari sumber yang independen menjadi satu mass function gabungan [15, 37]. *Dempster's rule of combination* dirumuskan seperti pada Persamaan (6).

$$(m_1 \oplus m_2)(A) = \frac{1}{1 - k} \sum_{(B \cap C = A)} m_1(B)m_2(C) \quad (6)$$

Nilai K menunjukkan nilai konflik atau ketidaksesuaian akan kedua informasi yang digabungkan yang dapat diformulasikan seperti pada Persamaan (7).

$$k = \sum_{(B \cap C = \emptyset)} m_1(B)m_2(C) \quad (7)$$

2.2.8. Feature Extraction Evidential Neural Network

Evidential Neural Network (ENN) merupakan pendekatan klasifikasi yang mengintegrasikan kerangka kerja Dempster-Shafer Theory (DST) ke dalam arsitektur jaringan saraf tiruan (*Neural Network*). ENN pertama kali diperkenalkan oleh Dencœux [37] sebagai upaya untuk menggabungkan kemampuan *Neural Network* dalam mengekstraksi pola dari data dengan kemampuan DST dalam merepresentasikan dan mengelola ketidakpastian secara eksplisit. Berbeda dengan *Neural Network* konvensional yang menghasilkan keluaran berupa distribusi probabilitas melalui lapisan softmax, ENN menghasilkan keluaran dalam bentuk *mass functions* yang mampu merepresentasikan tidak hanya keyakinan terhadap suatu kelas, tetapi juga

ketidakpastian yang melekat pada setiap keputusan klasifikasi yang dihasilkan [15, 37].

Secara arsitektur dasar dari ENN terbentuk dari empat lapisan. Lapisan pertama ialah vektor input, bekerja sebagai penerima input dan tidak melakukan perhitungan apapun. Pada arsitektur dasar ENN, lapisan tersembunyi biasanya terdiri dari dua lapisan (L1, L2), tiap lapisan memiliki tugasnya masing-masing. Lapisan tersembunyi (L1) bekerja dengan menghitung jarak antara vector fitur input x dengan prototipe p^i atau pola referensi yang bertindak sebagai item bukti dengan persamaan jarak $d^i = ||x - p^i||$, $i=1, \dots, n$. Jarak ini nantinya digunakan sebagai nilai aktivasi dimana semakin dekat input dengan prototipe, semakin kuat dukungan bukti yang diberikan. Nilai aktivasi ini ditunjukkan pada persamaan (8).

$$s^i = \alpha^i \exp(-\gamma^i (d^i)^2) \quad (8)$$

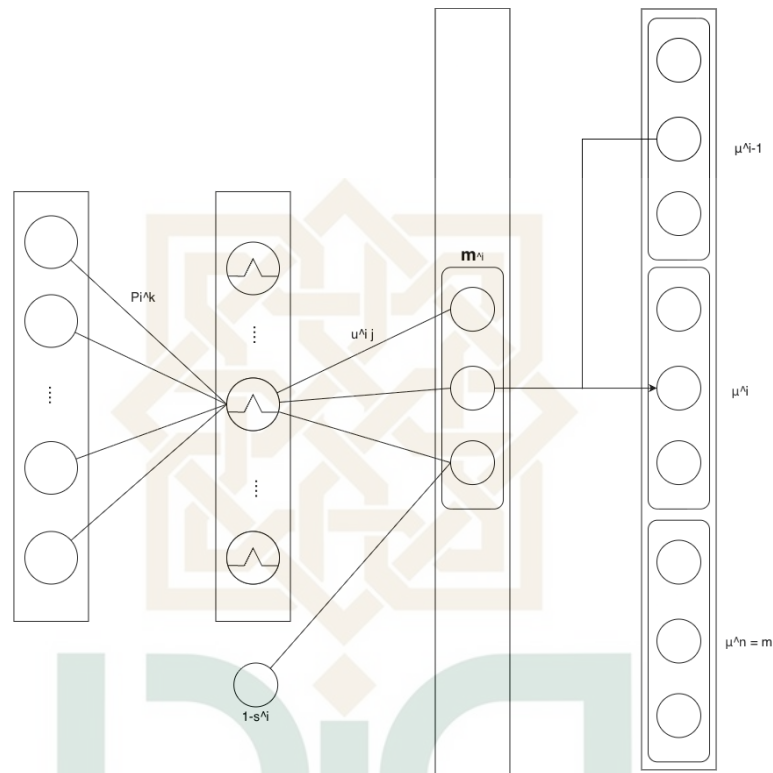
di mana $\gamma^i \in R$ dan $\alpha^i \in [0,1]$ adalah parameter yang diasosiasikan dengan prototipe p^i , sehingga s^i merupakan fungsi yang nilainya menurun dari 1 menuju 0. Fungsi ini memastikan bahwa semakin jauh vektor input dari suatu prototipe; Semakin besar massa yang dialokasikan ke Ω sebagai representasi ketidaktahuan total, dan semakin kecil massa yang dialokasikan ke kelas-kelas spesifik

Lapisan tersembunyi (L2) bekerja merubah nilai aktifasi yang didapatkan pada layer L1 menjadi nilai *Basic Belief Assignment* (BBA) atau *mass function*, pada tahap ini kepercayaan didistribusikan ke kelas-kelas spesifik berdasarkan keanggotaan prototipe ditunjukkan pada persamaan (9). Dimana s^i adalah nilai aktivasi yang didapatkan dari lapisan tersembunyi L1, dan u_j^i merupakan derajat keanggotaannya. Dengan nilai sisa kepercayaan yang tidak alokasi akan diberikan pada himpunan *universe* (Ω) untuk merepresentasikan ketidaktahuan (*ignorance*) direpresentasikan pada persamaan (10).

$$m^i(\{\omega_j\}) = u_j^i s^i, \quad j = 1, \dots, M \quad (9)$$

$$m^i(\Omega) = 1 - s^i \quad (10)$$

Pada lapisan terakhir atau lapisan output (L3), seluruh BBA dari prototipe digabungkan menggunakan *Dempster Rule of Combination* (Persamaan 6 - 7) untuk menghasilkan satu nilai BBA global merangkup semua bukti yang tersedia. Dengan hasil akhir memberikan deskripsi ketidakpastian yang memungkinkan untuk menjelaskan hasil dari klasifikasi yang diberikan.



Gambar 2.4 Arsitektur *Evidential Neural Network*

2.2.9. Evaluasi Model

Evaluasi kinerja model klasifikasi berbasis *Evidential Neural Network* (ENN) memerlukan metrik yang dirancang khusus untuk mengakomodasi karakteristik keluaran model berupa *mass functions*. Karena keluaran model berbasis belief functions berbeda dari model probabilistik, maka akan digunakan metrik evaluasi yaitu E_{pl} untuk pengukuran error prediksi dan U_{65} untuk memberikan *reward* terhadap prediksi yang mengandung kebenaran dengan indikator pada *ground truth*. Kedua metrik ini digunakan untuk menilai kualitas keputusan model berbasis *belief functions* secara lebih komprehensif [38].

a. Discrepancy measure

Untuk melakukan pengukuran error pada model yang diterapkan dalam penelitian ini, maka digunakan *discrepancy measure* yang dinotasikan dengan E_{pl}

yang merupakan ukuran ketidaksesuaian hasil prediksi dengan *ground truth*. Semakin kecil nilai E_{pl} semakin baik karena mengindikasikan tingkat kesalahan prediksi kecil dan performa semakin akurat. Metrik ini bekerja dengan menghitung selisih kuadrat antara fungsi kontur pl dan nilai kebenaran sebenarnya (*ground truth*) dari objek yang diuji seperti yang terlihat pada persamaan 11.

$$E_{pl} = \sum_{i=1}^n \sum_{k=1}^K (pl_{i(\omega_k)} - \delta_{i,k})^2 \quad (11)$$

di mana pl adalah fungsi kontur yang merupakan batasan *plausibility* pada nilai kelas individu dalam *frame of discrement* (Ω). Sedangkan n merupakan jumlah objek dalam set pembelajaran (learning set), K adalah jumlah kelas atau elemen atau kategori dalam semesta pembicaraan (Ω). Sedangkan *ground truth* dinotasikan dengan $\delta_{i,k}$ sebagai fungsi indikator kebenaran dan $pl_{i(\omega_k)}$ untuk objek ke-i pada kelas ω_k .

b. Utility Predictive Accuracy

Utility Predictive Accuracy atau U_{65} adalah metrik utilitas yang dikembangkan oleh Zaffalon et al. [39] untuk menilai kualitas dari keputusan parsial, khususnya dalam menghadapi informasi yang tidak pasti atau tidak presisi. Rata-rata nilai U_{65} yang lebih tinggi dianggap sebagai indikator performa yang lebih baik, karena menunjukkan kemampuan sistem dalam memberikan keputusan yang lebih berguna dan dapat diandalkan. Dirumuskan sebagai berikut:

$$U_{65} = 1.6x - 0,6x^2 \quad (12)$$

Di mana variabel x adalah akurasi terdiskon dari keputusan d, yang dihitung dengan rumus:

$$x = \frac{\mathbb{I}(\omega \in y)}{|y|} \quad (13)$$

$\mathbb{I}$: Fungsi indikator

ω : Kelas yang sebenarnya (true class)

$|y|$: Jumlah elemen dalam set keputusan y

c. K-Fold Cross Validation

K-Fold cross validation merupakan teknik evaluasi model yang digunakan untuk memastikan konsistensi dan generalitas hasil evaluasi terhadap berbagai subset data yang berbeda. Dengan cara ini, dataset keseluruhan dibagi secara acak menjadi K bagian (*fold*) yang berukuran sama. Pada setiap iterasi, satu fold digunakan sebagai data pengujian (*test set*) sementara K-1 *fold* sisanya digunakan sebagai data pelatihan (*training set*), sehingga setiap *fold* mendapat kesempatan menjadi data pengujian tepat satu kali [35]. Proses ini diulang sebanyak K kali hingga seluruh fold telah digunakan sebagai data pengujian, dan hasil evaluasi akhir diperoleh dengan merata-ratakan nilai metrik dari seluruh K iterasi tersebut. Keunggulan utama K-Fold cross validation dibandingkan pembagian data tunggal (*single split*) adalah kemampuannya dalam menghasilkan estimasi performa model yang lebih andal dan tidak bergantung pada satu pembagian data tertentu, sehingga mengurangi risiko *overfitting* maupun *underfitting* akibat pembagian data yang tidak representatif [35].

BAB III

METODE PENELITIAN

3.1. Alat dan Bahan Penelitian

3.1.1. Alat yang Digunakan

Dalam proses penelitian terdapat beberapa alat yang digunakan terbagi menjadi perangkat keras (*Hardware*) dan perangkat lunak (*Software*), diantaranya sebagai berikut:

1. Perangkat Keras (*Hardware*)

a. Laptop

Berikut adalah spesifikasi yang digunakan penulis:

- Prosesor: M1
- Penyimpanan: Solid State Drive 256 GB
- Memori: RAM 8 GB

2. Perangkat Lunak (*Software*)

- a. Sistem Operasi: MacOS Tahoe 26.1
- b. Bahasa pemrograman: Python
- c. Code Editor: Visual Studio Code
- d. Platform komputasi awan: Google Collaboratory
- e. Platform dataset: Kaggle
- f. Web Browser: Brave Browser
- g. Library: Numpy, Torch, Pandas, scikit-learn, dan Matplotlib

3.1.2. Dataset

a. Chest X-Ray Images (Pneumonia)

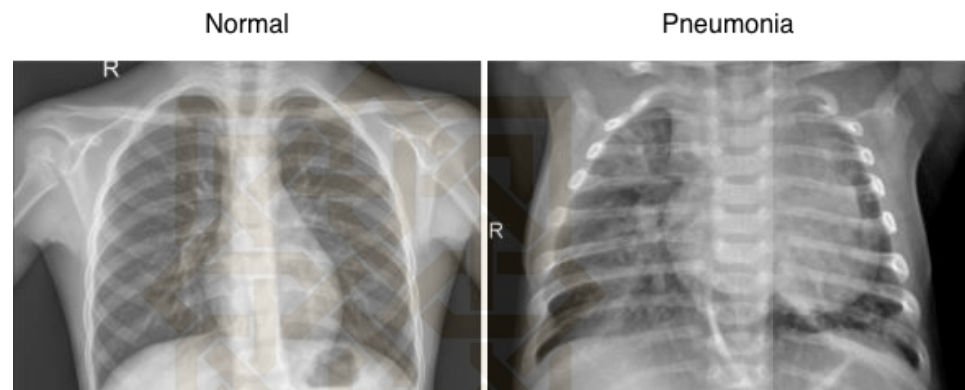
Sumber: Laman Kaggle

Karakteristik:

- Jenis Data: Citra Medis Chest X-Ray
- Total Data: 5856
- Jumlah Kelas: 2 Kelas (Normal dan Pneumonia)
- Distribusi Kelas:
 - Normal : 4273 (73%)
 - Pneumonia: 1583 (27%)

Deskripsi: Dataset ini berupa kumpulan citra hasil pemeriksaan X-ray pada bagian paru-paru pasien. Dataset terdiri pada citra X-ray paru-paru yang terindikasi pneumonia dan citra X-ray paru-paru normal. Dataset yang digunakan merupakan dataset publik berasal dari website Kaggle. Dataset ini memiliki jumlah kelas yang kurang seimbang dengan mayoritas data merupakan bagian dari kelas Normal.

Contoh Citra:



Gambar 3.1 Sample Citra CXR

b. Pneumonia Balanced CT-Scan

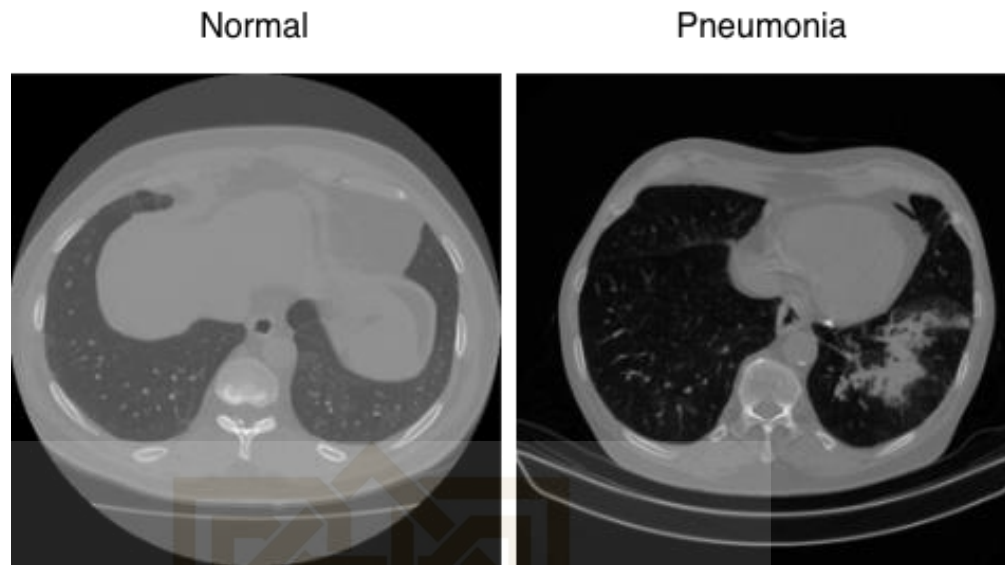
Sumber: Laman Kaggle

Karakteristik:

- Jenis Data: Citra Medis Chest X-Ray
- Total Data: 5236
- Jumlah Kelas: 2 Kelas (Normal dan Pneumonia)
- Distribusi Kelas:
 - Normal : 2618 (50%)
 - Pneumonia: 2618 (50%)

Dataset ini berupa kumpulan citra hasil pemeriksaan CT-Scan pada bagian paru-paru pasien. Dataset terdiri pada citra CT-Scan paru-paru yang terindikasi pneumonia dan citra CT-Scan paru-paru normal. Dataset yang digunakan merupakan dataset publik berasal dari website Kaggle. Dataset ini memiliki jumlah kelas yang seimbang antara kedua kelasnya.

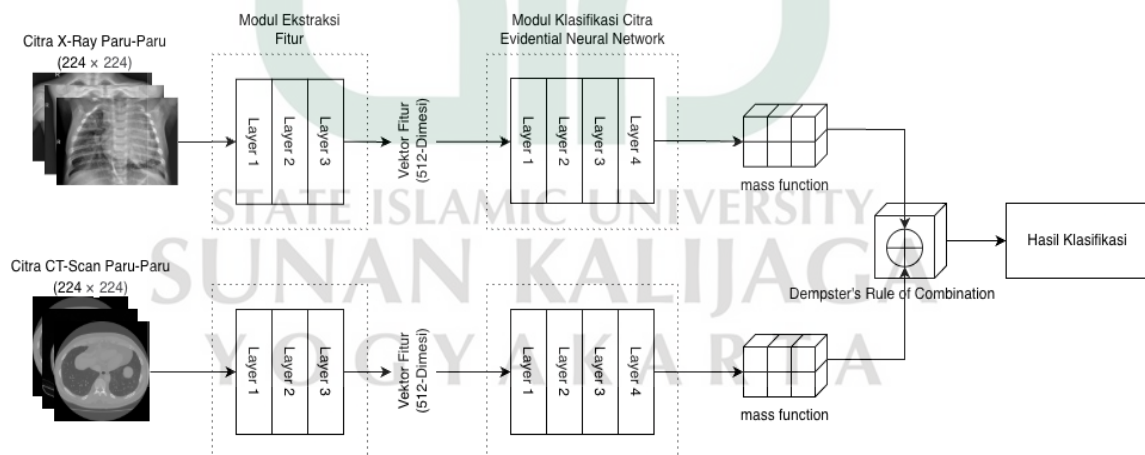
Contoh Citra:



Gambar 3.2 Sample Citra CT-Scan

3.2. Metode Penelitian

Melalui metode eksperimen dengan pendekatan kuantitatif, yang bertujuan untuk membangun serta mengevaluasi model *Evidential Neural Network* dalam mengidentifikasi paru-paru yang terindikasi penyakit pneumonia dengan memanfaatkan multimodalitas citra medis. Penelitian ini dilakukan melalui beberapa tahapan sistematis dari pengumpulan data hingga evaluasi model.



Gambar 3.3 Proses Tahapan Penelitian

3.2.1. Pengumpulan Data

Pada penelitian ini dataset yang digunakan merupakan dataset publik yang diperoleh dari platform Kaggle. Penelitian ini menggunakan dua dataset yang berbeda, dataset Chest X-Ray Images (Pneumonia) dan Pneumonia Balanced CT, penggunaan

kedua dataset ini merepresentasikan dua modalitas dalam pencitraan penyakit pneumonia. Kedua dataset masing-masing terdiri dari dua kelas, yaitu citra paru-paru normal dan citra paru-paru yang terindikasi penyakit pneumonia.

3.2.2. Preprocessing Data

Tahapan *preprocessing* atau pra-pemrosesan, merupakan tahapan yang dilakukan untuk mempersiapkan data sebelum digunakan dalam proses pelatihan model. Kedua dataset dalam penelitian ini akan melalui tahapan proses pra-pemrosesan yang sama, proses ini meliputi:

1. Penyeimbangan jumlah data pada tiap kelas
2. Penyesuaian ukuran citra ke ukuran yang seragam,
3. Normalisasi nilai piksel citra
4. Pembagian data menjadi data latih, data validasi, dan data test

Melalui proses penyeimbangan data, model tidak memiliki bias pada kelas tertentu. Penyesuaian ukuran citra dilakukan agar sesuai dengan dimensi input yang disyaratkan oleh arsitektur model, dengan normalisasi nilai piksel dilakukan untuk menyeragamkan skala data sehingga pelatihan model dapat berjalan lebih stabil dan efisien. Semua tahapan ini bertujuan untuk meningkatkan kualitas data dan membantu model dalam proses pembelajaran.

3.2.3. Ekstraksi Fitur

Data yang telah melalui proses *preprocessing* selanjutnya memasuki tahap ekstraksi fitur. Tahap ini bertujuan untuk mengekstraksi representasi fitur yang informatif dan relevan dari citra paru-paru pada masing-masing modalitas yang akan digunakan sebagai masukan dalam proses pelatihan model ENN. Memanfaatkan *model pre-trained* ResNet-34 proses akan dilakukan. Dalam penelitian ini, seluruh parameter ResNet-34 dibekukan pada seluruh lapisan, sehingga ResNet-34 berfungsi sebagai ekstraktor fitur tetap tanpa pembaruan bobot selama proses pelatihan. ResNet-34 digunakan dengan fokus merubah citra paru-paru menjadi representasi numerik yang nantinya digunakan dalam pelatihan model *Evidential Neural Network*.

3.2.4. Pelatihan Model Evidential Neural Network

Pada tahap ini dilakukan pembangunan dan pelatihan model *Evidential Neural Network* (ENN). Model neural network yang dirancang untuk menghasilkan nilai massa

keyakinan yang merepresentasikan kepercayaan dan ketidakpastian. Model ENN dilatih menggunakan fitur yang dihasilkan pada proses ekstraksi fitur. Pelatihan ini dilakukan secara terpisah pada masing-masing modalitas citra. Dengan cara ini setiap modalitas akan menghasilkan nilai massa keyakinan yang merepresentasikan informasi dan ketidakpastian masing-masing modalitas.

3.2.5. Penggabungan Hasil Prediksi Model ENN dari Tiap Modalitas

Mass function yang dihasilkan secara independen oleh model ENN dari masing-masing modalitas selanjutnya digabungkan menggunakan *Dempster's rule of combination*, sebagaimana ditunjukkan pada Persamaan (6-7), untuk menghasilkan satu *mass function* total yang merepresentasikan penggabungan informasi dari kedua modalitas. *Mass function* total yang dihasilkan merepresentasikan integrasi keyakinan dan ketidakpastian dari kedua modalitas menjadi dasar dalam pengambilan keputusan klasifikasi akhir.

3.2.6. Evaluasi Model

Kinerja model dievaluasi menggunakan dua metrik evaluasi yang dirancang khusus untuk model berbasis belief functions, yaitu pengukuran *discrepancy* (error) antara hasil prediksi dan *ground truth* menggunakan E_{pl} (lihat Persamaan 11) dan pengukuran *reward* U_{65} (lihat Persamaan 12 - 13). Untuk memastikan konsistensi dan generalitas hasil evaluasi, seluruh proses pelatihan dan evaluasi dilakukan menggunakan metode *Stratified K-Fold cross validation* dengan $K = 10$. Metode *Stratified K-Fold* dipilih untuk memastikan distribusi kelas yang proporsional pada setiap *fold*, sehingga evaluasi tidak dipengaruhi oleh ketidakseimbangan distribusi kelas pada pembagian data tertentu. Pada setiap iterasi, satu *fold* digunakan sebagai data uji sementara sembilan *fold* sisanya digunakan sebagai data latih, dan proses ini diulang sebanyak 10 kali hingga seluruh *fold* telah digunakan sebagai data uji. Hasil evaluasi akhir diperoleh dengan merata-ratakan nilai E_{pl} dan U_{65} dari seluruh 10 iterasi tersebut. Perbandingan performa dilakukan pada tiga konfigurasi model, yaitu model ENN berbasis CXR saja, model ENN berbasis CT-Scan saja, dan model ENN berbasis *multimodal* (fusi CXR dan CT-Scan),

BAB IV

HASIL DAN PEMBAHASAN

4.1. Pembangunan Model

Pada bagian ini, proses pembangunan model klasifikasi pneumonia berbasis multimodal yang mengintegrasikan modalitas CXR dan CT-Scan menggunakan kerangka *Evidential Neural Network* (ENN) berbasis *Dempster-Shafer Theory* akan dijelaskan. Proses pengembangan model dilakukan secara terstruktur dengan dua citra multimodal yaitu CXR dan CT-Scan dan penggunaan ENN dilanjutkan ke tahap penggabungan output dari ENN menggunakan *Dempster's Rule of combination* (Persamaan 6 - 7). Pembangunan model ini mengikuti alur yang disajikan pada Gambar 3.3, dimana pada bab 4 akan difokuskan pada pembahasan setelah dilakukan preprocessing, dikarenakan

4.1.1. Pengumpulan Data

Pada penelitian ini dataset yang digunakan merupakan dataset publik yang diperoleh dari platform Kaggle. Penelitian ini menggunakan dua dataset yang berbeda, dataset Chest X-Ray Images (Pneumonia) dan Pneumonia Balanced CT, penggunaan kedua dataset ini merepresentasikan dua modalitas dalam pencitraan penyakit pneumonia. Semua dataset dipastikan merupakan dataset yang berisi citra diagnosis untuk penyakit pneumonia. Kedua dataset masing-masing terdiri dari dua kelas, kelas normal dan kelas pneumonia. Dataset CXR terdiri dari 5856 citra dengan 4273 merupakan citra kelas normal dan 1583 citra kelas pneumonia. Pada dataset CT-Scan terdiri dari 5236 citra dengan 2618 citra pada kelas normal dan 2618 citra pada kelas pneumonia. Distribusi pada dataset modalitas CXR terlihat memiliki jumlah data lebih banyak pada kelas 'Normal' daripada kelas 'Pneumonia'. Distribusi pada dataset modalitas CT-Scan tiap kelas memiliki jumlah terbagi secara seimbang pada dua kelas.

Tabel 4.1 Distribusi Dataset yang Digunakan

Dataset	Normal	Pneumonia	Total
Chest X-Ray	4273	1583	5856
CT-Scan	2618	2618	5236

4.1.2. Preprocessing Data

Tahapan preprocessing dilakukan untuk mempersiapkan data sebelum digunakan dalam proses pelatihan model. Tahapan ini sangat penting dilakukan

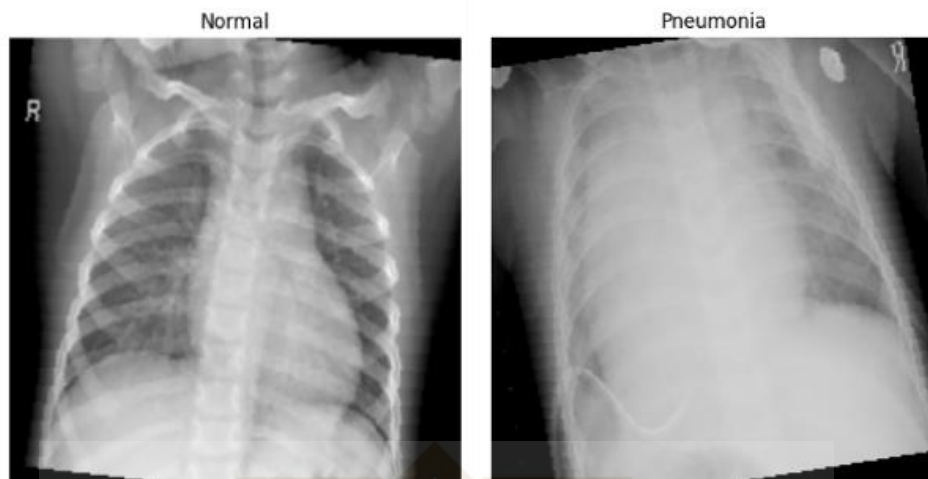
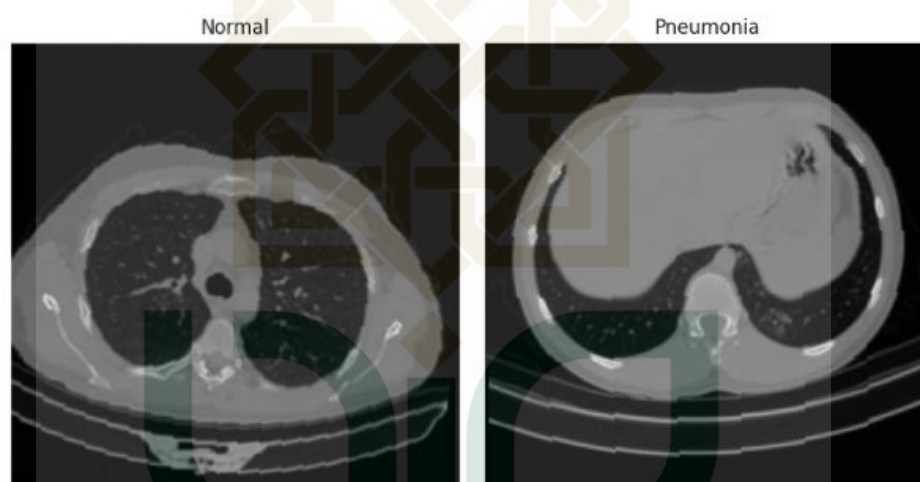
sebelum dataset digunakan dalam pembangunan model, mengingat berdasarkan Tabel 4.1, distribusi kelas pada modalitas CXR tidak seimbang, dengan 4273 citra pada kelas normal dan 1583 citra pada kelas pneumonia. Ketidakseimbangan antar kelas dapat menjadikan model memiliki bias akan kelas mayoritas, sehingga proses penyeimbangan antar kelas ini sangat penting untuk dilakukan. Kedua dataset yang digunakan melalui proses pembersihan data yang meliputi:

Dalam mengatasi permasalahan keterseimbangan kelas dapat digunakan teknik *undersampling*. Teknik ini menyeimbangkan kelas pada dataset dengan mengurangi jumlah data pada kelas mayoritas [40]. Dalam melakukan proses *undersampling* digunakan *random sampling* dengan mengambil sampel secara acak sebanyak 1550 citra dari kelas normal dan 1550 citra dari kelas pneumonia pada dataset CXR, menjadikan total data CXR yang digunakan 3100 citra. Untuk memastikan keseimbangan jumlah data antar modalitas, dataset CT-Scan juga disesuaikan menjadi 1550 citra per kelas melalui proses *random sampling* yang sama, sehingga total data CT-Scan yang digunakan juga menjadi 3100 citra. Detail distribusi setelah dilakukannya penyeimbangan dapat dilihat pada tabel 4.1.

Tabel 4.2 Distribusi Dataset Setelah Penyeimbangan

Dataset	Normal	Pneumonia	Total
Chest X-Ray	1550	1550	3100
CT-Scan	1550	1550	3100

Setelah proses penyeimbangan ini semua citra melalui proses penyamaan ukuran menjadi 224x224 piksel sesuai dengan ukuran yang dibutuhkan oleh arsitektur ResNet-34 yang akan digunakan pada tahap ekstraksi fitur. Normalisasi nilai piksel kemudian dilakukan menggunakan nilai rata-rata (mean) = [0,485, 0,456, 0,406] dan standar deviasi (std) = [0,229, 0,224, 0,225] untuk masing-masing kanal RGB, sesuai dengan standar normalisasi dataset ImageNet yang digunakan dalam bobot pre-trained ResNet-34. Hasil dari proses perubahan ukuran dan normalisasi citra ditunjukkan pada Gambar 4.1 dan 4.2.

Gambar 4.1 Hasil *Preprocessing* Citra CXRGambar 4.2 Hasil *Preprocessing* Citra CT-Scan

Proses *preprocessing* diakhiri dengan dilakukannya pembagian data. Data dibagi dengan rasio 80:10:10, dimana 80% akan digunakan sebagai data latih model, 10% sebagai data validasi, dan sisa 10% digunakan sebagai test dari latihan yang dilakukan. Dalam proses pembagian ini dipastikan bahwa tiap bagian memiliki 50% data citra pneumonia dan 50% sisanya merupakan data citra normal. Detail pembagian data pada masing-masing modalitas disajikan pada Tabel 4.3.

Tabel 4.3 Pembagian data setelah preprocessing

Dataset	Data Latih (80%)	Data Validasi (10%)	Data Test (10%)	Total
Chest X-Ray	2480	310	310	3100
CT-Scan	2480	310	310	3100

4.1.3. Feature Extraction

Proses ekstraksi fitur dilakukan memanfaatkan arsitektur ResNet-34. Dalam pemanfaatan arsitektur ini seluruh parameter dibekukan, sehingga ResNet-34 berkerja sebagai ekstraktor fungsi tetap tanpa ada perubahan bobot selama proses pelatihan. Proses ekstraksi fitur ini dilakukan secara independen untuk masing-masing modalitas. Setiap citra sudah digunakan dipastikan sudah melalui proses *preprocessing* untuk memastikan ukuran yang seragam berukuran $224 \times 224 \times 3$ piksel diproses melalui seluruh lapisan konvolusi ResNet-34 secara heirarkis. Melalui ekstraksi fitur tingkat paling rendah seperti bagian tepi dan tekstur pada lapisan awal, hingga fitur tinggi yang merepresentasikan pola pneumonia pada lapisan yang lebih dalam. Melalui semua proses ini ResNet-34 akan menghasilkan vektor fitur yang berdimensi 512 untuk semua citra yang dimasukan dan yang nantinya digunakan dalam pelatihan model ENN.

Hasil proses ekstraksi fitur menggunakan ResNet-34 menghasilkan representasi fitur yang siap digunakan sebagai masukan model ENN. Pada modalitas CXR, proses ekstraksi menghasilkan matriks fitur berukuran (2.480×512) , di mana 2.480 merupakan jumlah sampel data latih dan 512 merupakan dimensi vektor fitur yang dihasilkan oleh lapisan *global average pooling* ResNet-34 untuk setiap sampel. Setiap baris dalam matriks fitur merepresentasikan satu citra paru-paru dalam bentuk vektor numerik berdimensi 512 yang mengkodekan karakteristik visual citra tersebut secara kompak. Label kelas yang bersesuaian disimpan dalam vektor label berukuran (2.480,) dengan nilai 0 untuk kelas normal dan 1 untuk kelas pneumonia, sesuai dengan pemetaan label yang telah dilakukan pada tahap preprocessing. Detail dimensi fitur hasil ekstraksi pada masing-masing modalitas disajikan pada Tabel 4.4.

Tabel 4.4 Dimensi Hasil Ekstraksi Fitur

Dataset	Data Latih (80%)	Data Validasi (10%)	Data Test (10%)	Dimensi Feature Extraction
Chest X-Ray	2480	310	310	512
CT-Scan	2480	310	310	512

Setelah proses penyeimbangan ini semua citra melalui proses penyamaan ukuran menjadi 224×224 piksel sesuai dengan ukuran yang dibutuhkan oleh arsitektur

ResNet-34 yang akan digunakan pada tahap ekstraksi fitur. Normalisasi nilai piksel kemudian dilakukan menggunakan nilai rata-rata (mean) = [0,485, 0,456, 0,406] dan standar deviasi (std) = [0,229, 0,224, 0,225] untuk masing-masing kanal RGB, sesuai dengan standar normalisasi dataset ImageNet yang digunakan dalam bobot pre-trained ResNet-34. Proses preprocessing diakhiri dengan dilakukannya pembagian data. Data dibagi dengan rasio 80:10:10, dimana 80% akan digunakan sebagai data latih model, 10% sebagai data validasi, dan sisa 10% digunakan sebagai test dari latihan yang dilakukan. Dalam proses pembagian ini dipastikan bahwa tiap bagian memiliki 50% data citra pneumonia dan 50% sisanya merupakan data citra normal. Detail pembagian data pada masing-masing modalitas disajikan pada Tabel 4.3.

4.1.4. Pelatihan Model ENN

Pelatihan model ENN dilakukan menggunakan sejumlah *hyperparameter* yang telah ditentukan disajikan pada Tabel 4.5. Konfigurasi *hyperparameter* yang digunakan merupakan nilai *default* dari model ENN. Memanfaatkan algoritma optimasi Harris, yang mengoptimasi secara adaptif per-parameter. Nilai *learning rate* sebesar 0,1 dipilih sebagai nilai awal yang umum digunakan pada algoritma optimasi adaptif, karena memberikan keseimbangan antara kecepatan konvergensi dan stabilitas pelatihan. Jumlah iterasi maksimal sebesar 250 ditentukan berdasarkan observasi bahwa model telah menunjukkan konvergensi yang stabil sebelum iterasi ke-250, sebagaimana ditunjukkan oleh nilai *training loss* akhir yang rendah pada kedua modalitas. *Batch size* sebesar 64 dipilih karena merupakan nilai yang umum digunakan dalam pelatihan model berbasis CNN dan memberikan keseimbangan antara efisiensi komputasi dan stabilitas estimasi gradien.

Tabel 4.5 *Hyperparameter Model Evidential Neural Network*

Model	Parameter	Nilai
<i>Evidential Neural Network</i>	Optimizer	Harris
	Learning Rate	0,1
	Epoch	250
	Gain Min (Threshold)	1e-3
	Regularization (μ)	1e-3
	Lambda (λ)	1.0/M
	nproto	25

Semua hyperparameter ini digunakan pada setiap modalitas. Model ENN dilatih menggunakan fitur yang telah diekstraksi pada tahapan sebelumnya, di mana lapisan tersembunyi dalam arsitektur ENN bertugas mengonversi representasi fitur tersebut menjadi *mass functions* yang merepresentasikan keyakinan (*belief*) dan ketidakpastian (*uncertainty*) model terhadap setiap kelas secara eksplisit. Hasil pelatihan menunjukkan bahwa model ENN pada kedua modalitas berhasil konvergen dengan baik, terlihat dari nilai *training gain* yang relatif kecil pada kedua modalitas. Model ENN pada modalitas CXR mencapai nilai *training loss* sebesar 0,079681 pada iterasi ke-250, sementara model ENN pada modalitas CT-Scan mencapai nilai *training loss* yang lebih rendah yaitu sebesar 0,071668 pada iterasi yang sama. Nilai *training loss* yang kecil ini menandakan keberhasilan model dalam meminimalisasi kesalahan prediksinya. Dengan nilai *training loss* pada modalitas CT-Scan lebih rendah dibandingkan CXR mengindikasikan bahwa model ENN pada modalitas CT-Scan mampu mempelajari pola klasifikasi pneumonia dengan lebih baik, yang kemungkinan disebabkan oleh kualitas dan detail informasi yang lebih tinggi pada citra CT-Scan dibandingkan CXR. Detail nilai *training gain* dan nilai *loss* akhir per modalitas disajikan pada Tabel 4.6.

Tabel 4.6 Nilai *Gain* dan Nilai *Loss* dari Pelatihan Model ENN

Modalitas	Nilai <i>Training Gain</i>	Nilai <i>Loss</i>
Chest X-Ray	0,000665	0,079681
CT-Scan	0,000806	0,071668

Setelah pelatihan model ENN pada masing-masing modalitas selesai dilakukan, *mass function* yang dihasilkan secara independen akan dihasilkan. Ditunjukan pada Tabel 4.7 dan 4.8 10 hasil *mass function* dari masing-masing modalitas. Melalui *mass function* ini kita dapat melihat bagaimana model mengklasifikasi citra dari tiap modalitas.

Tabel 4.7 Nilai *Mass Function* pada Modalitas X-Ray

No	$m(\{\text{Normal}\})$	$m(\{\text{Pneumonia}\})$	$m(\Omega)$
1	0,963053	0,035809	0,001138
2	0,217014	0,780551	0,002435
3	0,894691	0,102975	0,002335

No	$m(\{\text{Normal}\})$	$m(\{\text{Pneumonia}\})$	$m(\Omega)$
4	0,969179	0,029419	0,001402
5	0,806240	0,184987	0,008773
6	0,891807	0,100175	0,008019
7	0,531113	0,427445	0,041442
8	0,353618	0,645350	0,001032
9	0,019994	0,978957	0,001048
10	0,023903	0,975530	0,000567

Tabel 4.8 Nilai *Mass Function* pada Modalitas CT-Scan

No	$m(\{\text{Normal}\})$	$m(\{\text{Pneumonia}\})$	$m(\Omega)$
1	0,934570	0,064379	0,001051
2	0,434404	0,560749	0,004847
3	0,931088	0,067306	0,001606
4	0,669595	0,319616	0,010789
5	0,908232	0,089750	0,002018
6	0,693539	0,293294	0,013167
7	0,036513	0,961660	0,001827
8	0,065047	0,932060	0,002893
9	0,291443	0,703296	0,005261
10	0,348725	0,649605	0,001670

4.1.5. Analisis Mass Function pada Tiap Modalitas

Untuk memberikan gambaran yang lebih jelas tentang *mass function* yang dihasilkan oleh model ENN pada masing-masing modalitas, Tabel 4.7 dan Tabel 4.8 memberikan 10 contoh nilai *mass function* yang dihasilkan pada data uji dari modalitas CXR dan CT-Scan. Setiap baris dalam tabel merepresentasikan satu sampel citra paru-paru dengan tiga komponen *mass function* dihasilkan, yaitu $m(\{\text{Normal}\})$ yang merepresentasikan keyakinan model terhadap kelas normal, $m(\{\text{Pneumonia}\})$ yang merepresentasikan keyakinan model terhadap kelas pneumonia, dan $m(\Omega)$ yang merepresentasikan ketidakpastian model ketika kurangnya informasi untuk menentukan kelas secara pasti.

Berdasarkan Tabel 4.7, model ENN pada modalitas CXR secara konsisten menghasilkan keputusan yang tegas pada seluruh sampel yang ada, ditunjukkan dengan

salah satu dari $m(\{\text{Normal}\})$ atau $m(\{\text{Pneumonia}\})$ selalu mendominasi dengan nilai yang jauh lebih dibandingkan komponen lainnya. Telihat nilai terbesar pada Tabel 4.7 terdapat pada sampel ke-sembilan dengan kelas $m(\{\text{Pneumonia}\})$ memiliki nilai 0,9790, sampel ini mengindikasikan keyakinan yang sangat tinggi terhadap kelas pneumonia dibandingkan yang lainnya. Sebaliknya, pada kelas $m(\{\text{Normal}\})$ nilai terbesar terdapat sampel keempat dengan nilai 0,9692, sampel ini mengindikasikan keyakinan pada kelas normal yang sangat kuat. Nilai $m(\Omega)$ pada modalitas CXR berada pada rentang 0,0005 hingga 0,0415 menunjukkan bahwa model memiliki tingkat ketidakpastian yang relatif kecil. Namun masih terdapat beberapa sampel seperti pada normor tujuh ($m(\Omega) = 0,0415$) yang menunjukkan nilai ketidakpastian cukup tinggi diantara seluruh sampel CXR, pada sampel ini ketidakpastian terlihat jelas dengan nilai *mass function* pada kelas ($m(\{\text{Normal}\}) = 0,5311$) dan ($m(\{\text{Pneumonia}\}) = 0,4275$) memiliki nilai yang hampir sama antara kedua kelasnya, hal ini menunjukkan model kurang yakin pada kelas yang dimiliki sampel ketujuh.

Pada modalitas CT-Scan, disajikan pada Tabel 4.8 model ENN menunjukkan karakteristik yang serupa namun dengan tingkat keyakinan yang lebih tinggi dan ketidakpastian yang lebih rendah dibandingkan modalitas CXR. Hal ini tercermin dari nilai $m(\Omega)$ yang berada pada rentang 0,0011 hingga 0,0131, lebih kecil dibandingkan rentang yang dimiliki modalitas CXR. Sampel ketujuh pada Tabel 4.8 menunjukkan nilai tertinggi pada $m(\{\text{Pneumonia}\}) = 0,9617$ dan $m(\{\text{Normal}\}) = 0,0365$, mengindikasikan keyakinan yang tinggi bahwa sampel termasuk pada kelas pneumonia. Dengan sampel satu, tiga, lima memberikan keyakinan pada kelas sebaliknya, pada kelas normal. Pada sampel kedua terlihat ketidakyakinan pada hasil model dengan dua kelas memiliki nilai ($m(\{\text{Normal}\}) = 0,4344$, $m(\{\text{Pneumonia}\}) = 0,5608$) yang hampir sama. Berbeda dengan ketidakyakinan yang terjadi pada modalitas CXR, terlihat bahwa nilai ketidakpastian pada sampel ini bukanlah yang tertinggi dari sepuluh sampel yang digunakan.

Secara keseluruhan, kedua model memiliki keyakinan yang kuat akan hasil yang diberikan, namun melihat nilai ketidakpastian yang diberikan dari kedua modalitas, CT-Scan menunjukkan keyakinan yang lebih baik dibandingkan modalitas CXR. Dengan rerata $m(\Omega)$ yang berada pada nilai 0,0045, dibandingkan CXR yang berada pada nilai 0,0068, namun dapat dilihat bahwa nilai dari kelas CXR memiliki nilai lebih tinggi dibandingkan modalitas CT-Scan, dilihat dari nilai yang diberikannya. Perbedaan karakteristik *mass function* antara kedua modalitas ini memperkuat urgensi pendekatan

multimodal, karena penggabungan informasi dari kedua modalitas menggunakan *Dempster's rule of combination* (Persamaan 6 - 7) diharapkan dapat menghasilkan *mass function* yang lebih yakin dan akurat dibandingkan masing-masing modalitas secara individual, sebagaimana akan dibahas pada bagian selanjutnya.

4.2. Penggabungan Hasil Prediksi

Tabel 4.9 menyajikan 10 contoh nilai *mass function* yang dihasilkan oleh model multimodal setelah proses penggabungan *mass function* dari dua modalitas menggunakan *Dempster's rule of combination* (Persamaan 6 - 7) pada data uji yang sama dengan sampel yang disajikan pada Tabel 4.7 dan Tabel 4.8. Melalui ketiga tabel tersebut dapat memberikan Gambaran tentang bagaimana penggabungan informasi dari kedua modalitas menggunakan kerangka DST berdampak pada klitas dan ketegasan *mass function* yang dihasilkan.

Tabel 4.9 Nilai *Mass Function* pada Model Multimodal

No	$m(\{\text{Normal}\})$	$m(\{\text{Pneumonia}\})$	$m(\Omega)$
1	0,997327	0,002671	0,000001
2	0,178737	0,821241	0,000022
3	0,991401	0,008595	0,000004
4	0,984813	0,015164	0,000023
5	0,976593	0,023384	0,000023
6	0,950435	0,049407	0,000158
7	0,046187	0,953653	0,000160
8	0,038337	0,961658	0,000005
9	0,008903	0,991089	0,000008
10	0,013307	0,986692	0,000001

4.2.1. Proses penggabungan Nilai Mass Function

Pada bagian ini akan diberikan contoh penggabungan dua *mass function* menggunakan *Dempster's Rule of Combination*. Contoh penggabungan ini akan menggunakan sampel pertama pada Tabel 4.7 dan 4.8. Kedua tabel ini menampilkan dua sumber yang berbeda, CXR dan CT-Scan secara berurutan. Pada tiap sampel yang digunakan pada masing-masing tabel menghasilkan keluaran berupa *mass function* dengan $m(\{\text{Pneumonia}\})$ untuk keadaan *pneumonia*, $m(\{\text{Normal}\})$ untuk keadaan

normal, dan $m(\Omega)$ yang menggambarkan ketidaktahuan. Setiap subset dari tiap sumber dapat memberikan output sesuai pada Tabel 4.10:

Tabel 4.10 Nilai *Mass Function* Contoh *Dempster's Rule of Combination*

CXR	CT-Scan
$m_1(\{\text{Pneumonia}\}) = 0,963053$	$m_2(\{\text{Pneumonia}\}) = 0,934570$
$m_1(\{\text{Normal}\}) = 0,035809$	$m_2(\{\text{Normal}\}) = 0,064379$
$m_1(\Omega) = 0,001138$	$m_2(\Omega) = 0,001051$

Dengan menggunakan Persamaan (6) hasil penggabungan m_1 dan m_2 dengan menggunakan *Dempster's Rule of Combination* dapat dilihat pada Tabel 4.11.

Tabel 4.11 Perkalian Silang *Mass Function* Contoh *Dempster's Rule of Combination*

CXR \ CT-Scan	$m_1(\{\text{Pneumonia}\}) = 0,963053$	$m_1(\{\text{Normal}\}) = 0,035809$	$m_1(\Omega) = 0,001138$
$m_2(\{\text{Pneumonia}\}) = 0,934570$	$m(\{\text{Pneumonia}\}) = 0,900041$	$m(\{\emptyset\}) = 0,033466$	$m(\{\text{Pneumonia}\}) = 0,001064$
$m_2(\{\text{Normal}\}) = 0,064379$	$m(\{\emptyset\}) = 0,062000$	$m(\{\text{Normal}\}) = 0,002305$	$m(\{\text{Normal}\}) = 0,000073$
$m_2(\Omega) = 0,001051$	$m(\{\text{Pneumonia}\}) = 0,001012$	$m(\{\text{Normal}\}) = 0,000038$	$m(\{\Omega\}) = 0,0000012$

Proses selanjutnya ialah menghitung nilai K (lihat Persamaan 7) didapatkan dari, $k = \sum m_1(A)m_2(B)$ dimana $A \cap B = \emptyset$ sehingga didapat nilai $k = 0,095466$ yang berarti $1-k = 0,904534$.

Persamaan (6) digunakan untuk melakukan perhitungan mass ketika S_1 dan S_2 digabungkan. Hasil penggabungan kedua mass function dapat dilihat seperti pada Tabel 3.6.

Tabel 4.12 Hasil Akhir Contoh *Dempster's Rule of Combination*

$m(\{\text{Pneumonia}\}) =$	$0,902117/0,904534 = 0,997327$
$m(\{\text{Normal}\}) =$	$0,002416/0,904534 = 0,002671$
$m(\Omega) =$	$0,0000012/0,904534 = 0,000001$

Dari Tabel 4.12 dapat disimpulkan bahwa melalui proses penggabungan *mass function* menggunakan *Dempster's Rule of Combination*, pada contoh kasus ini hipotesis dengan kepercayaan tertinggi terdapat pada hipotesis Pnemonia dengan mass sebesar 0,997327. Sedangkan mass pada Ω sebesar 0,000001, menunjukkan perhitungan *uncertainty* pada hasil penggabungan dari kedua modalitas. Untuk memastikan keberhasilan proses penggabungan nilai *mass function* menggunakan *Dempster's Rule of Combination* bisa dilakukan dengan menghitung nilai total keseluruhan mass: $0,997327 + 0,002671 + 0,000001 = 1$

4.2.2. Analisis Hasil Multimodal

Terdapat perbedaan yang mencolok dari model multimodal dibandingkan model modalitas tunggal. Terlihat dari nilai $m(\Omega)$ yang sangat mendekati nilai nol pada seluruh sampel yang disajikan, dengan rentang nilai 0,000001 hingga 0,000158 dan dengan rata-rata senilai 0,000041. Nilai ini jauh lebih kecil dibandingkan $m(\Omega)$ pada kedua modalitas tunggal CXR maupun CT-Scan, dengan penurunan sebesar 99,4% dari modalitas CXR dan 99,1% dari modalitas CT-Scan. Penurunan yang dratis ini mengkonfirmasi bahwa proses penggabungan *mass function* menggunakan *Dempster's rule of combination* (Persamaan 6 - 7) berhasil mengeliminasi hampir seluruh ketidakpastian yang muncul pada masing-masing modalitas tunggal dan menghasilkan model yakin dalam keputusan klasifikasinya.

Melihat dari sesi ketegasan pada keputusan yang diberikan, model multimodal secara konsisten menghasilkan *mass function* yang lebih terpolarisasi dibandingkan kedua modalitas tunggal. Terlihat pada sampel pertama pada Tabel 4.9 menunjukkan nilai $m(\{\text{Normal}\}) = 0,9973$ yang merupakan nilai keyakinan tertinggi pada seluruh sampel serta (Tabel 4.7, 4.8, 4.9) melampaui nilai tertinggi pada sampel CT-Scan (0,9346) maupun CXR (0,9692). Hal ini juga terdapat pada sampel nomor 9 yang menunjukkan nilai $m(\text{pneumonia})$ tertinggi dari semua keyakinan pada kelas tersebut. Peningkatan ketegasan keputusan ini merupakan bukti nyata bahwa informasi yang dibawa oleh CXR dan CT-Scan bersifat saling melengkapi (*complementary*), sehingga penggabungannya menghasilkan representasi belief yang jauh lebih akurat dan meyakinkan.

4.3. Hasil Evaluasi dengan K-Fold Cross Validation

Untuk menjawab rumusan masalah ketiga, dilakukan perbandingan performa antara tiga konfigurasi model, model ENN berbasis modalitas CXR, model ENN berbasis modalitas CT-Scan, dan Model ENN berbasis multimodal yang mengintegrasikan CXR dan CT-Scan menggunakan *Dempster's rule of combination* (Persamaan 6 - 7). Evaluasi dilakukan memanfaatkan K-fold cross validation dengan jumlah 10 *fold* menggunakan metrik E_{pl} (Persamaan 11) dan *reward* U_{65} (Persamaan 12 - 13). Hasil evaluasi pada setiap *fold* untuk kedua metrik tadi disajikan pada tabel 4.13 dan Tabel 4.14

Tabel 4.13 Nilai E_{pl} dari Semua Model ENN menggunakan 10-Fold

Fold	X-ray E_{pl}	CT-Scan E_{pl}	Fusion E_{pl}
1	11,89	7,72	1,28
2	11,32	7,69	1,91
3	9,90	7,84	1,14
4	12,87	9,67	3,21
5	11,68	8,62	3,23
6	11,88	5,87	1,64
7	11,78	7,81	2,12
8	9,78	9,77	2,23
9	14,96	9,14	3,66
10	11,82	8,91	2,18
Mean	11,79	8,31	2,26
Std	1,38	1,11	0,81

Tabel 4.14 Nilai *Reward* U_{65} dari Semua Model ENN menggunakan 10-Fold

Fold	X-ray Reward U_{65}	CT-Scan Reward U_{65}	Fusion Reward U_{65}
1	0,9387	0,9742	1.0000
2	0,9408	0,9742	0,9903
3	0,9452	0,9645	0,9968
4	0,9235	0,9516	0,9806

Fold	X-ray Reward U_{65}	CT-Scan Reward U_{65}	Fusion Reward U_{65}
5	0,9419	0,9548	0,9839
6	0,9344	0,9860	0,9935
7	0,9344	0,9581	0,9903
8	0,9484	0,9516	0,9903
9	0,9053	0,9581	0,9774
10	0,9323	0,9613	0,9968
Mean	0,9345	0,9634	0,9900
Std	0,0118	0,0108	0,0070

4.4. Analisis Perbandingan Performa

Berdasarkan hasil evaluasi pada Tabel 4.13 dan Tabel 4.14, terdapat pola yang jelas dan konsisten bahwa model multimodal secara signifikan mengungguli kedua model modalitas tunggal pada seluruh fold dan pada kedua metrik evaluasi yang digunakan. Pada metrik *Reward U_{65}* (2.12)-(2.13), model multimodal menghasilkan nilai rata-rata tertinggi sebesar $0,990 \pm 0,007$, dibandingkan model CT-Scan sebesar $0,963 \pm 0,011$ dan model CXR sebesar $0,935 \pm 0,011$. Peningkatan nilai U_{65} (Persamaan 12 - 13) dari modalitas CXR ke multimodal mencapai 5,94%, sementara peningkatan dari CT-Scan ke multimodal sebesar 2,76%. Mengamati nilai standar deviasi U_{65} yang dihasilkan, model *multimodal* (0,0070) yang jauh lebih kecil dibandingkan CXR (0,0181) dan CT-Scan (0,0118), mengindikasikan bahwa model *multimodal* tidak hanya menghasilkan performa terbaik tetapi juga paling konsisten dan stabil di seluruh *fold*.

Pada metrik E_{pl} , model multimodal menghasilkan nilai rata-rata terendah sebesar $2,26 \pm 0,81$, menghasilkan nilai jauh lebih baik dibandingkan model CT-Scan ($8,31 \pm 1,11$) maupun model CXR ($11,79 \pm 1,38$). Penurunan nilai E_{pl} dari modalitas CXR ke *multimodal* mencapai 80,82%, sementara penurunan dari CT-Scan ke *multimodal* mencapai 72,79%. Penurunan E_{pl} yang cukup drastis mengkonfirmasi bahwa penggabungan informasi dari dua modalitas menggunakan *Dempster's rule of combination* (Persamaan 6 - 7) mampu secara signifikan meningkatkan kualitas distribusi *mass function* yang dihasilkan, jauh melampaui apa yang dapat dicapai oleh masing-masing modalitas secara individual. Peningkatan pada kedua metrik terjadi secara konsisten pada seluruh *fold* tanpa pengecualian. Hal ini merupakan bukti kuat bahwa keunggulan model multimodal bukan

merupakan kebetulan pada pembagian data tertentu melainkan karakteristik pada pendekatan multimodal berbasis DST yang diterapkan pada penelitian ini. Hasil dari tiap modalitas dan multimodal disajikan pada Tabel 4.15.

Tabel 4.15 Tabel Hasil E_{pl} dan Reward U_{65} dari Kedua modalitas serta Multimodal

Modalitas	E_{pl} (Mean $\pm$ Std)	Reward U_{65} (Mean $\pm$ Std)
CXR	11,79 $\pm$ 1,38	0,9188 $\pm$ 0,0161
CT-Scan	8,31 $\pm$ 1,11	0,9500 $\pm$ 0,0103
Multimodal (CXR + CT)	2,26 $\pm$ 0,81	0,9816 $\pm$ 0,0056

4.5. Perbandingan dengan Penelitian Sejenis

Penelitian ini dan penelitian yang dilakukan oleh Huang et al. [10] sama-sama menggunakan data multimodal (Multi-MRI dan PET-CT) dalam melakukan tugas segmentasi/klasifikasi berbasis *neural network* yang mengimplementasikan *Dempster-Shafer Theory* dalam mempertimbangkan aspek ketidakpastian dalam prediksinya. Hal ini menjadi kedua penelitian berada pada domain permasalahan yang serupa, yaitu evaluasi keandalan model dalam kondisi data medis yang kompleks dengan melihat ketidakpastian yang diciptakannya. Hasil penelitian ini dapat dilihat pada Tabel 4.16.

Tabel 4.16 Hasil Penelitian dengan Penelitian sebelumnya

No	Penelitian	E_{pl}	Reward U_{65}	ECE	NLL	Brier Score	Mean Dice
1	Penelitian ini	2,26 $\pm$ 0,81	0,9816 $\pm$ 0,0056	-	-	-	-
2	Huang et al. [10] (Multi-MRI)	-	-	0,060 $\pm$ 0,0013	0,115 $\pm$ 0,0022	2,189 $\pm$ 0,041	0,864 $\pm$ 0,0058
3	Huang et al. [10] (PET-CT)	-	-	0,045 $\pm$ 0,0013	0,056 $\pm$ 0,0027	0,180 $\pm$ 0,013	0,811 $\pm$ 0,034

Meskipun terdapat kesamaan pada metode penelitian, perbedaan terdapat dari fokus penelitian serta kerangka evaluasi yang digunakan pada keduanya. Pada penelitian ini difokuskan pada proses klasifikasi citra pada paru-paru untuk mendeteksi penyakit pneumonia, dimana pada penelitian oleh Huang et al. [10] difokuskan pada segmentasi citra pada penyakit tumor otak dan limfoma dibedakan pada dataset yang digunakan dalam melakukan prosesnya (tumor otak pada data Multi-MRI dan limfoma pada data PET-CT).

Pada kerangka evaluasi metrik yang digunakan pada penelitian Huang et al. [10] adalah, metrik *Brier score*, *negative log-likelihood* (NLL), dan *Expected Calibration Error* (ECE), ketiga metrik digunakan untuk mengukur kuantifikasi ketidakpastian [10]. Berbeda dengan penelitian ini yang menggunakan metrik E_{pl} (Persamaan 11) dan *Reward* U_{65} (Persamaan 12) untuk mengukur kuantifikasi ketidakpastiannya. Selain itu, pada penelitian Huang et al. [10] terdapat metrik *mean dice* yang digunakan akurasi dari hasil segmentasi citra. Perbedaan focus dan kerangka evaluasi ini mengakibatkan kedua hasil penelitian tidak dapat dibandingkan secara langsung. Namun demikian, secara konseptual keduanya memiliki tujuan yang sama untuk menilai seberapa andal model dalam merepresentasi ketidakpastian prediksinya pada data medis multimodal. Melalui dua penelitian ini terbukti keberhasilan dari pemanfaatan Detspiter-Shafer Theory dalam melakukan proses klasifikasi maupun segmentasi pada data multimodal citra serta bagaimana penggunaan multimodal dalam proses ini memberikan hasil yang lebih baik.