

RISET INFORMATIKA D081
KLASIFIKASI PENYAKIT ESOFAGITIS PADA CITRA
ENDOSKOPI DENGAN EKSTRAKSI FITUR GRAY LEVEL
CO-OCCURENCE MATRIX (GLCM) DAN ALGORITMA
SUPPORT VECTOR MACHINE (SVM)



Oleh

Baktiar Yudha Yana / 21081010324

Dosen Pengampu:

Dr. Basuki Rahmat, S.Si, M.T

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2024

BAB I

PENDAHULUAN

1.1. Latar Belakang

Esofagitis adalah salah satu penyakit pada saluran pencernaan bagian atas yang terjadi karena adanya inflamasi atau iritasi pada lapisan esofagus (kerongkongan). Penyakit ini dapat disebabkan oleh berbagai faktor seperti refluks asam lambung, infeksi, alergi, atau efek samping obat-obatan. Jika tidak ditangani dengan baik, Peningkatan jumlah mediator inflamasi dan produksi *Reactive Oxygen Species* (ROS) berkaitan dengan esofagitis refluks dan berkontribusi dalam terjadinya risiko kanker esofagus terutama pada *barrett esofagus*[1]. Berdasarkan data dari Kementerian Kesehatan Republik Indonesia, Sekitar 5–15% pasien dengan penyakit refluks gastroesofageal (*GERD*) kronis dapat berkembang menjadi *barrett esofagus*, dan sekitar 0,5% dari kasus *barrett esofagus* berkembang menjadi kanker esofagus setiap tahunnya. Oleh karena itu, diagnosis yang tepat dan cepat sangat penting untuk menentukan langkah pengobatan yang efektif.

Dalam dunia medis, Endoskopi merupakan suatu sarana penunjang diagnostik dan terapeutik yang cukup handal dalam mendiagnosis penyakit dengan cara memasukan alat endoskopi kedalam saluran pencernaan[2]. Teknologi ini memungkinkan dokter untuk memperoleh gambaran langsung dari esofagus, namun interpretasi citra secara manual sering kali menjadi tantangan karena variasi tekstur dan pola yang kompleks, serta kemiripan gejala dengan penyakit lain seperti ulkus atau kanker esofagus. Dalam beberapa tahun terakhir, teknologi kecerdasan buatan semakin banyak diterapkan di bidang medis untuk mengatasi tantangan ini, khususnya untuk meningkatkan akurasi diagnosis dan efisiensi kerja tenaga medis. Salah satu cabang *Artificial Intelligence* yang menonjol adalah pembelajaran *machine learning*, yang dapat digunakan untuk menganalisis dan mengklasifikasikan citra medis secara otomatis. Kemampuan *Artificial Intelligence* ini membuka peluang baru untuk meningkatkan kualitas layanan kesehatan dan mempermudah akses kesehatan bagi semua orang[3].

GLCM (Gray Level Co-Occurrence Matrix) merupakan sebuah teknik untuk mendapatkan nilai statistik orde ke-2 dengan menghitung probabilitas hubungan kedekatan antara dua buah piksel pada jarak (d) dan sudut (θ) tertentu[4]. Metode ini sangat berguna untuk mengolah tekstur kompleks pada citra endoskopi esofagus dengan menghasilkan nilai array yang merepresentasikan hubungan spasial antar piksel. Pada tahap perhitungan *GLCM*, terbentuk matriks dengan sudut 0° , 45° , 90° dan 135° [5]. Fitur yang diekstraksi seperti kontras, homogeneity, energi dan correlation, yang kemudian digunakan sebagai input dalam proses klasifikasi.

SVM (Support Vector Machine) adalah salah satu algoritma machine learning yang digunakan untuk klasifikasi dan regresi[6]. *SVM (Support Vector Machine)* bekerja dengan mencari hyperplane optimal untuk memisahkan data ke dalam dua kelas berbeda. Kombinasi *GLCM (Gray Level Co-Occurrence Matrix)* dan *SVM (Support Vector Machine)* diharapkan mampu menghasilkan model klasifikasi yang unggul untuk mendeteksi penyakit esofagitis. Dengan pendekatan ini, penelitian berfokus pada tahapan preprocessing, ekstraksi fitur tekstur menggunakan *GLCM*, dan klasifikasi menggunakan *SVM* untuk membangun sistem yang akurat dan efisien.

Pada penelitian yang dilakukan oleh [5] yang membahas implementasi metode *Gray-Level Co-Occurrence Matrix (GLCM)* dan *Support Vector Machine (SVM)* untuk mengidentifikasi tingkat kematangan buah Monk secara otomatis. Dataset terdiri dari 991 gambar dengan dua kategori, yaitu “Matang” (635 gambar) dan “Belum Matang” (356 gambar). Proses melibatkan preprocessing berupa cropping, resizing, dan konversi ke grayscale, serta ekstraksi fitur menggunakan *GLCM* dengan parameter seperti contrast, homogeneity, energy, dan correlation pada sudut tertentu. Metode *SVM* dengan kernel linear menghasilkan akurasi terbaik sebesar 89% pada parameter $C=50$, menunjukkan potensi teknologi ini untuk mendukung efisiensi industri pertanian.

Pada peneliti lainnya [7] yang menggunakan metode *Gray Level Co-Occurrence Matrix (GLCM)* untuk ekstraksi fitur tekstur dan *Support Vector Machine (SVM)* dengan kernel RBF dalam klasifikasi kolitis ulseratif pada citra

endoskopi dari dataset *Kvasir*, yang terdiri dari 1990 citra (995 citra sehat dan 995 citra kolitis ulseratif). Dengan memanfaatkan parameter seperti contrast, energy, dan homogeneity, serta optimasi hyperparameter (C dan gamma), model ini berhasil mencapai akurasi 90,95%, presisi 91,58%, recall 90,68%, dan f1-score 91,12%, menunjukkan performa tinggi dalam deteksi otomatis kondisi medis ini.

Pada peneliti lainnya [8] yang menggunakan metode *Gray Level Co-Occurrence Matrix (GLCM)* untuk ekstraksi fitur tekstur dan *Support Vector Machine (SVM)* untuk deteksi penyakit kulit, dengan fokus pada klasifikasi jenis penyakit kulit seperti eksim, psoriasis, dan dermatitis. Dari total 200 citra yang digunakan, data dibagi menjadi 140 citra untuk pelatihan dan 60 citra untuk pengujian. Hasil pengujian menunjukkan bahwa penggunaan kernel *Radial Basis Function (RBF)* menghasilkan akurasi tertinggi sebesar 92%, lebih baik dibandingkan dengan kernel Linear yang hanya mencapai akurasi 75%.

Berdasarkan penelitian-penelitian sebelumnya, diketahui bahwa kombinasi metode *Gray Level Co-Occurrence Matrix (GLCM)* dengan algoritma *Support Vector Machine (SVM)* dapat memberikan tingkat akurasi yang cukup tinggi. Oleh karena itu, peneliti mengangkat judul penelitian "Klasifikasi Penyakit Esofagitis pada Citra Endoskopi dengan Ekstraksi Fitur *Gray Level Co-Occurrence Matrix (GLCM)* dan Algoritma *Support Vector Machine (SVM)*" untuk mengetahui tingkat efektivitas metode *SVM* dengan kombinasi ekstraksi fitur *GLCM* dalam klasifikasi penyakit esofagitis pada citra endoskopi. Hasil penelitian ini diharapkan dapat membantu tenaga medis dalam mendeteksi penyakit esofagitis dengan lebih cepat dan akurat, sehingga dapat meningkatkan kualitas diagnosis, mempercepat penanganan, dan mencegah komplikasi lebih lanjut.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka rumusan masalah yang didapat sebagai berikut:

1. Bagaimana penerapan ekstraksi fitur *Gray Level Co-Occurrence Matrix (GLCM)* dan algoritma *Support Vector Machine (SVM)* dalam melakukan klasifikasi penyakit esofagitis?
2. Bagaimana performa klasifikasi penyakit esofagitis menggunakan ekstraksi fitur *Gray Level Co-Occurrence Matrix (GLCM)* dan algoritma *Support Vector Machine (SVM)*?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, maka tujuan yang ingin dicapai dalam penelitian ini sebagai berikut:

1. Menerapkan metode ekstraksi fitur *Gray Level Co-Occurrence Matrix (GLCM)* dan algoritma *Support Vector Machine (SVM)* untuk melakukan klasifikasi penyakit esofagitis pada citra endoskopi.
2. Mengetahui performa klasifikasi penyakit esofagitis yang dihasilkan dari kombinasi ekstraksi fitur *Gray Level Co-Occurrence Matrix (GLCM)* dan algoritma *Support Vector Machine (SVM)*.

1.4. Manfaat penelitian

Penelitian ini memberikan manfaat bagi sejumlah pihak. Adapun manfaat penelitian ini bagi berbagai pihak terkait adalah sebagai berikut:

1. Bagi Akademis dan Penulis

Penelitian ini memberikan kesempatan bagi penulis untuk memperdalam pemahaman tentang teknologi kecerdasan buatan, khususnya dalam penerapan metode *Gray Level Co-Occurrence Matrix (GLCM)* dan algoritma *Support Vector Machine (SVM)* untuk analisis citra medis. Selain itu, penelitian ini juga membantu penulis mengembangkan kemampuan penelitian dan memecahkan masalah yang relevan dengan bidang informasi medis.

Menyediakan referensi untuk penelitian lebih lanjut di bidang analisis citra medis dan klasifikasi penyakit, khususnya dengan algoritma *Support Vector Machine (SVM)* dan metode ekstraksi fitur *Gray Level Co-Occurrence Matrix (GLCM)*.

2. Bagi Tenaga Medis

Hasil penelitian ini diharapkan dapat menjadi alat bantu yang meningkatkan akurasi dan kecepatan dalam proses diagnosis penyakit esofagitis. Dengan sistem berbasis *GLCM* dan *SVM*, tenaga medis dapat lebih mudah mengidentifikasi kondisi pasien secara objektif dan efisien, mengurangi risiko kesalahan diagnosis akibat faktor subjektivitas atau keterbatasan pengalaman klinis.

1.5. Batasan Masalah

Penelitian ini perlu adanya batasan masalah agar tidak keluar dari permasalahan yang dikaji, maka diperlukan batasan masalah sebagai berikut:

1. Dataset yang digunakan merupakan citra endoskopi esofagus yang diambil dari sumber *open source*, yaitu dataset yang sudah tersedia secara publik.
2. Data yang digunakan berjumlah 1000 citra endoskopi yang sudah diberi label esofagitis dan normal-z-line.
3. Penelitian ini menggunakan ekstraksi fitur *Gray Level Co-occurrence Matrix (GLCM)* pada citra endoskopi esofagus untuk mendeteksi pola tekstur yang terkait dengan esofagitis.
4. Penelitian ini menggunakan algoritma klasifikasi *Support Vector Machine (SVM)* untuk menganalisis dan mengklasifikasikan citra esofagitis.
5. Penelitian ini hanya berfokus pada pengembangan model klasifikasi untuk mendeteksi penyakit esofagitis pada citra endoskopi tanpa harus membangun sistem berbasis aplikasi atau alat medis. Dengan demikian, produk akhir dari penelitian ini adalah model klasifikasi penyakit esofagitis pada citra endoskopi.

BAB II

TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian-penelitian yang telah dilakukan oleh peneliti sebelumnya yang berkaitan dengan penelitian yang akan dilakukan adalah sebagai berikut:

1. Penelitian oleh Muhammad Thohir, Ahmad Zoebad Foeady, Dian Candra Rini Novitasari, Ahmad Zaenal Arifin, Bunga Yuwa Phiadelvira dan Ahmad Hanif Asyhar (2020).

Penelitian yang dilakukan oleh Muhammad Thohir, Ahmad Zoebad Foeady, Dian Candra Rini Novitasari, Ahmad Zaenal Arifin, Bunga Yuwa Phiadelvira dan Ahmad Hanif Asyhar yang berjudul “Classification of Colposcopy Data Using GLCM-SVM on Cervical Cancer”. Penelitian ini bertujuan untuk mengklasifikasikan data colposcopy dalam deteksi dini kanker serviks menggunakan metode Gray Level Co-Occurrence Matrix (GLCM) untuk ekstraksi fitur dan Support Vector Machine (SVM) sebagai metode klasifikasi. Proses penelitian dimulai dengan preprocessing gambar yang mencakup pengubahan citra menjadi grayscale, pengurangan noise menggunakan median filter, serta peningkatan kualitas gambar melalui histogram equalization. Setelah itu, fitur tekstur seperti contrast, correlation, energy, dan homogeneity diekstraksi menggunakan metode GLCM.

Pada tahap klasifikasi, metode SVM diuji dengan beberapa jenis kernel, yaitu linear, polynomial, dan gaussian. Hasil terbaik diperoleh menggunakan kernel polynomial dengan sudut 45° dari GLCM, yang menghasilkan tingkat akurasi sebesar 90%. Penelitian ini menunjukkan bahwa kombinasi GLCM dan SVM efektif dalam mendeteksi kanker serviks, terutama pada tahapan yang berbeda (normal, stage 1, stage 2, stage 3, dan stage 4), sehingga dapat membantu tenaga medis dalam mendiagnosis kanker serviks secara lebih cepat dan akurat. Hasil ini diharapkan dapat digunakan sebagai dasar pengembangan sistem pendukung keputusan untuk deteksi dini kanker serviks yang lebih efisien.

2. Penelitian oleh Dadang Iskandar Mulyana, Dewi Riyanti Wibowo (2023).

Penelitian yang dilakukan oleh Dadang Iskandar Mulyana, Dewi Riyanti Wibowo yang berjudul “Implementasi Tingkat Kematangan Buah Monk Dengan Menggunakan Ekstraksi Gray-Level Co-Occurrence Matrix (GLCM) Dan Support Vector Machine (SVM)”. Penelitian ini membahas implementasi metode Gray-Level Co-Occurrence Matrix (GLCM) dan Support Vector Machine (SVM) untuk menentukan tingkat kematangan buah Monk (*Siraitia grosvenorii*). Dataset yang digunakan dalam penelitian ini terdiri dari 991 citra, yang terbagi menjadi dua kelas: "Matang" (635 citra) dan "Belum Matang" (356 citra). Tahapan penelitian diawali dengan proses preprocessing, meliputi cropping citra, resize gambar ke ukuran 100x100 piksel, dan konversi citra menjadi grayscale. Proses ini bertujuan untuk menyederhanakan data gambar sekaligus mempertahankan informasi yang penting untuk analisis lebih lanjut.

Setelah preprocessing selesai, fitur tekstur dari gambar diekstraksi menggunakan metode GLCM, dengan parameter seperti contrast, homogeneity, energy, dan correlation yang dihitung pada empat sudut arah, yaitu 0°, 45°, 90°, dan 135°. Ekstraksi fitur ini bertujuan untuk mendapatkan karakteristik tekstur yang dapat membedakan tingkat kematangan buah Monk. Selanjutnya, metode SVM dengan kernel linear digunakan untuk melakukan klasifikasi. Hasil pengujian menunjukkan bahwa model yang dikembangkan berhasil mencapai akurasi tertinggi sebesar 89% dengan parameter $C=50$, yang menunjukkan efektivitas metode ini dalam mendeteksi kematangan buah.

3. Penelitian oleh Agni Nurrohmana, Irani Hoeronisa, dan Hen Hen Lukmana (2024).

Penelitian yang dilakukan oleh Agni Nurrohmana, Irani Hoeronisa, dan Hen Hen Lukmana dalam jurnal yang berjudul “Klasifikasi Kolitis Ulseratif pada Citra Endoskopi menggunakan Support Vector Machine dengan Ekstraksi Citra Gray Level Co-Occurrence Matrix”. Penelitian ini bertujuan

untuk mengembangkan sistem klasifikasi otomatis yang dapat mendeteksi kolitis ulseratif dari citra endoskopi menggunakan metode ekstraksi fitur Gray Level Co-Occurrence Matrix (GLCM) dan algoritma Support Vector Machine (SVM). Penelitian ini menggunakan dataset Kvasir yang terdiri dari 1990 citra endoskopi, dengan dua kelas yaitu citra sehat dan citra kolitis ulseratif. Fitur tekstur seperti energi, kontras, homogenitas, dan dissimilarity diekstraksi dari citra endoskopi yang telah diproses menjadi grayscale, kemudian dilakukan klasifikasi menggunakan kernel Gaussian Radial Basis Function (RBF) pada SVM.

Hasil penelitian menunjukkan bahwa model SVM dengan kernel RBF, setelah melakukan pencarian nilai parameter terbaik (C dan gamma), berhasil mencapai akurasi sebesar 90,95%, dengan nilai presisi 91,58%, recall 90,68%, dan f1-score 91,12%. Ini menunjukkan bahwa kombinasi GLCM dan SVM efektif dalam mendeteksi kolitis ulseratif pada citra endoskopi, dengan hasil yang cukup baik untuk diterapkan dalam diagnosa medis otomatis.

4. Penelitian oleh Rizky Adawiyah dan Dadang Iskandar Mulyana (2022).

Penelitian yang dilakukan oleh Rizky Adawiyah dan Dadang Iskandar Mulyana dalam jurnal berjudul “Optimasi Deteksi Penyakit Kulit Menggunakan Metode Support Vector Machine (SVM) dan Gray Level Co-Occurrence Matrix (GLCM)”. Penelitian ini bertujuan untuk meningkatkan kemampuan deteksi dini penyakit kulit melalui pengolahan citra digital. Dalam penelitian ini, metode GLCM digunakan untuk ekstraksi fitur tekstur seperti energi, kontras, korelasi, dan homogenitas, sedangkan SVM digunakan untuk klasifikasi. Data yang digunakan dalam penelitian ini terdiri dari citra berbagai jenis penyakit kulit yang diambil dari sumber daring.

Proses penelitian dimulai dengan preprocessing data citra, yang mencakup konversi ke grayscale, blurring menggunakan Gaussian blur, serta transformasi gamma untuk meningkatkan kualitas citra. Kemudian,

fitur tekstur diekstraksi menggunakan GLCM dengan berbagai orientasi piksel. Tahap akhir melibatkan klasifikasi menggunakan metode SVM dengan kernel yang sesuai. Hasil penelitian menunjukkan akurasi sebesar 90%, menegaskan bahwa kombinasi GLCM dan SVM efektif untuk mendeteksi penyakit kulit secara akurat dan dapat membantu diagnosis dini dalam dunia medis.

5. Penelitian oleh Rizal Amegia Saputra, Diah Puspitasari dan Taufik Baidawi (2022).

Penelitian yang dilakukan oleh Rizal Amegia Saputra, Diah Puspitasari, dan Taufik Baidawi dalam jurnal berjudul “Deteksi Kematangan Buah Melon dengan Algoritma Support Vector Machine Berbasis Ekstraksi Fitur GLCM”. Penelitian ini bertujuan untuk mengembangkan sistem yang dapat mendeteksi tingkat kematangan buah melon secara otomatis melalui analisis tekstur citra digital. Metode yang digunakan mencakup ekstraksi fitur tekstur menggunakan Gray Level Co-occurrence Matrix (GLCM) dengan dua pendekatan orientasi sudut (4 sudut dan 8 sudut), serta klasifikasi menggunakan algoritma Support Vector Machine (SVM) dengan empat jenis kernel, yaitu linear, polynomial, sigmoid, dan radial basis function (RBF).

Dataset penelitian terdiri dari 650 citra buah melon yang diklasifikasikan ke dalam tiga kategori: matang, setengah matang, dan tidak matang. Proses analisis melibatkan cropping untuk memfokuskan citra pada area buah, diikuti dengan ekstraksi fitur GLCM. Dari hasil pengujian, kernel linear dengan orientasi 8 sudut menunjukkan performa terbaik dengan akurasi 80%, presisi 81%, dan recall 80%. Penelitian ini memberikan kontribusi dalam bidang pertanian presisi, khususnya untuk membantu petani menentukan waktu panen optimal, sehingga kualitas hasil panen dapat ditingkatkan.

6. Penelitian oleh Risha Ambar Wati, Hafiz Irsyad dan M Ezar Al Rivan (2020).

Penelitian yang dilakukan oleh Risha Ambar Wati, Hafiz Irsyad, dan M. Ezar Al Rivan dalam jurnal berjudul “Klasifikasi Pneumonia Menggunakan Metode Support Vector Machine (SVM)”. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi berbasis citra rontgen paru-paru dalam mendeteksi pneumonia. Proses penelitian ini mencakup preprocessing citra rontgen seperti cropping, resizing, contrast stretching, dan thresholding untuk menghasilkan citra biner. Setelah itu, fitur tekstur diekstraksi menggunakan metode Gray Level Co-Occurrence Matrix (GLCM) dan hasil ekstraksi diklasifikasikan dengan algoritma SVM.

Penelitian menggunakan dataset sebanyak 3.000 hingga 3.150 citra rontgen yang dibagi ke dalam data latih dan uji. Ekstraksi GLCM dilakukan pada empat orientasi sudut (0° , 45° , 90° , dan 135°), menghasilkan fitur tekstur berupa energi, kontras, korelasi, dan homogenitas. Hasil pengujian menunjukkan akurasi terbaik sebesar 62,66% dengan 600 data uji dan 59,2% dengan 750 data uji.

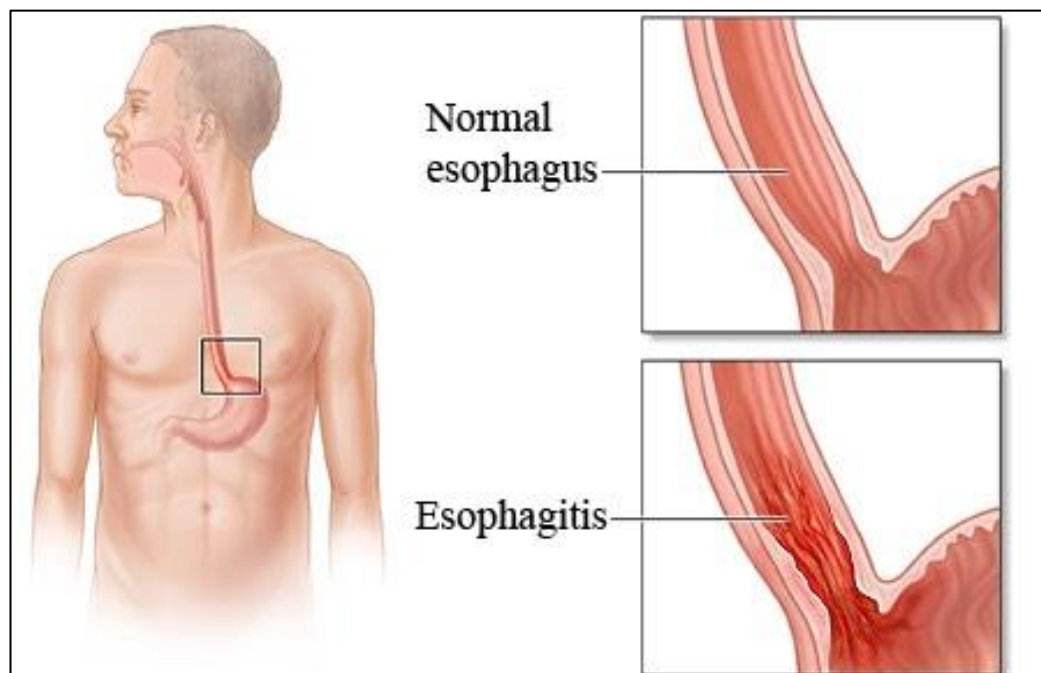
7. Penelitian oleh Anif Hanifa Setianingrum, Siti Umami Masruroh dan Syifa Fitratul M (2020).

Penelitian yang dilakukan oleh Anif Hanifa Setianingrum, Siti Umami Masruroh, dan Syifa Fitratul M dalam jurnal berjudul “Performance of Acne Type Identification Using GLCM and SVM”. Penelitian ini bertujuan untuk mengembangkan sistem berbasis komputer guna mengidentifikasi jenis jerawat berdasarkan citra kulit wajah. Sistem ini menggunakan metode Gray Level Co-Occurrence Matrix (GLCM) untuk ekstraksi fitur tekstur dan Support Vector Machine (SVM) sebagai algoritma klasifikasinya. Dalam penelitian ini, citra jerawat melewati tahapan preprocessing seperti cropping dan resizing, serta segmentation menggunakan Multi-Level Thresholding untuk meningkatkan akurasi analisis.

Data yang digunakan dalam penelitian terdiri atas 18 citra jerawat berukuran 256x256 piksel. Ekstraksi fitur GLCM menghasilkan empat karakteristik utama: energi, kontras, korelasi, dan homogenitas, yang kemudian digunakan oleh algoritma SVM untuk mengklasifikasikan jerawat menjadi tiga jenis: papula, pustula, dan nodul. Hasil pengujian menunjukkan bahwa sistem mampu mencapai akurasi hingga 89% untuk dataset 256x256 piksel, lebih baik dibandingkan akurasi 66,7% pada dataset berukuran 192x192 piksel.

2.2. Esofagitis

Esofagitis adalah kondisi peradangan yang terjadi pada esofagus, yaitu saluran yang menghubungkan mulut dengan lambung. Kondisi ini dapat disebabkan oleh berbagai faktor, seperti infeksi, refluks asam lambung, reaksi alergi, atau efek samping obat-obatan. Gejala umum esofagitis meliputi nyeri dada, kesulitan menelan dan sensasi terbakar di dada. Diagnosis esofagitis sering dilakukan melalui prosedur endoskopi untuk mengamati kondisi dinding esofagus dan pengambilan biopsi jika diperlukan. Berikut gambar 2.1 merupakan kondisi esofagus dalam keadaan normal dan kondisi terkena penyakit esophagitis.



Gambar 2. 1 Kondisi Esofagus

Salah satu penyebab utama esofagitis adalah gastroesophageal reflux disease (GERD), di mana kondisi naiknya isi lambung ke kerongkongan yang dapat mengakibatkan kerusakan pada saluran pencernaan mukosa esofagus[13]. Selain GERD, esofagitis juga dapat dipicu oleh konsumsi obat-obatan tertentu seperti antibiotik, aspirin, atau NSAID yang dapat merusak lapisan pelindung esofagus.

Esofagitis dapat dibedakan menjadi beberapa tipe, salah satunya adalah esofagitis refluks, yang terjadi akibat naiknya asam lambung ke esofagus. Kondisi ini biasanya disertai dengan gejala heartburn, atau sensasi terbakar di dada, yang sering kali terjadi setelah makan atau saat berbaring. Selain itu, esofagitis juga dapat disebabkan oleh infeksi jamur, bakteri, atau virus, yang sering kali terjadi pada individu dengan sistem kekebalan tubuh yang lemah, seperti penderita HIV/AIDS atau pasien yang menjalani terapi imunosupresan. Infeksi ini memerlukan pengobatan yang sesuai dengan jenis patogen yang terlibat.

Pengobatan esofagitis tergantung pada penyebab yang mendasarinya. Untuk esofagitis refluks, perubahan gaya hidup seperti menghindari makanan pemicu, tidur dengan kepala lebih tinggi, dan mengonsumsi obat-obatan antasid atau penghambat pompa proton (PPI) dapat membantu mengurangi gejala dan mempercepat penyembuhan. Pada kasus yang lebih parah, seperti esofagitis yang disebabkan oleh infeksi atau alergi, pengobatan dengan antibiotik, antijamur, atau obat-obatan anti alergi mungkin diperlukan. Dalam beberapa kasus yang sangat serius, prosedur pembedahan atau dilatasi esofagus mungkin diperlukan untuk memperbaiki kerusakan yang terjadi pada saluran pencernaan.

2.3. Endoskopi

Endoskopi merupakan suatu sarana penunjang diagnostik dan terapeutik yang cukup handal dalam mendiagnosis penyakit dengan cara memasukkan alat endoskop ke dalam saluran pencernaan[2]. Dalam konteks diagnosis esofagitis, endoskopi memungkinkan dokter untuk mengamati langsung kerusakan pada lapisan esofagus akibat refluks atau peradangan.

Dalam perkembangannya, teknologi endoskopi telah mengalami kemajuan yang signifikan sejak pertama kali diperkenalkan. Endoskopi modern dilengkapi dengan sistem optik berkualitas tinggi dan kemampuan perekaman video digital yang memungkinkan dokumentasi prosedur secara detail. Selain itu, beberapa endoskop juga dilengkapi dengan kanal kerja yang memungkinkan pengambilan sampel jaringan (biopsi) dan pelaksanaan prosedur terapeutik minimal invasif.



Gambar 2. 2 Endoskopi

Endoskopi juga telah berkembang dengan adanya teknologi endoskopi kapsul, yang memungkinkan pemeriksaan saluran pencernaan tanpa memerlukan alat endoskopi tradisional. Pada prosedur ini, pasien menelan kapsul kecil yang dilengkapi dengan kamera dan lampu LED, yang kemudian akan merekam gambar sepanjang perjalanan kapsul di dalam saluran pencernaan. Teknologi ini memberikan keuntungan karena lebih nyaman bagi pasien dan dapat mengakses area-area yang sulit dijangkau dengan endoskop tradisional, seperti usus kecil. Meskipun teknologi ini masih dalam tahap pengembangan dan memiliki beberapa

keterbatasan, endoskopi kapsul memberikan alternatif dalam diagnosis beberapa kondisi saluran pencernaan, termasuk esofagitis.

Selain itu, penggunaan teknologi endoskopi dengan sistem pencitraan canggih, seperti endoskopi berbantuan ultrasound (endoscopic ultrasound), juga semakin populer. Endoscopic Ultrasound memungkinkan visualisasi lapisan dalam saluran pencernaan dengan resolusi yang lebih tinggi dan mampu memberikan gambaran lebih jelas mengenai struktur jaringan di sekitar saluran pencernaan. Teknologi ini sangat berguna untuk mengevaluasi kondisi lebih lanjut dari peradangan atau komplikasi pada esofagitis, serta membantu dalam perencanaan terapi yang lebih tepat. Endoscopic Ultrasound dapat memberikan informasi lebih mendalam mengenai kedalaman peradangan atau keberadaan lesi, serta membantu dalam prosedur biopsi yang lebih akurat.

2.4. Pengolahan 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 [14]. Teknik ini melibatkan berbagai tahapan, seperti perbaikan kualitas terhadap suatu gambar (meningkatkan kontras, perubahan warna, restorasi citra) dan transformasi gambar (translasi, rotasi transformasi, skala, geometrik)[15]. Dalam bidang medis, pengolahan citra digital digunakan untuk membantu diagnosis penyakit dengan mengidentifikasi pola atau karakteristik tertentu pada citra, yang mungkin sulit diamati secara langsung.

Proses pengolahan citra dapat dikerjakan setelah melakukan input citra. Kemudian citra yang sudah di input ditukar sebagai bentuk matriks yang selanjutnya diproses dengan pengolahan citra. Hasil dari pengolahan citra yang dilakukan berupa output citra. Pengolahan citra mempunyai sejumlah teknik seperti penyaringan, modifikasi, pemetakan dan lain-lain[15].

2.5. Machine Learning

Machine learning adalah cabang dari kecerdasan buatan (AI) dan ilmu komputer yang fokus pada pengembangan algoritma dan model untuk menciptakan

sistem yang dapat belajar dan berkembang secara mandiri. Penerapan algoritma dan matematika memungkinkan sistem untuk mempelajari pola dari data dan membuat prediksi di masa depan.

Menurut [16], algoritma machine learning dapat dibagi menjadi tiga kategori utama:

1. Supervised Learning

Supervised Learning adalah metode machine learning yang menggunakan data berlabel untuk melatih algoritma dalam membuat estimasi terbaik terhadap output (Y) dari input (X). Proses ini efektif untuk tugas klasifikasi dan regresi, dengan contoh algoritma seperti K-NN, Naive Bayes, Decision Tree, dan Support Vector Machine. Meski sederhana dan mudah dipahami, supervised learning membutuhkan data pelatihan yang akurat dan komputasi yang lebih lama dibanding unsupervised learning karena memerlukan pelabelan pada setiap input.

2. Unsupervised Learning

Unsupervised learning adalah algoritma yang bekerja tanpa memerlukan data berlabel. Dalam pendekatan ini, algoritma tidak membutuhkan data pelatihan, melainkan digunakan untuk menemukan pola atau membuat model deskriptif tanpa perlu kategori atau output yang sudah ditentukan. Algoritma unsupervised learning banyak diterapkan dalam clustering dan asosiasi aturan. Keunggulan utamanya adalah fleksibilitasnya dalam mendeteksi pola yang sebelumnya mungkin tidak dikenali. Namun, kekurangannya adalah sulitnya mengidentifikasi informasi spesifik di dalam data karena tidak ada label, sehingga sulit juga membandingkan output dengan input.

3. Semi Supervised dan Reinforcement Learning

Semi-supervised learning adalah algoritma yang menggabungkan pendekatan supervised dan unsupervised, dengan bekerja pada data besar yang sebagian berlabel dan sebagian tidak. Keunggulannya adalah lebih hemat biaya karena hanya sebagian data yang memerlukan pelabelan dan

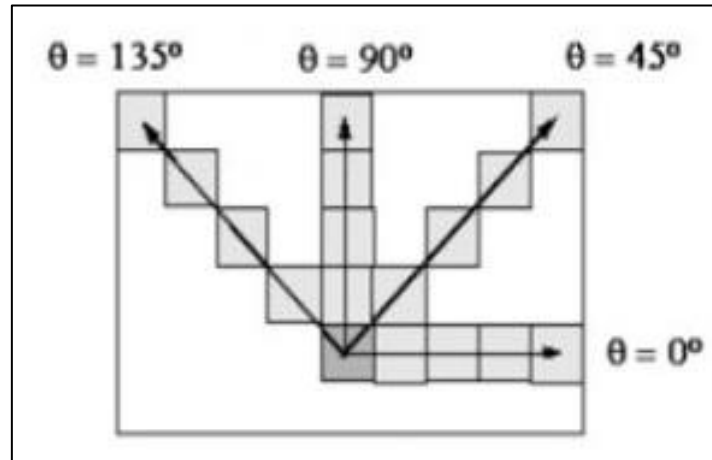
tidak membutuhkan tenaga ahli untuk pemrosesan. Sementara itu, reinforcement learning bertujuan memaksimalkan hasil dan mengurangi risiko dengan mengamati interaksi agen dengan lingkungannya. Algoritma ini belajar secara berulang, di mana agen mengamati data input, mengambil tindakan, dan menerima "reward" atau umpan balik dari lingkungan. Dengan mengamati 15 input ulang dan mendapatkan umpan balik tambahan, agen memperbaiki keputusannya secara bertahap untuk hasil yang lebih akurat.

2.6. GLCM

Gray Level Co-occurrence Matrix (GLCM) adalah metode analisis tekstur dalam citra digital yang dirancang untuk mengevaluasi hubungan spasial antara nilai intensitas piksel dalam berbagai sudut dan jarak tertentu. Metode ini pertama kali diperkenalkan oleh Haralick pada tahun 1973 dengan pengenalan 28 fitur yang menggambarkan pola spasial[17]. GLCM sangat efektif dalam menangkap informasi tekstur, terutama dalam aplikasi seperti analisis medis, pengolahan citra satelit, dan deteksi objek. Dengan menganalisis hubungan antar piksel, GLCM dapat memberikan wawasan mendalam tentang struktur dan pola yang terkandung dalam citra.

Proses pembentukan GLCM dimulai dengan mengukur hubungan spasial antar dua piksel tetangga berdasarkan intensitasnya. Hubungan ini dihitung untuk sudut 0° , 45° , 90° dan 135° [5], serta untuk jarak antar piksel yang telah ditentukan. Pasangan nilai intensitas piksel ini kemudian direpresentasikan dalam bentuk matriks co-occurrence, di mana setiap elemen matriks mencerminkan frekuensi kemunculan pasangan nilai intensitas tertentu.

Matriks ini dirancang untuk menangkap karakteristik tekstur, seperti distribusi tingkat keabuan dalam area tertentu dari citra. Berikut adalah ilustrasi atau gambaran arah dan sudut terhadap piksel bertetangga dapat dilihat pada Gambar 2.3.



Gambar 2. 3 Ilustrasi arah sudut GLCM

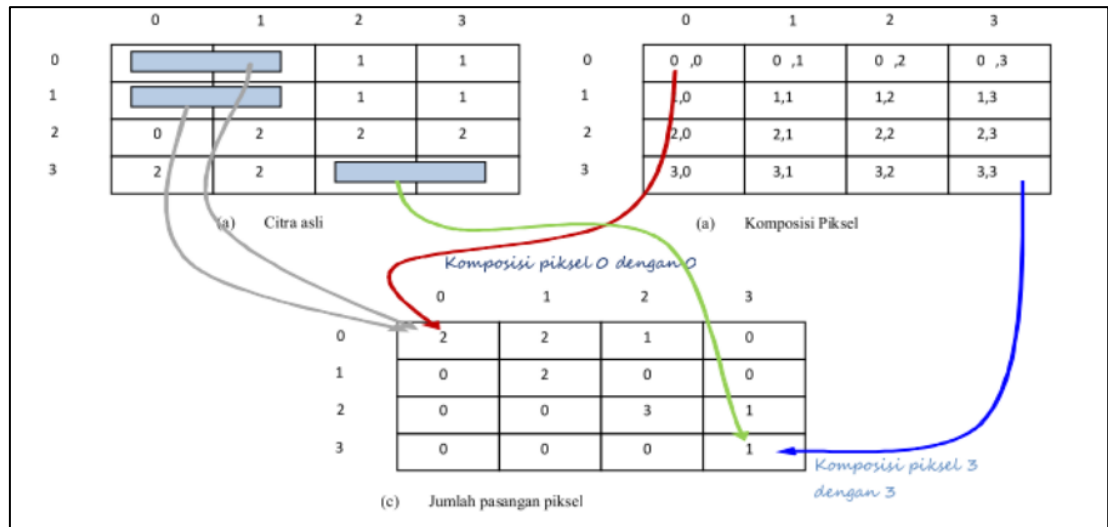
Pada setiap pasangan piksel, nilai intensitasnya direpresentasikan dalam bentuk matriks co-occurrence. Setelah matriks ini terbentuk, dilakukan normalisasi dengan cara membagi setiap elemen matriks dengan jumlah total elemen, sehingga menghasilkan probabilitas relatif dari setiap pasangan piksel.

Pada umumnya, sebuah citra terdiri dari tiga lapisan warna utama, yaitu Red, Green, dan Blue (RGB). Namun, setiap pengolahan yang melibatkan ketiga lapisan warna ini membutuhkan perhitungan terpisah untuk masing-masing lapisan, sehingga menghasilkan tiga kali perhitungan yang sama. Oleh karena itu, satu citra direpresentasikan melalui 3 buah matrik tingkat grayscale berupa matrik uRed (R-layer), matrik Green (G-layer) dan matrik untuk Blue(B-layer)[18]. Proses konversi ini dapat dilakukan menggunakan persamaan berikut:

$$I(i, y) = R * \alpha + G * \beta + B * \gamma$$

Pada persamaan tersebut, variabel R, G, dan B masing-masing merepresentasikan intensitas warna merah, hijau, dan biru. Konstanta α (0,2989), β (0,5870), dan γ (0,1140) adalah nilai yang digunakan untuk menghitung tingkat keabuan pada koordinat tertentu. Setelah dikonversi menjadi grayscale, metode GLCM digunakan untuk membentuk matriks co-occurrence berdasarkan data citra grayscale tersebut.

Matriks co-occurrence mencerminkan hubungan spasial antar piksel, di mana setiap elemen matriks merepresentasikan jumlah kemunculan pasangan piksel dengan level intensitas tertentu pada jarak dan orientasi sudut yang telah ditentukan. Hubungan ini, yang disebut sebagai co-occurrence atau kejadian bersama, mengacu pada frekuensi piksel bertetangga dengan nilai intensitas tertentu.



Gambar 2. 4 Ilustrasi Piksel Matriks GLCM

Gambar 2.4 memberikan ilustrasi proses pembentukan matriks co-occurrence yang digunakan untuk menganalisis tekstur pada citra. Panel (a) merepresentasikan citra asli dalam bentuk matriks tingkat keabuan, di mana setiap angka menunjukkan nilai intensitas piksel. Panel (b) memperlihatkan berbagai kemungkinan pasangan nilai intensitas piksel yang dihasilkan berdasarkan arah dan jarak tertentu. Selanjutnya, panel (c) menampilkan matriks GLCM, yaitu matriks yang mencatat frekuensi kemunculan setiap pasangan nilai intensitas piksel dalam citra.

Matriks GLCM ini menggambarkan hubungan spasial antara piksel dalam citra, yang dapat digunakan untuk mengekstraksi fitur tekstur. Dengan demikian, informasi dalam matriks ini mencakup pola distribusi intensitas yang relevan untuk analisis tekstur lebih lanjut.

$$\begin{bmatrix} 2 & 2 & 1 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 1 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

Transpos

GLCM sebelum dinormalisasi

Gambar 2. 5 Pembentukan Matriks

Membentuk matriks yang simetris dengan menjumlahkan matriks awal GLCM dengan nilai transposnya. Pada Gambar 2.5 dapat dilihat contoh pembentukan matriks[19].

$$\begin{bmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \xrightarrow{\text{normalisasi}} \begin{bmatrix} \frac{4}{24} & \frac{2}{24} & \frac{1}{24} & 0 \\ \frac{2}{24} & \frac{4}{24} & 0 & 0 \\ \frac{1}{24} & 0 & \frac{6}{24} & \frac{1}{24} \\ 0 & 0 & \frac{1}{24} & \frac{2}{24} \end{bmatrix}$$

Gambar 2. 6 Normalisasi Matriks GLCM

Gambar 2.6 dapat dilihat contoh nilai matriks GLCM yang telah dilakukan normalisasi dengan cara menjumlah semua pasangan piksel lalu dilakukan pembagian[19]. Proses ini bertujuan untuk menghilangkan pengaruh ukuran citra dan memastikan bahwa analisis hanya bergantung pada karakteristik tekstur.

GLCM digunakan untuk menghitung frekuensi kombinasi tingkat kecerahan piksel yang ditentukan. Matriks GLCM memiliki jumlah baris dan kolom yang sesuai dengan jumlah tingkat keabuan dalam citra. Setiap elemen matriks mencerminkan frekuensi kemunculan pasangan piksel tertentu berdasarkan jarak dan arah yang ditentukan. Kombinasi pasangan piksel ini dapat bervariasi tergantung pada area sekitar piksel dalam citra.

Setelah matriks dinormalisasi, hasilnya digunakan sebagai masukan untuk pengukuran probabilitas yang merepresentasikan fitur tekstur citra. Dalam analisis tekstur menggunakan GLCM, terdapat empat parameter utama yang sering digunakan, yaitu energy, contrast, correlation, dan homogeneity[20]. Parameter-parameter ini memberikan wawasan yang mendalam tentang pola dan karakteristik tekstur dalam citra.

1. Energi

Contrast pada fitur GLCM menunjukkan ukuran penyebaran elemen – elemen matriks ciri. Jika letaknya jauh dari diagonal utama, nilai contrast besar. Pada persamaan dibawah ini dijelaskan persamaan untuk menghitung nilai contrast, yaitu[20]:

$$\text{contrast} = \sum_i \sum_j (i - j)^2 P_{(i,j)}$$

Dimana, i adalah baris matriks, j adalah kolom matriks, P(i, j) adalah elemen matriks kookurensi baris (i) dan kolom (j).

2. Correlation

Correlation menyatakan ukuran ketergantungan linear derajat keabuan citra sehingga dapat memberikan petunjuk adanya struktur linear dalam citra. Pada persamaan dibawah ini dijelaskan persamaan untuk menghitung nilai correlation, yaitu:

$$\text{correlation} = \sum_i \sum_j \frac{(i - \mu_i)(j - \mu_j)P_{(i,j)}}{\sigma_i \sigma_j}$$

Dengan i adalah baris matriks, j adalah kolom matriks, $P(i,j)$ adalah elemen matriks kookurensi baris (i) dan kolom (j), μ_i, μ_j adalah rata – rata elemen pada baris dan kolom matriks, σ_i, σ_j adalah standar deviasi pada baris dan kolom matrik.

3. Energy

Energy mengukur tentang keseragaman atau sering disebut angular second moment, energy akan bernilai tinggi ketika nilai piksel mirip dengan piksel yang lain, sebaliknya akan bernilai kecil yang berarti nilai dari GLCM normalisasi adalah heterogen. Pada persamaan dibawah ini dijelaskan persamaan untuk menghitung nilai energy, yaitu:

$$correlation = \sum_i \sum_j P(i,j)^2$$

Dimana, i adalah baris matriks, j adalah kolom matriks, $P(i,j)$ adalah elemen matriks kookurensi baris (i) dan kolom (j).

4. Homogeneity

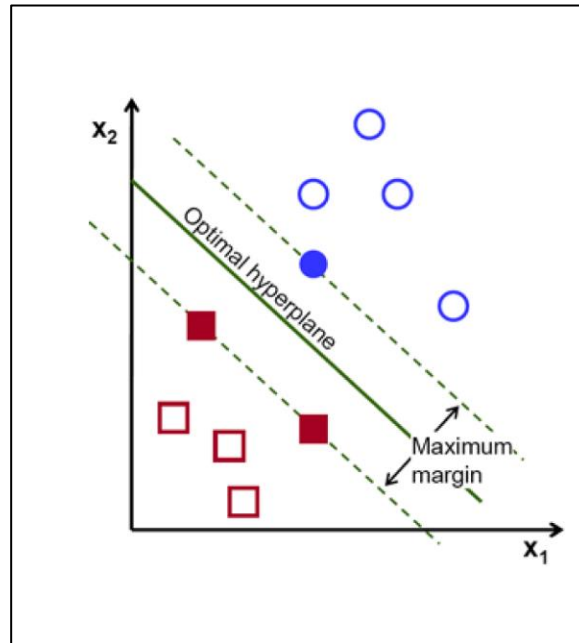
Homogeneity menunjukkan kehomogenan variasi intensitas dalam citra, homogeneity akan bernilai tinggi jika semua piksel mempunyai nilai yang uniform. Pada persamaan dibawah ini dijelaskan persamaan untuk menghitung nilai homogeneity, yaitu:

$$homogeneity = \sum_i \sum_j \frac{P(i,j)}{1 + |i - j|}$$

Dengan i adalah baris matriks, j adalah kolom matriks, $P(i,j)$ adalah elemen matriks kookurensi baris (i) dan kolom (j).

2.7. Support Vector Machine (SVM)

(*Support Vector Machine*) adalah salah satu algoritma machine learning yang digunakan untuk klasifikasi dan regresi[6], dengan pendekatan yang kuat untuk memisahkan data ke dalam kelas-kelas yang berbeda. *SVM* bekerja dengan menemukan hyperplane optimal yang memisahkan data dalam ruang berdimensi tinggi. Hyperplane adalah fungsi yang digunakan untuk membedakan antar fitur [21]. Untuk mencari hyperplane dapat dilakukan dengan mencari margin hyperplane dan mencari titik maksimum[22].



Gambar 2. 7 Support Vector Machine (SVM)

Salah satu keunggulan *SVM* adalah kemampuannya untuk menangani data non-linear melalui penggunaan kernel, yang memungkinkan pemetaan data ke dalam ruang yang lebih tinggi di mana pemisahan kelas menjadi lebih mudah dilakukan. Gambar 2.7 merupakan ilustrasi konsep Support Vector Machine (SVM), Grafik menunjukkan dua kelas data yang direpresentasikan dengan simbol warna merah (kelas pertama) dan biru (kelas kedua), dengan sumbu x_1 dan x_2 sebagai dua fitur atau variabel data. Garis yang berada di tengah disebut sebagai optimal hyperplane, yaitu garis pemisah yang memaksimalkan margin antara kedua kelas. Margin tersebut ditunjukkan oleh garis putus-putus di kedua sisi hyperplane. Margin ini diukur sebagai jarak terdekat antara hyperplane dan titik-titik data dari masing-masing kelas, yang disebut support vectors berada di tepi margin dan memiliki peran penting dalam menentukan posisi optimal hyperplane. Tujuan utama SVM adalah menemukan hyperplane dengan margin maksimum untuk menghasilkan pemisahan kelas yang optimal dan meminimalkan kesalahan klasifikasi.

2.8. Klasifikasi

Klasifikasi merupakan suatu proses untuk menemukan sekumpulan model maupun fungsi yang menjelaskan dan membedakan data kedalam kelas – kelas

tertentu, yang dimana bertujuan untuk Menggunakan model itu dalam menentukan kelas dari suatu objek yang belum diketahui kelasnya. Terdapat dua proses dalam Klasifikasi, yakni proses learning/training dengan melakukan pembangunan model Menggunakan data training. Proses testing dilakukan terhadap data testing Menggunakan model dari hasil data training[23].

2.9. Confusion Matriks

Confusion matrix adalah alat evaluasi yang digunakan untuk mengukur kinerja suatu model klasifikasi, khususnya dalam hal seberapa baik model dapat memprediksi kelas yang benar dari data yang ada. Matrix ini menyajikan perbandingan antara prediksi yang dihasilkan oleh model dengan nilai kebenaran dan biasanya terdiri dari empat komponen utama: True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN).

		<i>Observed</i>	
		<i>True</i>	<i>False</i>
<i>Predicted Class</i>	<i>True</i>	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
	<i>False</i>	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Gambar 2. 8 Confussion Matrix

Keterangan[24]:

1. TP = True Positif
Merupakan jumlah data positif yang terklasifikasi dengan benar oleh sistem
2. TN = True Negatif
Merupakan jumlah data negative yang terklasifikasi dengan benar oleh sistem.
3. FN = False Negatif
Merupakan jumlah data negative Namun terklasifikasi salah oleh sistem.
4. FP = False Positif

Merupakan jumlah data positif Namun terklasifikasi salah oleh sistem.

Dari confusion matrix tersebut, terdapat empat kategori data yang dapat digunakan untuk mengukur performa dari sebuah model atau algoritma, yaitu Accuracy, Precision, Recall, dan F-1 Score seperti berikut [25]:

1. Accuracy merupakan persentase benar model dalam melakukan Klasifikasi.

Rumus accuracy adalah sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \times 100\%$$

2. Precision merupakan persentase model memprediksi benar positif dengan keseluruhan data yang diprediksi positif.

Rumus precision adalah sebagai berikut :

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

3. Recall merupakan persentase prediksi benar positif ketika kelas actual data tersebut positif.

Rumus recall adalah sebagai berikut:

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

4. F-1 Score merupakan rata – rata antara precision dengan recall yang dibobotkan.

Rumus F-1 Score adalah sebagai berikut:

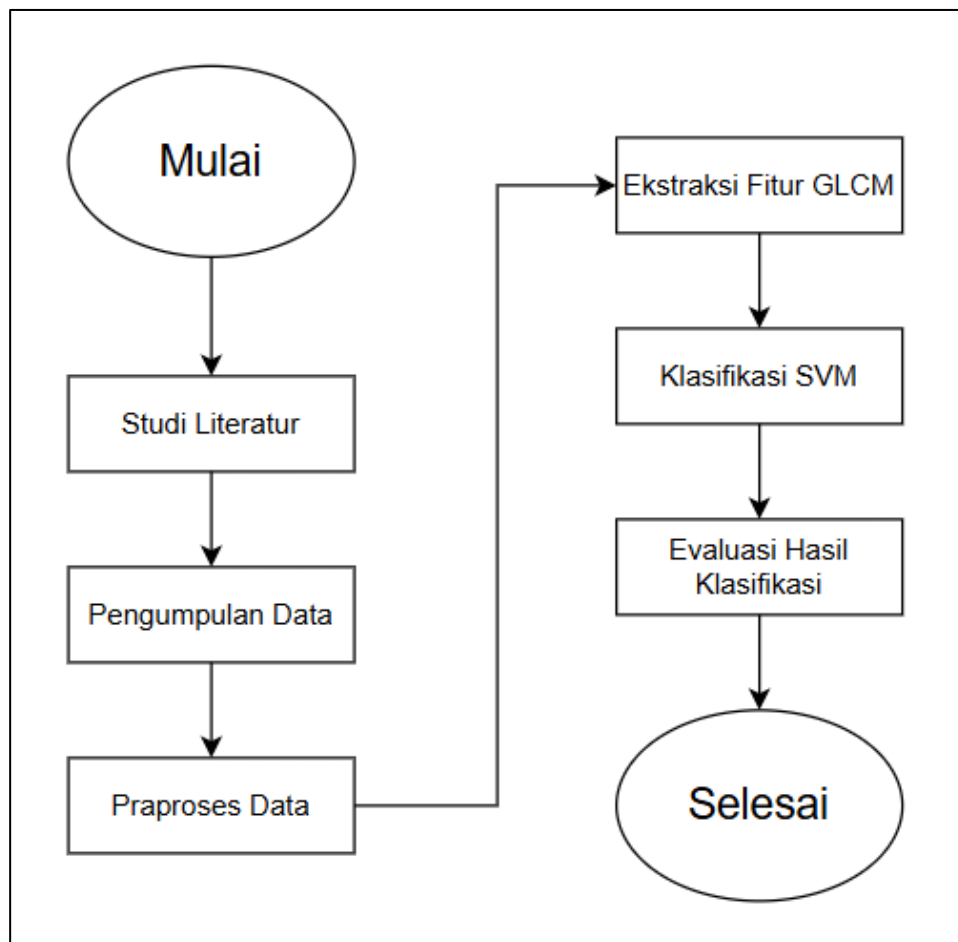
$$F - 1 \text{ Score} = \frac{(2 * Recall * Precision)}{Recall + Precision} \times 100\%$$

BAB III

DESAIN DAN IMPLEMENTASI SISTEM

3.1 Tahapan Penelitian

Dalam sebuah penelitian, diperlukan langkah-langkah yang terstruktur dan sistematis agar tujuan penelitian dapat dicapai secara efektif, sehingga mampu menjawab rumusan masalah serta memenuhi tujuan yang telah dijabarkan sebelumnya.



Gambar 3. 1 Tahapan Penelitian

Pada Gambar 3.1 menjelaskan alur proses Klasifikasi Penyakit Esofagitis Pada Citra Endoskopi Dengan Ekstraksi Fitur Gray Level Co-Occurrence Matrix (GLCM) Dan Algoritma Support Vector Machine (SVM). Penelitian ini diawali dengan studi literatur, yang berfungsi sebagai dasar untuk memahami konsep-konsep teoretis serta metodologi yang relevan dengan topik penelitian. Studi literatur membantu

peneliti dalam menentukan langkah-langkah penelitian yang sesuai serta memastikan pendekatan yang digunakan sudah mendukung pencapaian tujuan penelitian.

Tahap berikutnya adalah pengumpulan data, yang mencakup pengumpulan citra endoskopi sebagai data utama dalam penelitian ini. Data yang terkumpul akan melalui praproses data sehingga data menjadi siap untuk dianalisis lebih lanjut. Tahapan ini penting untuk memastikan data dalam kondisi optimal untuk proses klasifikasi berikutnya.

Setelah data diproses, dilakukan ekstraksi fitur menggunakan metode Gray Level Co-Occurrence Matrix (GLCM). Metode ini bertujuan untuk menggali informasi pola tekstur yang terkandung dalam citra, yang merupakan salah satu aspek penting dalam membedakan kondisi kesehatan saluran pencernaan. Fitur tekstur yang dihasilkan dari proses ini akan menjadi input bagi algoritma klasifikasi.

Langkah akhir adalah klasifikasi menggunakan Support Vector Machine (SVM). SVM berfungsi untuk memetakan data fitur ke dalam kategori tertentu, seperti Esofagitis dan kondisi normal. Setelah model klasifikasi selesai dibangun, dilakukan evaluasi terhadap performa model dengan mengukur tingkat akurasi, presisi, dan sensitivitasnya. Evaluasi ini memastikan bahwa model yang dikembangkan mampu memberikan hasil yang akurat dan sesuai dengan tujuan penelitian.

3.2 Studi Literatur

Dalam tahap awal penelitian ini, peneliti melakukan studi literatur untuk mengumpulkan dan memahami berbagai sumber teori yang relevan dengan permasalahan yang diangkat. Tahap ini sangat penting karena memberikan landasan pemahaman yang kuat terkait konsep, metode, dan istilah teknis yang digunakan. Salah satu fokus utama dalam literatur ini adalah metode ekstraksi fitur Gray Level Co-Occurrence Matrix (GLCM) dan algoritma Support Vector Machine (SVM). Untuk itu, peneliti memanfaatkan berbagai sumber, seperti jurnal, artikel ilmiah,

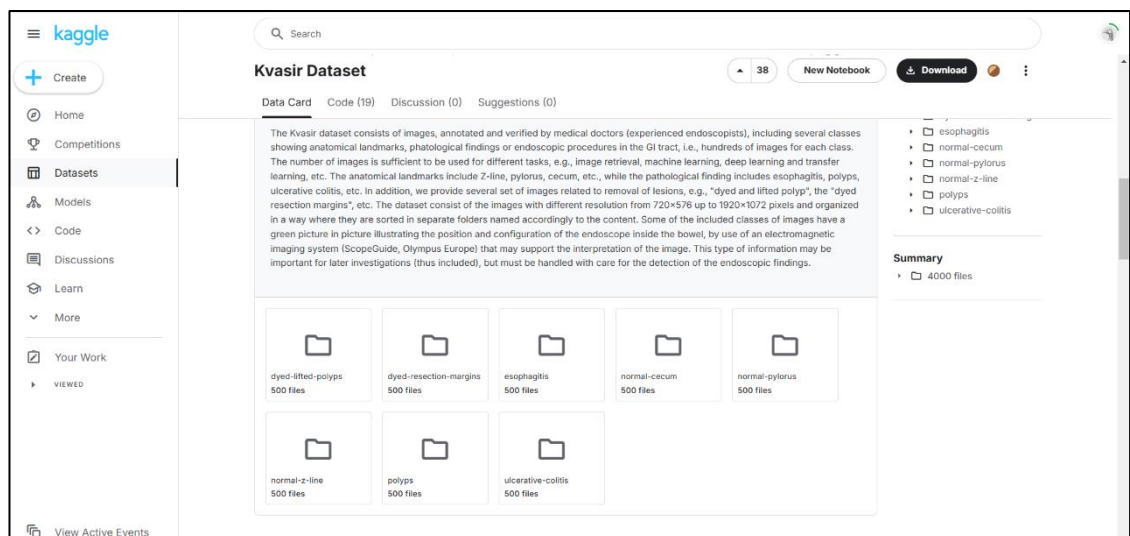
buku, dan penelitian terdahulu, guna memperoleh landasan teoritis yang mendalam dalam mendukung klasifikasi penyakit esofagitis melalui analisis tekstur citra endoskopi.

Studi literatur memberikan wawasan tentang penerapan GLCM dalam mengekstraksi pola tekstur dari citra endoskopi serta prinsip kerja algoritma SVM dalam proses klasifikasi. Penelitian sebelumnya menunjukkan bahwa kombinasi GLCM dan SVM memiliki keunggulan dalam hal keakuratan dan efisiensi. Pendekatan ini membantu mengidentifikasi pola tekstur spesifik yang menjadi indikator esofagitis, sehingga memberikan kontribusi besar terhadap proses diagnosis yang lebih cepat dan akurat di bidang medis.

Sebagian besar referensi yang digunakan dalam penelitian ini telah dicantumkan pada bagian Daftar Pustaka dalam laporan skripsi. Referensi-referensi tersebut tidak hanya menjadi rujukan utama, tetapi juga memberikan pengetahuan yang relevan untuk memahami dasar teoritis dari metode yang diterapkan.

3.3 Pengumpulan Data

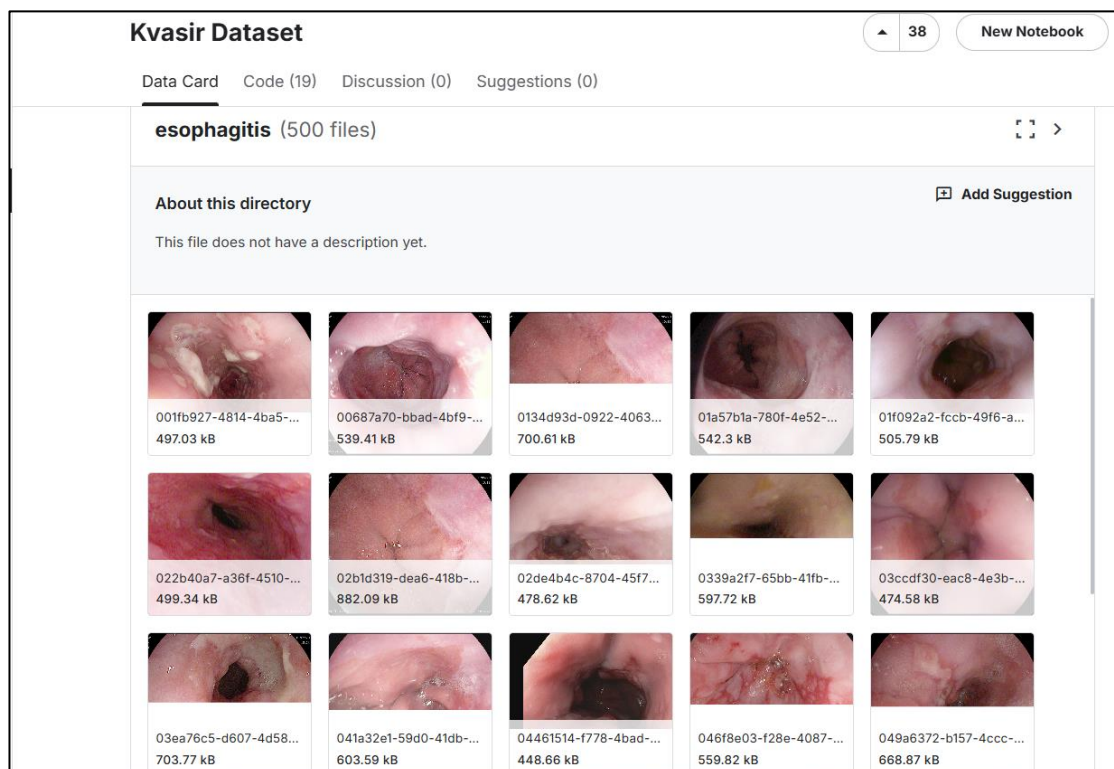
Dalam penelitian ini, data yang digunakan berasal dari dataset sekunder yang diperoleh dari sumber terbuka Kaggle, yaitu Kvasir Dataset, yang berisi 4.000 gambar citra endoskopi saluran pencernaan dalam format .jpg. Dataset ini mencakup berbagai kategori, seperti kondisi normal serta penyakit esofagitis, polip,



Gambar 3. 2 Dataset dari Kaggle

dan ulcerative colitis, dan dirancang untuk mendukung penelitian di bidang pengolahan citra medis, khususnya untuk deteksi dan klasifikasi penyakit. Dokumentasi pengambilan data kaggle dapat dilihat pada Gambar 3.10.

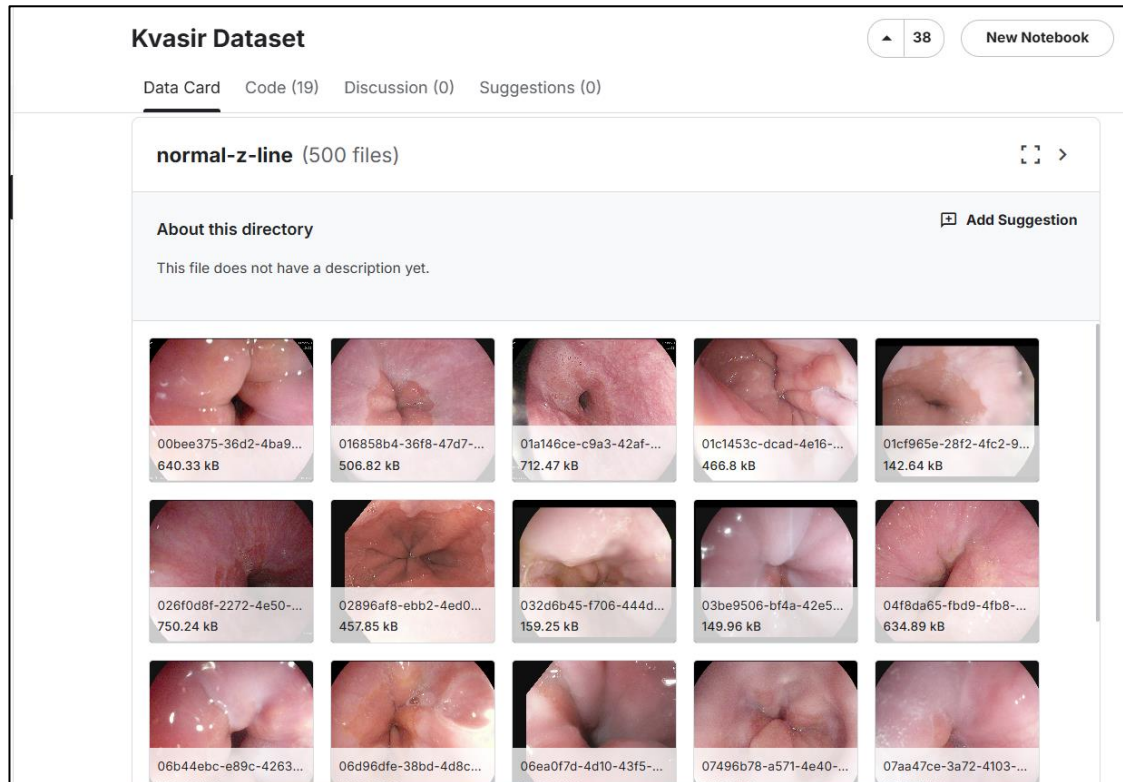
Penelitian ini berfokus pada dua kategori utama, yaitu esofagitis dan normal-z-line, dengan jumlah total 1.000 citra yang digunakan sebagai dataset. Setiap kategori memiliki jumlah data yang seimbang, sehingga memungkinkan analisis yang lebih akurat terhadap perbedaan pola tekstur antara kondisi normal dan patologis. Penggunaan dataset yang terstandarisasi memberikan kemudahan dalam proses pengembangan model klasifikasi, karena data sudah siap untuk digunakan tanpa perlu banyak penyesuaian tambahan.



Gambar 3. 3 Citra Endoskopi Esopagitis

Untuk memberikan gambaran yang lebih jelas, sampel citra dari kedua kategori tersebut disajikan pada Gambar 3.3 dan Gambar 3.4. Visualisasi ini menunjukkan variasi data yang digunakan dalam penelitian, sehingga dapat membantu mengilustrasikan pola dan karakteristik yang khas dari masing-masing kategori.

Dengan memanfaatkan data ini, peneliti dapat menganalisis perbedaan antara kondisi normal dan esofagitis.

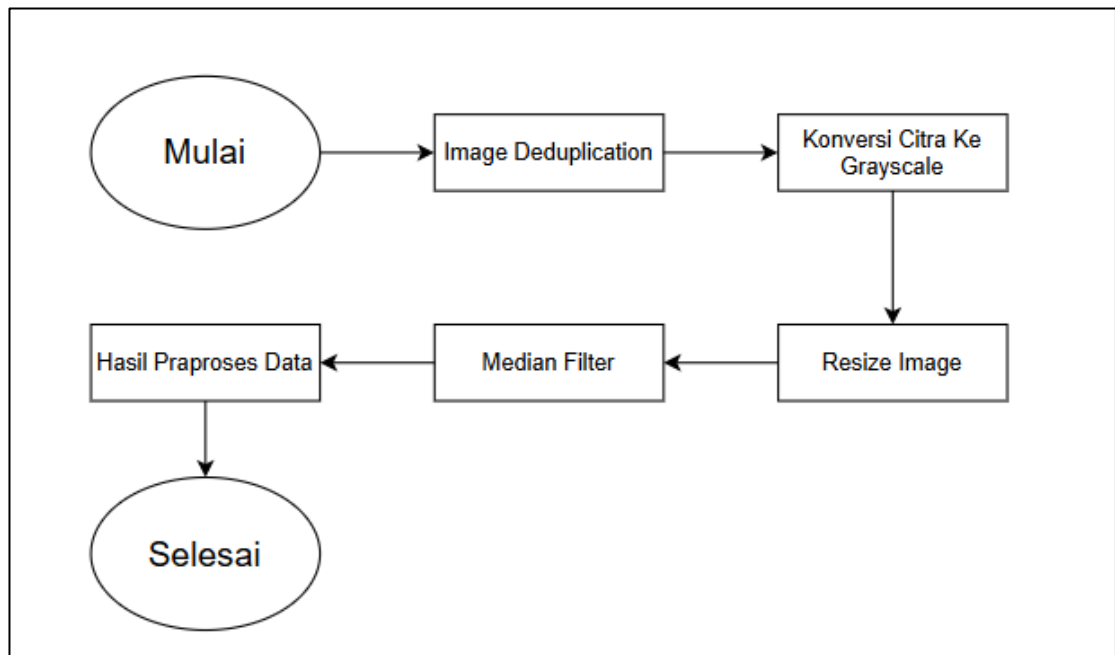


Gambar 3. 4 Citra Endoskopi Normal-Z-Line

Dataset yang terstruktur dan representatif ini memainkan peran penting dalam memastikan validitas penelitian. Selain memudahkan analisis awal, visualisasi ini juga membantu dalam proses pengembangan model klasifikasi dengan memberikan wawasan tentang tantangan dan pola-pola yang mungkin memengaruhi hasil. Harapannya, penggunaan dataset ini dapat menghasilkan model klasifikasi yang memiliki akurasi tinggi dan mampu mengidentifikasi pola-pola secara andal.

3.4 Praproses Data

Dari data citra penyakit esofagitis pada endoskopi, dilakukan tahapan praproses data untuk meningkatkan kualitas citra dan mempermudah proses analisis serta klasifikasi. Tahapan-tahapan tersebut terdapat pada gambar 3.5



Gambar 3. 5 Praproses Data

Praproses data dilakukan untuk mengoptimalkan data citra sehingga analisis fitur menggunakan metode Gray Level Co-occurrence Matrix (GLCM) dan algoritma Support Vector Machine (SVM) menjadi lebih efektif.

a. Image Deduplication

Tahapan pertama adalah Image Deduplication, yaitu proses identifikasi dan penghapusan citra duplikat. Langkah ini bertujuan untuk menghemat ruang penyimpanan, meningkatkan efisiensi pengelolaan data, dan mengurangi risiko analisis yang terganggu akibat data yang redundan. Data yang unik memungkinkan analisis yang lebih akurat dan efisien.

b. Konversi Citra ke Grayscale

Langkah selanjutnya bertujuan untuk mengubah citra berwarna menjadi citra skala abu-abu dengan rentang intensitas piksel dari 0 hingga 255. Proses ini dilakukan untuk menyederhanakan informasi visual pada citra, mengurangi kompleksitas data, dan mempermudah ekstraksi fitur tekstur menggunakan metode Gray Level Co-occurrence Matrix (GLCM).

Citra grayscale lebih efisien dalam analisis karena hanya berfokus pada informasi intensitas piksel tanpa dipengaruhi oleh komponen warna.

c. Resize Image

Tahap berikutnya adalah Resize Image, di mana semua citra diubah ke dimensi yang seragam. Ukuran yang seragam diperlukan untuk memastikan konsistensi dalam analisis dan efisiensi pemrosesan data. Dengan mengurangi perbedaan resolusi, langkah ini juga membantu mencegah kesalahan saat fitur diekstraksi atau saat data dimasukkan ke dalam model klasifikasi.

d. Median Filter

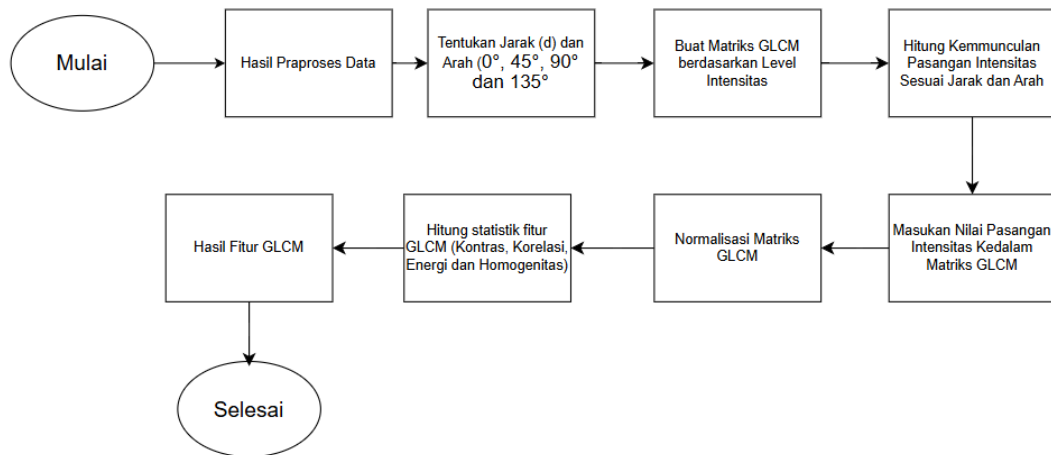
Langkah terakhir, dilakukan Median Filter, yaitu teknik pengurangan noise untuk meningkatkan kualitas citra. Filter ini bekerja dengan menggantikan setiap piksel dengan median nilai piksel di sekitarnya, sehingga hasilnya lebih halus tanpa kehilangan detail penting. Tahap ini sangat berguna untuk meningkatkan akurasi analisis fitur, terutama dalam lingkungan citra yang cenderung memiliki gangguan atau artefak visual.

Secara keseluruhan, setiap tahapan dalam proses praproses ini saling mendukung untuk memastikan data citra yang digunakan memenuhi syarat analisis, baik dari segi kualitas maupun konsistensi. Proses yang sistematis ini memungkinkan analisis fitur dan klasifikasi penyakit menjadi lebih akurat dan andal.

3.5 GLCM

Setelah melakukan praproses data, tahap selanjutnya adalah ekstraksi fitur Gray Level Co-occurrence Matrix (GLCM). Ekstraksi fitur GLCM bertujuan untuk membantu dalam membedakan citra endoskopi yang menunjukkan penyakit Esophagitis dan yang normal. GLCM menghitung frekuensi pasangan intensitas piksel yang berdekatan dengan jarak dan arah tertentu, dan menghasilkan informasi tekstur seperti kontras, korelasi, energi, dan homogenitas. Fitur-fitur ini akan menjadi vektor fitur GLCM yang mewakili tekstur pada citra endoskopi. Fitur-fitur

ini nantinya akan digunakan sebagai input untuk klasifikasi penyakit menggunakan algoritma Support Vector Machine (SVM). Proses ekstraksi fitur GLCM dapat dilihat pada Gambar 3.6.



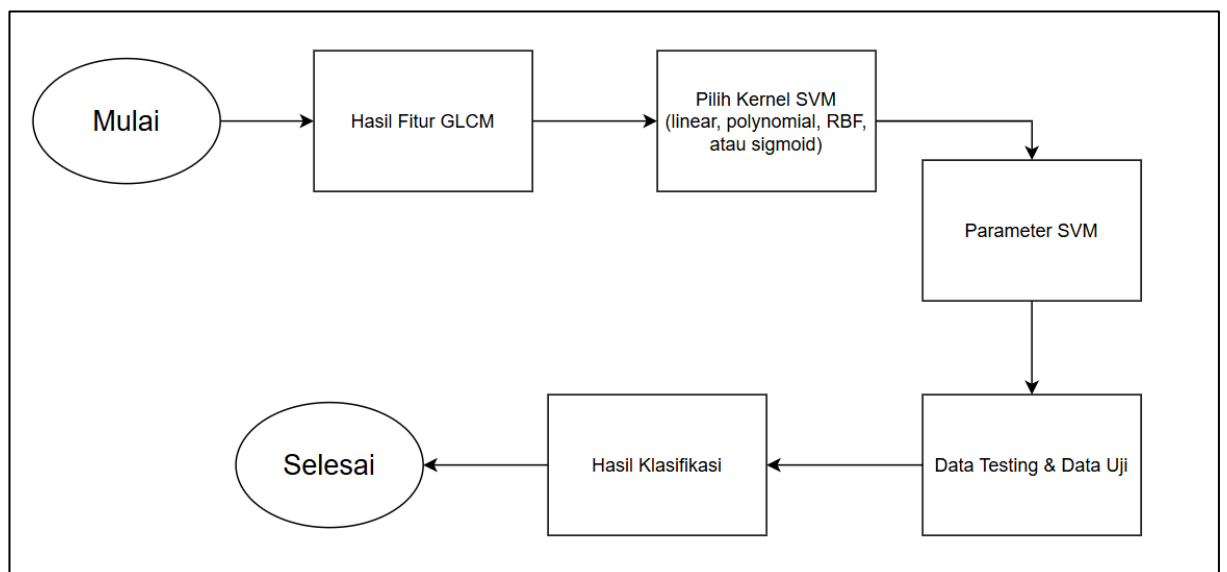
Gambar 3. 6 Alur Proses GLCM

Pada gambar 3.6 merupakan alur dari ekstraksi fitur GLCM. Langkah awal dimulai dengan melakukan praproses data citra. Selanjutnya, ditentukan arah (0° , 45° , 90° , dan 135°) serta jarak (d) antar piksel untuk mengidentifikasi pasangan intensitas yang akan dihitung. Setelah parameter ini ditentukan, dilakukan pembuatan matriks GLCM berdasarkan tingkat intensitas piksel dalam citra. Jika intensitas citra memiliki rentang (0–255). Langkah berikutnya adalah menghitung kemunculan pasangan intensitas piksel sesuai dengan jarak dan arah yang telah ditetapkan sebelumnya. Hasil perhitungan tersebut kemudian dimasukkan ke dalam matriks GLCM untuk membentuk hubungan spasial antar piksel dalam citra. Setelah matriks GLCM terbentuk, dilakukan proses normalisasi matriks GLCM untuk memastikan nilai dalam matriks memiliki skala probabilitas yang seragam dengan rentang antara 0–1. Normalisasi ini juga membantu meningkatkan stabilitas dalam perhitungan fitur statistik. Dari matriks GLCM yang telah dinormalisasi, dilakukan perhitungan fitur statistik GLCM, seperti Kontras, Korelasi, Energi dan Homogenitas. Empat fitur statistik ini kemudian menjadi hasil dari proses ekstraksi

fitur GLCM. Fitur-fitur tersebut nantinya dapat digunakan dalam proses klasifikasi, menggunakan metode Support Vector Machine (SVM), untuk membedakan antara citra yang normal dan citra yang mengalami penyakit esofagitis.

3.6 Klasifikasi SVM

Setelah menyelesaikan ekstraksi fitur GLCM, proses selanjutnya adalah klasifikasi menggunakan metode Support Vector Machine (SVM). Data yang digunakan dalam proses klasifikasi ini berupa vektor fitur yang diperoleh dari hasil ekstraksi fitur GLCM, seperti Kontras, Korelasi, Energi, dan Homogenitas. Seluruh data latih yang telah disiapkan memiliki kelas tertentu yang digunakan sebagai acuan untuk membangun model klasifikasi SVM. Model ini akan memprediksi kelas dari data uji berdasarkan fitur yang diberikan. Gambaran umum dari proses klasifikasi menggunakan SVM ditunjukkan pada gambar 3.7.



Gambar 3. 7 Klasifikasi SVM

Pada gambar 3.7, tahap pertama dari proses klasifikasi adalah pemilihan kernel SVM. Kernel merupakan fungsi matematika yang menentukan cara data ditransformasikan ke dimensi yang lebih tinggi untuk menemukan hyperplane yang optimal. Beberapa kernel yang sering digunakan adalah linear, polynomial, Radial Basis Function (RBF), dan sigmoid. Pemilihan kernel yang tepat sangat penting

karena memengaruhi kinerja model, tergantung pada kompleksitas dan karakteristik data.

Setelah kernel dipilih, langkah berikutnya adalah mengatur parameter SVM, seperti nilai regularisasi (C) yang mengontrol tingkat overfitting dan gamma untuk kernel non-linear. Parameter ini harus dioptimalkan agar model dapat bekerja secara efisien dan memberikan hasil klasifikasi yang akurat. Proses ini sering kali melibatkan teknik seperti grid search atau cross-validation untuk mencari kombinasi parameter terbaik.

Setelah parameter diatur, data latih digunakan untuk membangun model SVM. Model ini akan belajar dari data latih, di mana setiap vektor fitur memiliki label kelas tertentu. Selanjutnya, data uji digunakan untuk mengukur kinerja model dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya. Proses ini menghasilkan prediksi kelas berdasarkan pola yang telah dipelajari oleh model.

Hasil akhir dari proses klasifikasi adalah prediksi kelas untuk data uji. Prediksi ini menunjukkan kategori data, seperti normal atau penyakit esofagitis. Langkah ini merupakan bagian terakhir dari sistem berbasis fitur GLCM dan SVM, yang menghasilkan informasi penting untuk analisis dan pengambilan keputusan medis. Dengan pendekatan ini, akurasi diagnosis dapat ditingkatkan, sehingga mendukung penanganan penyakit yang lebih efektif.

3.7 Evaluasi Hasil Klasifikasi

Confusion matrix merupakan alat yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan hasil prediksi model terhadap nilai sebenarnya dari data uji. Matriks ini terdiri dari empat komponen utama: True Positive (TP), False Negative (FN), False Positive (FP), dan True Negative (TN). True Positive (TP) menggambarkan kondisi ketika model memprediksi positif dan hasil sebenarnya juga positif. Sebaliknya, False Negative (FN) terjadi ketika model memprediksi negatif, tetapi hasil sebenarnya adalah positif.

False Positive (FP) muncul ketika model memprediksi positif, namun hasil sebenarnya adalah negatif. Situasi ini sering disebut sebagai kesalahan tipe I.

Sementara itu, True Negative (TN) terjadi ketika model memprediksi negatif dan hasil sebenarnya juga negatif. Keempat komponen ini memberikan dasar untuk memahami di mana kesalahan model terjadi dan seberapa sering model membuat prediksi yang benar.

Tabel 3. 1 Confussion Matrix

		PREDIKSI	
		POSITIF	NEGATIF
AKTUAL	POSITIF	True Positif (TP)	False Negatif (FN)
	NEGATIF	False Positif (FP)	True Negatif (TN)

Melalui confusion matrix, berbagai metrik evaluasi dapat dihitung untuk mengukur kinerja model. Akurasi, misalnya, mengukur seberapa sering model memberikan prediksi yang benar secara keseluruhan. Presisi digunakan untuk mengetahui seberapa akurat model dalam memprediksi kelas positif, sementara recall menggambarkan seberapa baik model dalam menemukan semua data positif yang sebenarnya. F1-Score, di sisi lain, adalah rata-rata harmonis antara presisi dan recall, yang berguna ketika ada ketidakseimbangan kelas dalam data.

Dengan memahami setiap komponen pada confusion matrix, peneliti dapat mengidentifikasi kelemahan model dan menentukan langkah perbaikan yang diperlukan. Secara keseluruhan, confusion matrix bukan hanya alat untuk mengevaluasi performa model klasifikasi, tetapi juga memberikan wawasan yang lebih mendalam tentang perilaku prediksi model. Dengan memahami elemen-elemen dalam confusion matrix, peneliti dapat meningkatkan kualitas dan efektivitas model dalam menyelesaikan tugas klasifikasi yang dihadapi.

3.8 Skenario Uji Coba

Dalam penelitian ini, dilakukan upaya untuk mencapai hasil klasifikasi dengan tingkat akurasi setinggi mungkin. Oleh karena itu, dilakukan serangkaian skenario uji coba untuk menguji algoritma Support Vector Machine (SVM) berdasarkan pembagian dataset menjadi data latih dan data uji. Skenario pembagian dataset yang digunakan meliputi data latih sebesar 80% dengan data uji 20%, serta data latih sebesar 70% dengan data uji 30%. Pembagian ini bertujuan untuk mengevaluasi performa algoritma SVM dalam berbagai konfigurasi dataset guna memperoleh hasil klasifikasi yang optimal.

Tabel 3. 2 Skenario Uji Coba

Metode	Data Latih	Data Uji	Akurasi (%)	Presisi (%)	Recall (%)	F1 – Score (%)
GLCM & Support Vector Machine (SVM)	80	20				
	70	30				

Tabel 3.2 menunjukkan perbandingan hasil klasifikasi pada kedua skenario pembagian data latih dan uji. Pada skenario pertama, di mana 80% data digunakan untuk pelatihan dan 20% untuk pengujian, diharapkan model dapat mencapai akurasi yang lebih tinggi karena lebih banyak data digunakan untuk melatih model.

Pada skenario kedua menggunakan 70% data latih dan 30% data uji. Dengan proporsi data uji yang lebih besar, model diuji dengan lebih banyak data baru, sehingga performanya mencerminkan kemampuan generalisasi yang lebih baik. Jika model tetap menunjukkan akurasi, presisi, recall, dan F1-score yang konsisten, ini menandakan bahwa model cukup stabil dan tidak mengalami overfitting pada data latih.

Setiap skenario akan menghasilkan nilai metrik evaluasi yang meliputi akurasi, presisi, recall, dan F1-score. Akurasi mengukur seberapa sering prediksi model benar, sedangkan presisi melihat sejauh mana prediksi positif benar-benar akurat. Recall mengukur kemampuan model untuk menangkap semua sampel positif, sementara F1-score menggambarkan keseimbangan antara presisi dan recall.

Melalui analisis hasil dari kedua skenario, peneliti dapat membandingkan performa model dan memilih konfigurasi terbaik untuk diterapkan dalam klasifikasi citra endoskopi. Dengan pendekatan ini, penelitian diharapkan dapat memberikan kontribusi yang signifikan dalam pengembangan sistem pendeteksian penyakit berbasis citra medis menggunakan metode SVM dan fitur GLCM.

REFERENSI

- [1]. Bernanthos, I. N. O., & Handjari, D. R. Hubungan antara Profil Histomorfologik dengan Gambaran Endoskopi Esofagitis Refluks pada Dewasa.
- [2]. Rusiana, D. (2023). HUBUNGAN PELAKSANAAN EDUKASI PERSIAPAN ENDOSKOPI TERHADAP KEPATUHAN PASIEN MELAKSANAKAN PERSIAPAN ENDOSKOPI DI RS MANDAYA ROYAL PURI. *JURNAL KESEHATAN*, 9(1), 50-70.
- [3]. Thaariq, M. A., Baskara, M. D. M., Chaniago, R. A., Christin, D., & Ernawati, I. (2024, August). Systematic Literature Review: Analisis Penerapan Kecerdasan Buatan Dalam Bidang Kesehatan. In *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya* (Vol. 5, No. 1, pp. 168-173).
- [4]. Salsabila, A., Yunita, R., & Rozikin, C. (2021). Identifikasi Citra Jenis Bunga menggunakan Algoritma KNN dengan Ekstraksi Warna HSV dan Tekstur GLCM. *Technomedia Journal*, 6(1 Agustus), 124-137.
- [5]. Mulyana, D. I., & Wibowo, D. R. (2023). Implementasi Tingkat Kematangan Buah Monk Dengan Menggunakan Ekstraksi Gray-Level Co-Occurrence Matrix (GlcM) Dan Support Vector Machine (SVM). *Jurnal Informatika Teknologi dan Sains (Jinteks)*, 5(3), 334-339.
- [6]. Eldo, H., Ayuliana, A., Suryadi, D., Chrisnawati, G., & Judijanto, L. (2024). Penggunaan Algoritma Support Vector Machine (SVM) Untuk Deteksi Penipuan pada Transaksi Online. *Jurnal Minfo Polgan*, 13(2), 1627-1632.
- [7]. Nurrohman, A. (2024). *KLASIFIKASI KOLITIS ULSERATIF PADA CITRA ENDOSKOPI MENGGUNAKAN SUPPORT VECTOR MACHINE DENGAN EKSTRAKSI CITRA GRAY LEVEL CO-OCCURANCE MATRIX* (Doctoral dissertation, Universitas Siliwangi).
- [8]. Adawiyah, R., & Mulyana, D. I. (2022). Optimasi Deteksi Penyakit Kulit Menggunakan Metode Support Vector Machine (SVM) dan Gray Level Co-occurrence Matrix (GLCM). *INFORMASI (Jurnal Informatika dan Sistem Informasi)*, 14(1), 18-33.
- [9]. Thohir, M., Foeady, A. Z., Novitasari, D. C. R., Arifin, A. Z., Phiadelvira, B. Y., & Asyhar, A. H. (2020, February). Classification of colposcopy data using GLCM-SVM on cervical cancer. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)* (pp. 373-378). IEEE.
- [10]. Saputra, R. A., Puspitasari, D., & Baidawi, T. (2022). Deteksi kematangan buah melon dengan algoritma support vector machine berbasis ekstraksi fitur glcm. *Jurnal Infortech*, 4(2), 200-206.

- [11]. Wati, R. A., Irsyad, H., & Rivan, M. E. A. R. (2020). Klasifikasi Pneumonia Menggunakan Metode Support Vector Machine. *J. Algoritm*, 1(1), 21-32.
- [12]. Setianingrum, A. H., & Masruroh, S. U. (2020, October). Performance of acne type identification using glcm and svm. In *2020 8th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-4). IEEE.
- [13]. Ajjah, B. F. F., Mamfaluti, T., & Putra, T. R. I. (2020). Hubungan pola makan dengan terjadinya gastroesophageal reflux disease (GERD). *Journal of Nutrition College*, 9(3), 169-179.
- [14]. Ratna, S. (2020). Pengolahan citra digital dan histogram dengan phyton dan text editor pycharm. *Technologia: Jurnal Ilmiah*, 11(3), 181-186.
- [15]. Munantri, N. Z., Sofyan, H., & Florestiyanto, M. Y. (2020). Aplikasi Pengolahan Citra Digital Untuk Identifikasi Umur Pohon. *Telematika: Jurnal Informatika dan Teknologi Informasi*, 16(2), 97-104.
- [16]. Wijoyo, A., Saputra, A. Y., Ristanti, S., Sya'Ban, S. R., Amalia, M., & Febriansyah, R. (2024). Pembelajaran Machine Learning. *OKTAL: Jurnal Ilmu Komputer dan Sains*, 3(02), 375-380.
- [17]. Sarimole, F. M., & Syaeful, A. (2022). Classification of Durian Types Using Features Extraction Gray Level Co-Occurrence Matrix (Glcm) and K-Nearest Neighbors (Knn). *Journal of Applied Engineering and Technological Science (JAETS)*, 4(1), 111-121.
- [18]. Subandi, R., & Yudhana, A. (2023). Pre-Processing Pada Klasifikasi Citra Medis Pneumonia. *Decode: Jurnal Pendidikan Teknologi Informasi*, 4(1), 86-93.
- [19]. Achmad, Y. F., Yulfitri, A., & Ulum, M. B. (2021). Identifikasi Jenis Jerawat Berdasarkan Tekstur Menggunakan GLCM dan Backpropagation. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)*, 20(2), 139-146.
- [20]. Marwan, F. A. T., Ramadhanti, H. N., Wahid, N. C., & Rimantho, D. (2023). Aplikasi Grey Level Coocurent Matrix (Glcm) Menggunakan Matlab Gui Dan Ann Dalam Identifikasi Penyakit Pada Tanaman Porang. *Jurnal Teknologi Elekerika*, 20(2), 128-132.
- [21]. Wafa, H. S., Hadiana, A. I., & Umbara, F. R. (2022). Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM). vol, 4, 40-45.
- [22]. Yanti, N. P. D. T., & Asana, I. M. D. P. (2023). Sistem Klasifikasi Pengajuan Kredit dengan Metode Support Vector Machine (SVM). *Jurnal Sistem Cerdas*, 6(2), 123-133.
- [23]. Raysyah, S., Arinal, V., & Mulyana, D. I. (2021). Klasifikasi tingkat kematangan buah kopi berdasarkan deteksi warna menggunakan metode knn dan pca. *JSiI (Jurnal Sistem Informasi)*, 88-95.

- [24]. Pratiwi, B. P., Handayani, A. S., & Sarjana, S. (2020). Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix. *Jurnal Informatika Upgris*, 6(2).
- [25]. Christopher, A., & Mulyana, T. M. S. (2022). Klasifikasi Tumbuhan Angiospermae Menggunakan Algoritma K-Nearest Neighbor Berdasarkan Pada Bentuk Daun. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 7(4), 1233-1243.