الجمهورية العربية السورية
جامعة دمشق
كلية الهندسة الميكانيكية والكهربائية
قسم هندسة الحواسيب والأتمتة

تطوير خوارزمية لتجزئة صور العقيدات الرئوية من أجل الكشف المبكر عن سرطان الرئة

رسالة مقدمة لنيل درجة الماجستير في هندسة الحواسيب وشبكاتها في قسم هندسة الحواسيب والأتمتة بكلية الهندسة الميكانيكية والكهربائية

إعداد الطالب:

م. محمد جميل عيوش

المشرف العلمي:

أ.م.د. مأمون يونس

أستاذ مساعد في قسم هندسة الحواسيب والأتمتة

المشرف المشارك:

أ.م.د. أيمن صابوني

أستاذ مساعد في قسم الهندسة الطبية

٢٠٢٢-٢٠٢٣ م

الشكر والتقدير

عرفاناً بالجميل وتقديراً للجهد والبذل والعطاء أتوجه بأسمى آيات التقدير والاحترام للأساتذة والسادة أعضاء الهيئة التدريسية في قسم هندسة الحواسيب والأتمتة بجامعة دمشق وأخص بالذكر:

الدكتور المهندس **مأمون يونس** الذي تفضل مشكوراً بالإشراف على هذا البحث وتقديم الكثير من الدعم والتوجيهات التي ساهمت في إغنائه، فله مني كل الشكر والتقدير وعسى الله أن يجزيه عني كل خير.

والشكر موصول للدكتور المهندس أيمن صابوني على الدعم المقدم من قبله وللسادة أعضاء اللجنة العلمية الموكل إليها تحكيم هذا البحث على ما أولوه من اهتمامات وملاحظات تخدم البحث.

كما أتقدم بالشكر الجزيل إلى عائلتي وأصدقائي على حبهم ودعمهم وتشجيعهم الدائم وإيمانهم بي، وإلى كل من وقف إلى جانبي خلال مراحل عملي.

الملخص

أدى دخول الحاسب إلى تطوير تقنيات جديدة في العديد من المجالات كالمجال الطبي، وساعدت الأطباء في كشف الأمراض وتشخيصها بدقة ومصداقية، إذ تؤدي خبرة الطبيب بالإضافة إلى دقة الحاسب إلى قدرة تشخيص عالية وتسهم بشكل كبير في نجاح الجراحات وإنقاذ الكثير من الأرواح.
تم تصنيف سرطان الرئة في المرتبة الأولى من قائمة السرطانات الأكثر انتشاراً في العالم، ويتطلب اتخاذ قرار علاجي كاستئصال العقيدات الرئوية المعرفة المسبقة بأبعاد ومواصفات هذه العقيدات، لذلك كان لابد من العمل على تطوير طرق جديدة آلية لاستخراج هذه الأورام وتحديد خصائصها.

يهدف البحث إلى تطوير خوارزمية جديدة لاكتشاف أورام الرئة وتصنيفها في الصور المقطعية المحوسبة بالاعتماد على تقنيات معالجة الصور الرقمية وخوارزميات الذكاء الاصطناعي. وتتضمن **الخوارزمية المقترحة** أربع مراحل أساسية، إذ يتم أولاً إجراء المعالجة الأولية للصور وذلك باستخدام تقنية تحويل المويجات لتقليل الضجيج في الصور، بينما يتم تجزئة الصور المحسنة في المرحلة الثانية لفصل الأجزاء المختلفة، ما يزيد من قدرة تقسيم الورم في المراحل اللاحقة، وهنا يتم تطبيق خوارزمية العنقدة Fuzzy C-Mean، وتحديد عتبة otsu التي تضمن الاستخراج الصحيح لجميع الكائنات في الصور نظراً لأن الورم له مستويات رمادية مختلفة في كل صورة. وتم تطبيق العمليات المورفولوجية (التمدد والتعرية) هنا أيضاً لإزالة بعض الكائنات غير المرغوبة من صور العتبة. وأخيراً، فإنه يتم الحصول على منطقة الورم عن طريق طرح الصورتين الناتجتين عن عمليتي التعرية والتمدد. وفي المرحلة الثالثة تم استخلاص أهم 6 سمات مثلى لكي يتم الاعتماد عليها في التصنيف والتشخيص. أما المرحلة الرابعة، فتتضمن تصنيف الأورام المستخرجة بوصفها عقيدات سليمة أو خبيثة وذلك عن طريق خوارزمية متجه دعم الآلة SVM.
وتم اختبار خوارزمية التجزئة على قاعدة بيانات تتكون من 137 صورة ورمية وغير ورمية وحققت دقة عالية وصلت حتى 96.2٪، كما وجدت النتائج التجريبية لتطبيق نظام التصنيف على 24 حالة عشوائية (تحوي 29 ورماً)، لم يتم التدريب عليها أن النظام حقق دقة عالية أيضا بنسبة 93.1%.

الكلمات المفتاحية:

الكشف الآلي عن سرطان الرئة، تحويل المويجات، معالجة الصور الرقمية وتجزئتها، خوارزمية العنقدة FCM، خوارزمية متجه الدعم الآلي SVM، تصنيف العقيدات الرئوية.

الفهرس

الموضوع	الصفحة
الملخص	١
الفهرس	٢
فهرس الأشكال	٦
فهرس الجداول	٨
قائمة المصطلحات	٩
قائمة المختصرات	١٠
الفصل الأول: المقدمة INTRODUCTION	١١
١-١-مقدمة	١٢
١-٢-الدراسات المرجعية	١٣
١-٢-١-نتائج الدراسات المرجعية السابقة في البحث	١٧
١-٣-مشكلة البحث	١٧
١-٤-فرضيات البحث	١٧
١-٥-أهمية البحث	١٨
١-٦-أهداف البحث	١٨
١-٧-تنظيم البحث	١٨
١-٨-خلاصة	١٩
الفصل الثاني: لمحة نظرية عن سرطان الرئة وطرائق تشخيصه	٢٠
٢-١-مقدمة	٢١
٢-٢-تشريح الرئتين والجهاز التنفسي	٢١
٢-٣-سرطان الرئة Lung Cancer	٢١
٢-٤-الفروقات الشكلية بين أنواع الأورام	٢٢
٢-٥-تشخيص سرطان الرئة	٢٣
٢-٥-١-الفحص الطبي والتاريخ المرضي	٢٣

٢-٥-٢-تنظير القصبات الهوائية Bronchoscopy	٢٣
٢-٥-٣-الخزعة biopsy	٢٤
٢-٥-٤-التشخيص باستخدام الأشعة السينية X-Ray	٢٤
٢-٥-٥-التشخيص بالتصوير المقطعي المحوسب Computerized Tomography (CT)	٢٤
٢-٥-٦-التصوير بالرنين المغناطيسي Magnetic resonance imaging (MRI)	٢٥
٢-٥-٧-التصوير المقطعي بالإصدار البوزيتروني Positron emission tomography (PET)	٢٦
٢-٦-خلاصة	٢٦
الفصل الثالث: المبادئ النظرية للبحث Theoretical Principles of Research	٢٧
٣-١-مقدمة	٢٨
٣-٢-التشخيص بمساعدة الحاسب Computer-Aided Diagnosis (CAD)	٢٨
٣-٣-نظام تشخيص أو تحديد الأمراض بمساعدة الحاسب CAD	٢٨
٣-٤-معالجة الصور الرقمية Digital Image processing	٢٩
٣-٥-تحويل المويجات Wavelet Transform	٢٩
٣-٥-١-تحليل الصورة وإعادة تركيبها في التحويل المويجي	٣٠
٣-٥-٢-تحويل الصورة متعدد المستويات (شجرة التحليل المويجي)	٣٢
٣-٥-٣-أشهر أنواع دوال المويجة	٣٣
٣-٥-٤-أهمية تحويل المويجات في الصورة	٣٤
٣-٥-٥-إزالة الضجيج باستخدام المويجات	٣٤
٣-٦-المعالجة باستخدام العمليات المورفولوجية Morphological operations	٣٦
٣-٧-تجزئة الصورة Image Segmentation	٣٦
٣-٧-١-التجزئة القائمة على الحواف Edge based Thresholding	٣٧
٣-٧-٢- التجزئة باستخدام العتبة Thresholding Segmentation	٣٧
٣-٧-٢-١-أهم طرق التعتيب الأكثر شيوعاً	٣٨
٣-٧-٣-التجزئة القائمة على المنطقة Region based Thresholding	٤٠
٣-٧-٣-١-عنقدة البيانات Data Clustering	٤٠
٣-٧-٣-٢-خوارزميات عنقدة البيانات	٤٠

٣-٨-الذكاء الاصطناعي	٤٢
٣-٨-١-تعلم الآلة Machine Learning	٤٣
٣-٨-٢-أنواع تعلم الآلة	٤٣
٣-٨-٣-خوارزمية متجه دعم الآلة Support Vector Machine	٤٣
٣-٨-٣-١-أنواع متجهات دعم الآلة	٤٦
٣-٨-٣-٢-إيجابيات خوارزمية متجه الدعم الآلي وسلبياتها	٤٧
٣-٨-٣-٣-استخدامات خوارزمية متجه الدعم الآلي	٤٨
٣-٩-خلاصة	٤٨
الفصل الرابع: الخوارزمية المقترحة للكشف عن العقيدات الرئوية	٤٩
٤-١-مقدمة	٥٠
٤-٢-استحصال الصورة Image Acquisition	٥١
٤-٣-المعالجة الأولية للصورة	٥١
٤-٣-١-إزالة الضجيج باستخدام تحويل المويجات	٥١
٤-٣-٢-خطوات إزالة الضجيج باستخدام تحويل المويجات	٥٢
٤-٤-تجزئة الصورة Image Segmentation	٥٤
٤-٤-١-التعتيب باستخدام Otsu Thresholding	٥٨
٤-٤-٢-استخلاص الأورام من الصور المجزأة باستخدام العمليات المورفولوجية	٥٨
٤-٤-٣-خوارزمية العنونة Labeling	٦١
٤-٥-استخراج السمات Feature Extraction	٦٢
٤-٥-١-تقليل متجه السمات المستخرجة	٦٥
٤-٦- مرحلة التصنيف Classification	٦٦
٤-٦-١-خطوات التصنيف باستخدام مصنف متجه دعم الآلة SVM	٦٦
٤-٧-خلاصة	٦٧
الفصل الخامس: تقييم الأداء Performance Evaluation	٦٨
٥-١-مقدمة	٦٩
٥-٢-نموذج تقييم الأداء	٦٩

٥-٢-١-تقييم أداء خوارزمية التجزئة	٧١
٥-٢-١-١-تقييم أداء خوارزمية التجزئة على مستوى الصورة	٧١
٥-٢-١-٢-تقييم أداء خوارزمية التجزئة على مستوى قاعدة البيانات	٨٣
٥-٢-٢-تقييم أداء خوارزمية التصنيف	٨٤
٥-٣-النتائج والمناقشة	٩٣
٥-٤-الاستنتاجات	٩٣
٥-٥-المقترحات والتوصيات	٩٤
٥-٦-خلاصة	٩٤
المراجع	٩٥

فهرس الأشكال

الشكل	الصفحة
الشكل (١-١) أمثلة عن التصوير المحوسب CT للرئة a) المقطع الجانبي، b) المقطع الرأسي	١٢
الشكل (٢-١) الفروقات الشكلية بين الأورام الحميدة والخبيثة	٢٣
الشكل (٢-٢) تصوير الرئة بالأشعة السينية	٢٤
الشكل (٢-٣) التصوير المقطعي المحوسب للرئتين	٢٥
الشكل (٢-٤) الفرق بين التصوير بالأشعة السينية والتصوير المقطعي المحوسب للرئتين	٢٥
الشكل (٣-١) التحويل المويجي	٣٠
الشكل (٣-٢) مرحلة تفكيك الصورة إلى مركباتها الفرعية	٣١
الشكل (٣-٣) مرحلة التركيب والحصول على الصورة الأصلية	٣١
الشكل (٣-٤) نتيجة تطبيق تقنية تحويل المويجات على صورة ما	٣١
الشكل (٣-٥) التحويل المويجي متعدد المستويات (ثلاث مراحل)	٣٢
الشكل (٣-٦) تقنية التحويل المويجي متعدد المستويات	٣٢
الشكل (٣-٧) التابع المويجي لمويجة Haar	٣٣
الشكل (٣-٨) التوابع المويجية لمويجات Daubechies	٣٣
الشكل (٣-٩) العتبة القاسية والمرنة	٣٥
الشكل (٣-١٠) خطوات خوارزمية عتبة Otsu	٣٩
الشكل (٣-١١) العناقيد الناتجة عن تطبيق خوارزمية FCM	٤٢
الشكل (٣-١٢) آلية عمل SVM (a) للفصل بين صنفين، (b) للفصل بين أكثر من صنفين	٤٤
الشكل (٣-١٣) متجهات الدعم الواقعة على الحد الفاصل بين البيانات في الفضاء	٤٤
الشكل (٣-١٤) متجهات الدعم	٤٥
الشكل (٣-١٥) الهوامش	٤٥
الشكل (٣-١٦) تقييم المستوي الفائق الأفضل واختباره	٤٦
الشكل (٣-١٧) فصل مجموعة البيانات خطيا (بواسطة خط مستقيم)	٤٧
الشكل (٣-١٨) فصل مجموعة بيانات غير قابلة للفصل الخطي بواسطة مستوٍ	٤٧
الشكل (٤-١) خطوات نظام الكشف عن العقيدات المقترح	٥٠
الشكل (٤-٢) عينة بسيطة من قاعدة البيانات المستخدمة	٥١

الشكل (٤-٣) مراحل إزالة الضجيج باستخدام تقنية تحويل المويجات	٥٢
الشكل (٤-٥) ناتج تطبيق عدة أنواع للمويجات على إحدى عينات قاعدة البينات	٥٣
الشكل (٤-٦) ناتج إجراء عملية التحليل المويجي باستخدام نوع المويجة db4 عند عدة مستويات	٥٤
الشكل (٤-٧) خطوات خوارزمية FCM المقترحة	٥٥
الشكل (٤-٨) مخطط بياني يبين قيم تابع التكلفة خلال تكرار تنفيذ خطوات تابع FCM	٥٧
الشكل (٤-٩) ناتج تطبيق العنقدة FCM على الصورة	٥٧
الشكل (٤-١٠) ناتج تطبيق عتبة Otsu على الصورة	٥٨
الشكل (٤-١١) مراحل استخلاص المنطقة الورمية باستخدام العمليات المورفولوجية	٥٩
الشكل (٤-١٢) تطبيق عمليتي (b) التعرية و(c) التمدد على الصورة المجزأة (a)	٦٠
الشكل (٤-١٣) إجراء عملية الطرح من أجل استخلاص الورم	٦٠
الشكل (٤-١٤) ناتج تطبيق عملية الفتح على الصورة المجزأة	٦٠
الشكل (٤-١٥) تطبيق الخوارزمية المقترحة على صورة طبيعية (من دون ورم)	٦١
الشكل (٤-١٦) تطبيق الخوارزمية المقترحة على صورة تحوي ورمين معاً	٦١
الشكل (٤-١٧) مرحلة استخراج السمات	٦٢
الشكل (٤-١٨) استرجاع السويات الرمادية للمنطقة المرشحة	٦٣
الشكل (٤-١٩) السمات المستخدمة في تصنيف الأورام المكتشفة	٦٥
الشكل (٤-٢٠) خطوات استخراج السمات المثلى تمهيداً لتدريب المصنف	٦٦
الشكل (٤-٢١) خطوات التصنيف باستخدام خوارزمية SVM	٦٧
الشكل (٥-١) بارامترات مصفوفة الارتباك	٦٩

فهرس الجداول

الجدول	الصفحة
الجدول (٤-١) قيم تابع التكلفة وقيم التحسن خلال تكرارات التابع FCM	٥٦
الجدول (٤-٢) السمات المستخدمة في تصنيف الأورام المكتشفة	٦٣
الجدول (٤-٣) قيم متجهات السمات الثنائية لبعض الأورام في صور قاعدة البيانات	٦٤
الجدول (٤-٤) قيم متجهات السمات الرمادية لبعض الأورام في صور قاعدة البيانات	٦٤
الجدول (٥-١) صور المناطق المجزأة ذات الورم الحميد	٧٢
الجدول (٥-٢) صور المناطق المجزأة ذات الورم الخبيث	٧٤
الجدول (٥-٣) صور المناطق المجزأة ذات الأورام المتعددة	٧٦
الجدول (٥-٤) صور المناطق السليمة بدون أورام	٧٨
الجدول (٥-٥) قيم عناصر مصفوفة الارتباك لصور المناطق المجزأة ذات الورم الحميد، مع قيم معايير تقييم نموذج التجزئة	٨١
الجدول (٥-٦) قيم عناصر مصفوفة الارتباك لصور المناطق المجزأة ذات الورم الخبيث، مع قيم معايير تقييم نموذج التجزئة	٨١
الجدول (٥-٧) قيم عناصر مصفوفة الارتباك لصور المناطق المجزأة ذات الأورام المتعددة، مع قيم معايير تقييم نموذج التجزئة	٨٢
الجدول (٥-٨) قيم عناصر مصفوفة الارتباك لقاعدة البيانات جميعها	٨٣
الجدول (٥-٩) قيم محددات تقييم أداء خوارزمية التجزئة	٨٣
الجدول (٥-١٠) صور مجموعة بيانات الاختبار ذات الورم الحميد، مع التشخيص الطبي والآلي	٨٦
الجدول (٥-١١) صور مجموعة بيانات الاختبار ذات الورم الخبيث، مع التشخيص الطبي والآلي	٨٨
الجدول (٥-١٢) صور مجموعة بيانات الاختبار ذات الأورام المتعددة، مع التشخيص الطبي والآلي	٩٠
الجدول (٥-١٣) قيم عناصر مصفوفة الارتباك لصنفي الأورام	٩١
الجدول (٥-١٤) قيم محددات تقييم أداء خوارزمية التصنيف	٩١
الجدول (٥-١٥) مقارنة بين الأبحاث المنشورة بين عامي 2016 و2020 مع النظام المقترح	٩٢

قائمة المصطلحات

المصطلح باللغة العربية	المصطلح باللغة الإنكليزية
العقيدات الرئوية	Pulmonary Nodules
الأورام الحميدة / الخبيثة	Benign / Malignant Tumors
الشبكات العصبية التلافيفية	Convolutional Neural Networks
التنظير المجهري للصور الفلورية	Fluorescence LifeTime Endomicroscopic
التصوير بالأشعة السينية	X-Radiography imaging
التصوير المقطعي المحوسب	Computerized Tomography scan
التصوير بالرنين المغناطيسي	Magnetic resonance imaging
التشخيص بمساعدة الحاسب	Computer –Aided Diagnosis
معالجة الصور الرقمية	Digital Image Processing
تحويل فورييه	Fourier Transform
بتحويل فورييه قصير الأمد	Short-Time Fourier Transform
تحويل المويجات	Wavelet Transform
العمليات المورفولوجية	Morphological operations
التجزئة القائمة على الحواف	Edge based Thresholding
التجزئة باستخدام العتبة	Thresholding Segmentation
التجزئة القائمة على المنطقة	Region based Thresholding
عنقدة البيانات	Data Clustering
العنقدة باستخدام C-Mean الضبابية	Fuzzy C-Mean Clustering
الذكاء الاصطناعي	Artificial Intelligence
تعلم الآلة	Machine Learning
متجه دعم الآلة	Support Vector Machine
خوارزمية العنونة	Labeling Algorithm
استخراج السمات	Feature Extraction
التصنيف	Classification
تقييم الأداء	Performance Evaluation

قائمة المختصرات

الاختصار	المصطلح بالإنكليزية
CNN	**C**onvolutional **N**eural **N**etworks
RF	**R**andom **F**orest
DT	**D**ecision **T**ree
MRI	**M**agnetic **R**esonance **I**maging
FLIM	**F**luorescence **L**ifeTime **I**maging **M**icroscopy
X-Ray	**X**-**R**adiography
CT	**C**omputerized Axial **T**omography scan
CAD	Computer-**A**ided **D**iagnosis
FT	**F**ourier **T**ransform
STFT	**S**hort-**T**ime **F**ourier **T**ransform
AI	**A**rtificial **I**ntelligence
ML	**M**achine **L**earning
FCM	**F**uzzy **C**-**M**ean
SVM	**S**upport **V**ector **M**achine

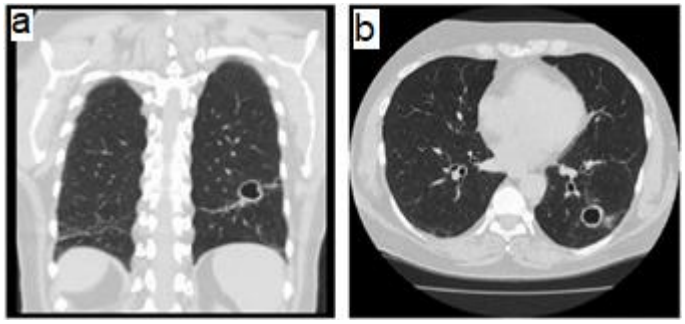
الفصل الأول

المقدمة Introduction

١ – ١ –مقدمة:

السرطان هو عبارة عن مرض يمكن أن يحدث في أي مكان في جسم الإنسان، إذ تبدأ خلايا الجسم في الانقسام من دون توقف وتنتشر في الأنسجة المحيطة.
وسرطان الرئة هو النوع الذي لا يمكن فيه التحقق من نمو الخلايا غير الطبيعية في إحدى الرئتين أو كلتيهما، عندما لا تؤدي هذه الخلايا الشاذة وظائف الخلايا البشرية السليمة، إذ تستمر في النمو والانقسام، مشكلة ما يسمى بالأورام Tumors أو العقيدات Nodules، ما يؤثر في الأداء السليم المنظم للرئة في تزويد جسم الإنسان بالأوكسجين عن طريق الدم [1].

وتظهر هذه العقيدات على شكل تكتلات نسيجية صغيرة في الرئة، وتبدو في صور الأشعة السينية أو فحوصات التصوير المقطعي المحوسب (CT) على شكل ظلال دائرية بيضاء، عادة ما يبلغ حجم العقيدات 30 ميلليمترا. أما العقيدات الأكبر من ذلك، فتعتبر سرطانية أكثر من العقيدات الأصغر.



الشكل (١–١) أمثلة عن التصوير المحوسب CT للرئة a) المقطع الجانبي، b) المقطع الرأسي [1].

تم تصنيف سرطان الرئة أو سرطان الجهاز التنفسي في المرتبة الأولى من قائمة السرطانات الأكثر تشخيصاً وانتشاراً وفتكاً من قبل مؤسسة أبحاث السرطان العالمي الدولي مسبباً عدداً كبيراً من الوفيات، إذ يؤدي إلى 1.61 مليون حالة وفاة كل عام. ومن المتوقع أن يستمر هذا العدد في الزيادة، وسيصل إلى 17 مليون حالة سرطان الرئة بحلول عام 2030 [2].

يزداد خطر الإصابة بسرطان الرئة نتيجة التعرض طويل الأمد لدخان التبغ أو ملوثات الهواء، كما يمكن أن ينمو وينتشر إلى أجزاء أخرى من الجسم، بينما يصبح معدل النجاة أعلى اذا تم تشخيص السرطان في المراحل المبكرة، ولكن لا تعد هذه المهمة بالأمر السهل بسبب عدم قدرة المتخصصين على اكتشاف المنطقة المصابة حتى النقطة التي يصل فيها المرض إلى مرحلة متقدمة، وذلك بسبب صغر وعدم تجانس عقيدات الرئة من حيث التركيب Texture والشكل Shape والكثافة Intensity، بالإضافة إلى درجة التشابه العالية بين العقيدات وتكوينات الأنسجة الحميدة، يتطلب اتخاذ قرار علاجي كاستئصال العقيدات الخبيثة من الرئتين

المعرفة المسبقة والدقيقة بمواصفات وأبعاد هذه العقيدات، لذلك كان لابد من العمل على تطوير طرائق جديدة آلية لاستخراج هذه الأورام وتحديد خصائصها.

وتعد تقنيات الكشف الآلي باستخدام نتائج التحليل الآلي المحوسب للصور الطبية طريقة مساعدة في زيادة دقة تشخيص الطبيب الشعاعي Radiologist، وفي تحقيق الكشف المبكر وتشخيص الآفات المرضية التي تظهر في الصور الطبية مثل أورام الرئة، ونظراً لأهمية الرئة العضوية والوظيفية في جسم الإنسان، فإن الكشف المبكر عن أورامه الخبيثة يزيد من احتمالات نجاح العلاج وبقاء المريض على قيد الحياة. وبالتالي تقدم هذه التقنيات رأياً ثانياً في كشف الأورام والوصول إلى التشخيص الصحيح، مع التأكيد على حقيقة أن الطبيب هو من يقوم بإعطاء التشخيص النهائي [3].

ويتكون نظام التشخيص بمساعدة الحاسوب الذي يتضمن تشخيص أورام الرئة بشكل عام من أربع مراحل، إذ يتم في المرحلة الأولى المعالجة الأولية للصورة ويجري في المرحلة الثانية تحديد المناطق ذات الأهمية (العقيدات الرئوية)، وفي المرحلة الثالثة يتم تحديد الخصائص التي سيتم الاعتماد عليها واستخلاصها في التصنيف والتشخيص، وأما في المرحلة الرابعة، فيتم تصنيف العقيدات فيما إذا كانت سليمة أم خبيثة، وبالتالي تشخيص الحالة المرضية [4].

١-٢-الدراسات المرجعية:

١-**الدراسة الأولى** [5]:

في عام ٢٠٢٠ قام Qiang Wango وآخرون بتصميم تقنية لتمييز سرطان الرئة القائم على التنظير المجهري للصور الفلورية Fluorescence LifeTime Imaging Microscopy (FLIM) باستخدام تعلم الآلة، وذلك بالاعتماد على سمة عمر البكسل بدلاً من السمات الهندسية، ووجد أن خوارزمية الغابات العشوائية تحقق أعلى دقة في تقدير الأنسجة السرطانية أو الطبيعية، بينما كان لخوارزمية الجار الأقرب-K Nearst Neighbour الدقة الأسوأ. على الرغم من النتائج الجيدة التي تم التوصل لها، كان هناك مشكلة فقدان ارتباط البكسلات المجاورة الناتجة عن التجزئة، لذلك كان لا بد من تحسين جودة الصورة المجزأة لزيادة الدقة.

٢-الدراسة الثانية [6]:

في عام ٢٠١٩ قام Ummadi Janardhan Reddy وآخرون بتصنيف سرطان الرئة من الصور الرقمية مباشرة باستخدام الشبكات العصبية Neural Networks (NN) وخوارزمية الجار الأقرب K-Nearst Neighbour (KNN)، وبدون إجراء المعالجة الأولية عليها، كما أنهم اقتصروا على استخراج السمات الرمادية من أجل عملية التصنيف، والتي ستحقق دقة جيدة فقط في حال كانت الأورام تملك أحجام متقاربة.

٣-الدراسة الثالثة [7]:

في عام ٢٠١٩ قام كل من Israa Al_Barazanchi وآخرون بنشر بحث بعنوان تصنيف سرطان الرئة باستخدام التنقيب عن البيانات Data Mining وخوارزميات التعلم الخاضع للإشراف على مجموعة بيانات متعددة الأبعاد. واعتمدت منهجية الدراسة على تجزئة الرئة أولاً ثم استخدام الترشيح المورفولوجي لتصحيح أي ثقوب قد تنجم عن التجزئة. وبعدها تم استخراج مناطق العقيدات المرشحة من خلال سمات الشكل والتركيب. وفي النهاية مرحلة تصنيف العقيدات المرشحة باستخدام مصنف شجرة القرار Tree Decision (DT). وعلى الرغم من أن نتائج مرحلة تجزئة العقيدات كانت دقيقة جداً بما يكفي للتنبؤ بمواقع العقيدات، إلا أنه في بعض الأحيان لم يكن النظام قادراً على توقع ملامح Contours العقيدات بشكل أكثر دقة نظراً لتطبيق بعض العمليات المورفولوجية التي أسفرت عن معالم متآكلة في العقيدات.

٤-الدراسة الرابعة [8]:

في عام ٢٠١٩ قام Murat UCAR بتحسين نتائج التعلم العميق باستخدام مرشح لابلاسيان Laplacian كخطوة أولية في معالجة الصورة، ووجد أن عملية تحسين الصور قبل إدخالها إلى الشبكات العصبية التلافيفية ستزيد من دقة التصنيف.

٥-الدراسة الخامسة [9]:

في عام ٢٠١٩ قام Lakshmanaprabu وآخرون باقتراح طريقة آلية لتصنيف أورام الرئة باستخدام الشبكات العصبية العميقة، حيث تم استخدام تحليل التمييز الخطي Linear Discriminant Anaysis (LDA) لتقليل أبعاد السمات المستخرجة والذي حقق دقة عالية في عملية التصنيف، لكنهم لم يعتمدوا على تجزئة منطقة الورم والتي ستشكل خطوة مهمة وتجعل عملية التشخيص أكثر بساطة.

٦-الدراسة السادسة [10]:

في عام ٢٠١٨ قام كل من Jing Gong وآخرون باقتراح طريقة للكشف التلقائي عن العقيدات الرئوية من خلال دمج ترشيح tensor ثلاثي الأبعاد مع تحليل ميزة الصورة المحلية، إذ تم إجراء معالجة أولية للصور. وتمت تجزئة الرئة من خلال تطبيق عتبة Otsu وخوارزمية نمو المنطقة Region Growing. وبعدها تم تطبيق طريقة ترشيح ثلاثية الأبعاد لتعزيز العقيدات الكروية أو النقطية، مع تحليل ميزة الصور المحلية (باستخدام مؤشري الشكل والانحناء) لتقييم طوبولوجيا سطح كل عقيدة وإزالة المناطق غير المرغوب بها، وأخيراً تم استخراج سمات العقيدات المرشحة لتدريب مصنف الغابات العشوائية Random Forest (RF) على تصنيف المرشحات، واعتمد هذا الأسلوب على فرضية أن العقيدات لها هياكل شبيهة بالكرة ball أو تشبه النقاط dot، لكن يوجد بعض العقيدات غير منتظمة الشكل، وبالتالي سوف يتم حذفها.

٧-الدراسة السابعة [11]:

في عام ٢٠١٨ قام Gur Amrit Pal Singh وآخرون بتصميم وتطوير تقنية لتشخيص أورام الرئة، حيث بدأت الطريقة المقترحة بإزالة الضجيج من الصور الرقمية باستخدام عدة مرشحات في المجال الحيزي (Median, Gaussian, Bilateral)، ثم تمت تجزئة الصور الناتجة وصولاً إلى مرحلة التصنيف باستخدام خوارزمية Multi-Layer Perceptron (MLP)، حققت هذه الطريقة نتائج جيدة ومفيدة والتي من الممكن أن يتم تحسينها بالتحويل إلى الترشيح الترددي كونه يستغرق وقتاً أقصراً مقارنة بالترشيح الحيزي.

٨-الدراسة الثامنة [12]:

في عام ٢٠١٧ أيضاً قام كل من Darren Baker وآخرون بإجراء دراسة حول التنبؤ بحدوث سرطان الرئة من التصوير المقطعي المحوسب CT. وكان الهدف الأولي من الدراسة هو استنتاج خطر الإصابة مباشرة، لكن تم تعديل البحث للتركيز على تحديد مواقع عقيدات الرئة. وعلى الرغم من أن الباحثين لم يتمكنوا من تدريب الشبكة التلافيفية من الصفر على البيانات، إلا أنهم وجدوا أن استخدام نمط نقل التعلم Transfer Learning يعطي دقة جيدة.

ولم ينجح البحث في تدريب كل من الشبكات التلافيفية ثنائية وثلاثية الأبعاد على صور الرئة، ربما بسبب انخفاض نسبة الإشارة إلى الضجيج التي تجعل التشخيص البشري صعباً، والسبب الثاني أن حجم بيانات التدريب كانت صغيرة جداً نسبياً.

٩-الدراسة التاسعة [13]:

في عام ٢٠١٧ قام كل من Lei Fan وآخرون باقتراح نظام تشخيص حاسوبي لاكتشاف سرطان الرئة بالاعتماد على الشبكات العصبية التلافيفية Convolutional Neural Networks (CNN)، وقد تم استخدام خوارزمية التعلم العميق بعد تجزئة منطقة الرئة، وجرى تشكيل قواعد التشخيص من السمات المستخرجة للتصنيف. وأعطى النظام دقة منخفضة، ويتوقع أن سبب تدني هذه الدقة عدم تطبيق تقنيات معالجة الصور على الصور المقطعية للرئتين.

١٠-الدراسة العاشرة [14]:

في عام ٢٠١٧ قام Qing Wu وآخرون باقتراح أسلوب قائم على معالجة الرسم البياني Histogram والشبكات العصبية Neural Networks (NN) للتعرف على سرطان الرئة، حقق الباحثان من خلال هذا الأسلوب دقة لا تعتبر كافية للتنبؤ الدقيق بالمرض، ويفسر ذلك بسبب تدريب الشبكة على حجم بيانات صغير جداً (لا يتجاوز ٥٠ صورة)، بالإضافة إلى استخدام بنية شبكة بسيطة وليست عميقة.

١١-الدراسة الحادية عشر [15]:

في عام ٢٠١٦ قام Ashis Kumar Dhara وآخرون بتصميم نظام CAD وذلك بالاعتماد على خوارزمية تجزئة قائمة على العتبة، كما تم استخلاص العديد من السمات القائمة على الشكل shape والهامش margin والنسيج texture، والتي تم إدخالها كمتجه الدخل إلى مصنف متجه دعم الآلة Support Vector Machine (SVM).

١٢-الدراسة الثانية عشر [16]:

في عام ٢٠١٦ قام كلا من Elmar Rendon-Gonzalez وآخرون بتطوير نظام CAD مشابه للدراسة السابقة، حيث تم إجراء التعتيب أيضاً واستخدام مصنف SVM لتمييز عقيدات الرئة الحميدة عن الخبيثة.

تعاني الدراستان الحادية عشر والثانية عشر من استخدام عدد كبير من السمات غير المفيدة (تصل إلى ١٥ سمة)، والتي ستؤثر سلباً على أداء المصنف SVM كونه لا يحقق الدقة المرغوبة في حال كان متجه السمات كبيراً جداً، لذا كان هناك مجالاً كبيراً للتحسين عن طريق تحديد السمات الأكثر أهمية فقط، بالإضافة إلى اتخاذ خطوة معالجة الصورة كإجراء أمثلي لتحسين جودة الصور المستخدمة في الدراستين السابقتين.

١-٢-١-نتائج الدراسات المرجعية السابقة في البحث:

كل من هذه الدراسات السابقة له ميزاته وسلبياته، لكن بعضها لم يستخدم خوارزمية التجزئة المناسبة بسبب أوجه التشابه في كثافة السويات الرمادية للمناطق الورمية وغير الورمية، وعدم التجانس بين منطقة الرئة Lung Region والهياكل الرئوية Pulmonary structures مثل الشعب الهوائية Bronchi والقصبات Bronchioles والشرايين Arteries والأوردة Veins، وبعضها الآخر لم يستخدم أفضل مصنف لتصنيف عقيدات الرئة من عدد محدود من السمات.

بالإضافة إلى أن هناك أبحاثاً لم تتخذ تقنيات معالجة الصور كخطوة أساسية وهذا يؤثر في دقة النتائج المرجوة لما لها من دور كبير في تحليل الصور الطبية، ويجعل عمليات التصنيف أكثر دقة.

١-٣-مشكلة البحث:

إن أكبر مشكلة واجهت الباحثين في أثناء تنفيذ تقانة تجزئة الصورة هي تداخل الكثافة، إذ تزداد الصعوبة في كشف العقيدات مع ازدياد كثافة الأنسجة كون بنى الأعضاء المختلفة تملك خصائص أنسجة مختلفة، وأيضاً تزيد بعض التغيرات التشريحية التي تختلف من شخص إلى آخر في صعوبة الكشف.

وتحتوي الصور المقطعية المحوسبة وصور الأشعة السينية دائماً على ضجيج ناجم عن أداء المشغل أو بعض التشويش الذي يمكن أن يؤدي إلى أخطاء خطيرة في التشخيص، إذ تختلف دقة هذه الصور بطرق تعيق التحديد، كما يعد اختيار المصنف من التحديات الأساسية في عملية تشخيص عقيدات الرئة الحميدة أو الخبيثة من عدد محدود من السمات، إذ يؤثر اختيار عدد كبير من السمات سلباً على أداء المصنف، وبالتالي الوصول إلى الدقة المرغوبة في التصنيف.

١-٤-فرضيات البحث:

انطلاقاً من النتائج التي تم التوصل لها في الدراسات السابقة، يمكن التغلب على التحديات المذكورة في أنظمة تشخيص سرطان الرئة من دون المساس بمزاياها وفق الطرق التالية:

استخدام طرق تجزئة أكثر كفاءة من الطرق السابقة التي أنتجت صور مجزأة بشكل غير صحيح، إذ احتوت الصور الناتجة مناطق غير مرغوب بها، وأدى بعضها إلى حدوث تشوه في الأشكال. ومن التقنيات التي تعد حلاً لهذه المشاكل هي طريقة **العنقدة** Clustering، إذ تقوم هذه الطريقة بتجزئة المناطق من خلال تنظيم

الصورة في مجموعات البكسل التي لديها معدل تشابه أعلى. وتتميز بأن الشكل الناتج عن عملية التجزئة يشابه الشكل الأصلي لأنه يستخدم قيم العضوية، كما أنها ناجحة في الكشف عن الكائنات الصغيرة جداً دون التأثير في خصائص شكلها.

واختيار مصنفات تعطي نتائج أكثر دقة في الكشف، كمصنف **SVM** الذي يعد خوارزمية تصنيف قوية جداً ذات كلفة حسابية أقل. كما يفضل استخدامه عندما يكون للبيانات صنفان فقط، كما في مثالنا.

١-٥-أهمية البحث:

إن استخدام الأنظمة الحاسوبية مع تقانات التصوير الطبية أصبح أمراً في غاية الأهمية نظراً لأنه سيعطي نتائج أكثر دقة تساعد العنصر البشري في تحليل البيانات الطبية وتحسين النتائج وتخفيض معدل الخطأ، والذي يمكن بدوره أن يحل محل طرائق الفحص البصرية الحالية من قبل الطبيب، ويسهم في تخفيض عدد الخزعات الجراحية غير الضرورية وزيادة الوثوقية بالمساعدة على التشخيص الصحيح لسرطان الرئة في مراحله المبكرة قبل أن يتقدم الورم إلى درجة كبيرة.

١-٦-أهداف البحث:

الهدف الأساسي من هذا البحث هو تطوير خوارزمية لاكتشاف العقيدات السرطانية في الرئة من خلال تحليل صور الأشعة للرئتين باستخدام تقنيات معالجة الصور الرقمية، واستخدام طرائق تجزئة تعطي نتائج أكثر كفاءة ومصداقية، لإجراء عملية التشخيص الآلي بأعلى دقة وبأقل زمن ممكن.

١-٧-تنظيم البحث:

تتكون الرسالة من عدة فصول وجدول بالمصطلحات والأشكال، إذ جرى تقسيمها وفق الآتي:

- الفصل الأول: يتضمن هذا الفصل عرضاً لعدد من الأبحاث الحديثة المنشورة في مجال البحث نفسه، ويبين أهمية البحث وأهدافه.
- الفصل الثاني: يقدم هذا الفصل تعريفاً للرئتين، بالإضافة إلى سرطان الرئة وتعريفه ومخاطره وطرائق تشخيصه، وكذلك يتناول بالتفصيل تقنية التصوير المقطعي المحوسب واستخدامها في تشخيص أورام الرئة.

- الفصل الثالث: يقدم شرحاً نظرياً موجزاً لكل من الخوارزميات والتقنيات المستخدمة في البحث.
- الفصل الرابع: يقدم هذا الفصل شرحاً لخطوات الخوارزمية المقترحة ومراحلها للكشف عن العقيدات الرئوية.
- الفصل الخامس: يلخص هذا الفصل نتائج تطبيق النظام المقترح على الصور المقطعية المحوسبة CT المستخدمة في البحث، ويقدم مجموعة من النقاط التحليلية والاستنتاجات المرتكزة على هذه النتائج التجريبية.

١-٨-خلاصة:

تحدثنا في هذا الفصل عن المشكلة العلمية المدروسة في البحث، وذكر أهداف البحث وأهميته العلمية، كما نوقشت أهم الدراسات المرجعية التي تبحث في إيجاد حل للمشكلة المدروسة في البحث، وأخيراً تم عرض ملخص يشرح تنظيم البحث وتقسيمه إلى عدة فصول.

الفصل الثاني

لمحة نظرية عن سرطان الرئة وطرائق تشخيصه

٢ – ١ –مقدمة:

سيكون لدينا في هذا الفصل نظرة عامة على تشريح الرئتين، وهو أمر مهم جداً لفهم العديد من الهياكل التي سنراها في أثناء تفسير التصوير المقطعي المحوسب للرئة، سنلقي نظرة سريعة على الجهاز التنفسي، ما سيساعد في فهم البنية الداخلية للرئة. بعد ذلك، سنتحدث عن سرطان الرئة، المرض الأساسي الذي نعمل فيه لتطوير نظام التشخيص.

٢ – ٢ –تشريح الرئتين والجهاز التنفسي:

تعد الرئة أحد أعضاء الجهاز التنفسي، إذ تعد عضواً تنفسياً مهماً في الكائنات الحية التي تستخدم الهواء الجوي للتنفس. وإن الوظيفة الرئيسية للرئة هي نقل الأوكسجين إلى المجرى الدموي وأخذ ثاني أكسيد الكربون للتخلص منه خارج الجسم، وذلك ما يعرف بعملية الشهيق والزفير. وتوجد الرئتان داخل التجويف الصدري محاطتين بالغشاء البلوري الحشوي داخل حجرة جدارها من الضلوع والقفص والعمود الفقري ودعامتهما الحجاب الحاجز. وهما عضوان إسفنجيان مرنان يشتملان على شجرة القصبات التي نتجت عن الحويصلات الرئوية. وينقسم جوف كل حويصلة إلى عدد من التحدبات هي الأسناخ الهوائية التي تزيد من سعة السطح الداخلي للهواء. وتجتمع الأسناخ لتشكل حويصلات، وتجتمع الحويصلات لتشكل كتلاً هرمية الشكل تدعى الفصوص الرئوية، وتجتمع الفصوص الرئوية وعددها ثلاثة في الرئة اليمنى، وفصان فقط في الرئة اليسرى [17].

٢ – ٣ –سرطان الرئة Lung Cancer:

هو عبارة عن ورم خبيث في الرئة ناتج عن تجمع غير طبيعي من الخلايا التي تنمو منتشرة بدرجة عالية في أنسجة tissues الرئة [18].

تسمى الأورام المكونة من الخلايا السرطانية Cancer بالأورام أو العقيدات الخبيثة (Malignant)، ويشار إلى الأورام المكونة من الخلايا غير السرطانية Non Cancer بالأورام الحميدة (Benign).

ويمكن إزالة الأورام الحميدة ومنع انتشارها إلى أجزاء أخرى من الجسم، بينما تنمو الأورام الخبيثة بقوة وتغزو أنسجة الجسم الأخرى، ما يسمح بدخول الخلايا السرطانية إلى مجرى الدم أو الجهاز اللمفاوي، ثم إلى مواقع أخرى في الجسم، وتسمى عملية الانتشار هذه بالانبثاق Metastasis، بينما تسمى مناطق نمو الورم في

هذه المواقع البعيدة بالنقائل. نظراً لأن سرطان الرئة يميل إلى الانتشار أو الانتقال في وقت مبكر جداً من الإصابة به، فهو يعد سرطان شديد الخطورة على الحياة وواحداً من أصعب السرطانات في العلاج.

وهناك عدة أشكال لسرطان الرئة، ولكن الأنواع الأكثر شيوعاً (والأكثر سرعة في الزيادة)، هي تلك الأنواع التي تشمل الخلايا الظهارية المبطنة للشعب الهوائية والقصيبات، إذ تتكون بطانة هذه الممرات الهوائية من طبقتين من الخلايا. ويؤدي التعرض المزمن للمهيجات إلى زيادة عدد الطبقات. يمكن أن يحدث هذا بشكل خاص في فروع القصيبات والممرات الهوائية، إذ تختفي الخلايا الهدبية والمخاطية لها وتستبدل بكتلة غير منظمة من الخلايا بنواة غير طبيعية، إذا استمرت هذه العملية فإن الكتلة متزايدة النمو سوف تخترق الغشاء القاعدي الأساسي للرئتين، وبالتالي تتفكك الخلايا الخبيثة وتنتقل عبر الدم والخلايا الليمفاوية إلى أجزاء أخرى من الجسم، وقد تستقر وتستمر في التكاثر. وهذا ما يعرف بالورم الخبيث للرئة، الذي يقتل المريض في النهاية [19].

٢-٤-الفروقات الشكلية بين أنواع الأورام:

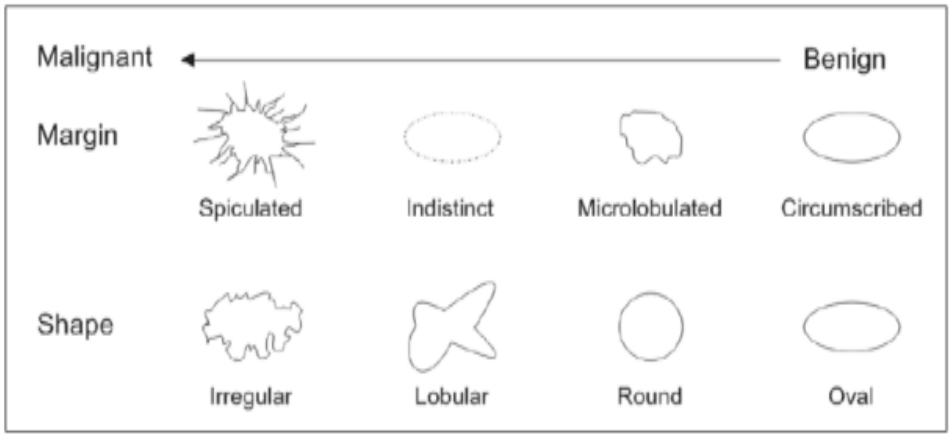
الأورام الخبيثة:

تظهر كتلة الورم الخبيث في الصور الشعاعية للرئة على شكل كثافة غير متجانسة المحتوى، وغالباً ما تكون بمحيط مشرشر على شكل أسنان المشط Spiculated حوالي 60% من حالات الأورام الخبيثة، أو تكون من دون مظهر وصفي محدد، أو بشكل غير منتظم [20].

الأورام السليمة:

تظهر بشكل كثافة ورمية صغيرة بشكل دائري أو بيضوي بحدود واضحة ومحتوى متجانس. ويعتمد أطباء الأشعة في تشخيص عقيدات الرئة على أعينهم وعلى خبراتهم السابقة. ويستخدم أطباء الأشعة بشكل رئيسي صفات مميزة مستقاة من شكل الورم [20].

يظهر الشكل التالي الاختلاف بين الأورام الحميدة والخبيثة:



الشكل (٢-١) الفروقات الشكلية بين الأورام الحميدة والخبيثة [20].

٢-٥-تشخيص سرطان الرئة:

يستخدم الأطباء مجموعة واسعة من إجراءات التشخيص والاختبارات لتشخيص سرطان الرئة، ويتضمن ذلك:

٢-٥-١-الفحص الطبي والتاريخ المرضي:

يمكن أن يكشف التاريخ والفحص البدني عن وجود أعراض أو علامات مشبوهة لسرطان الرئة، بالإضافة إلى السؤال عن الأعراض وعوامل الخطر لتطور السرطان مثل التدخين، كما يمكن أن يكتشف الأطباء علامات صعوبات في التنفس أو انسداد مجرى الهواء أو التهابات في الرئتين. وبالتالي سوف تشير التغييرات في أنسجة المرضى إلى الإصابة بسرطان الرئة [21].

٢-٥-٢-تنظير القصبات الهوائية Bronchoscopy:

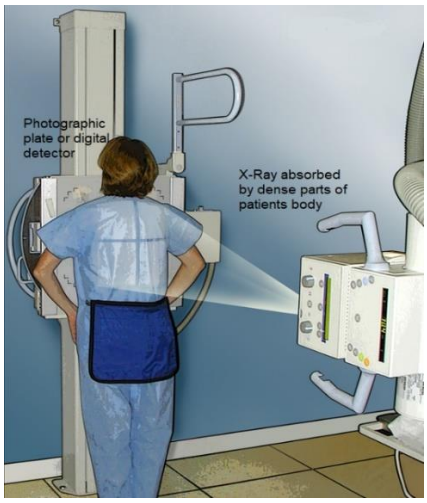
يمكن إجراء تنظير القصبات باستخدام منظار القصبات الهوائية الذي يتم إدخاله عن طريق الأنف أو الفم للكشف عن مناطق الورم التي تم أخذ الخزعات منها، ويمكن أن يكون الإجراء غير مريح، ويتطلب التخدير، في حين أنه يعد آمناً نسبياً، ويجب أن يتم إجراؤه بواسطة أخصائي الرئة [21].

٢-٥-٣-الخزعة biopsy:

يتم اللجوء إلى إجراء الخزعات عندما لا يمكن الوصول إلى محيط ورم الرئة باستخدام تنظير القصبات، ويتم إعطاء كمية صغيرة من التخدير الموضعي قبل إدخال إبرة رفيعة عبر جدار الصدر في المنطقة الورمية في الرئة، تسحب الخلايا إلى المحقنة، وبعدها يجري فحصها تحت المجهر من أجل التأكد من خباثة الورم.

٢-٥-٤-التشخيص باستخدام الأشعة السينية X-Ray:

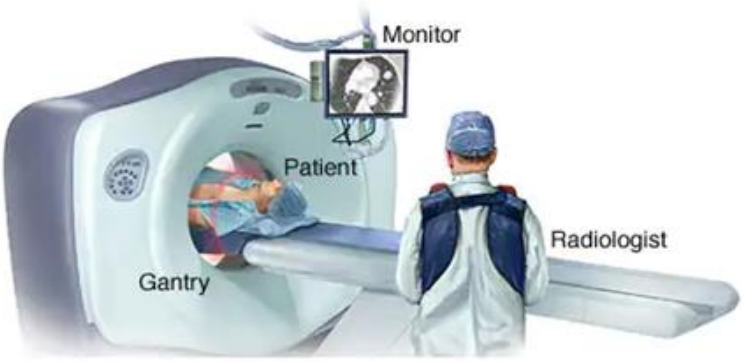
تعد الأشعة السينية للصدر هي الخطوة التشخيصية الأولى الأكثر شيوعاً عند ظهور أية أعراض جديدة لسرطان الرئة، وغالباً ما يتضمن إجراء تصوير الصدر بالأشعة السينية منظراً من الخلف باتجاه مقدمة الصدر، بالإضافة إلى منظر من الجانبين، ويمكن أن تكشف الأشعة السينية للصدر عن مناطق مشبوهة أو عقيدات متكلسة في الرئتين، ولكنها غير قادرة على تحديد ما إذا كانت هذه المناطق سرطانية أم لا [22].



الشكل (٢-٢) تصوير الرئة بالأشعة السينية [22].

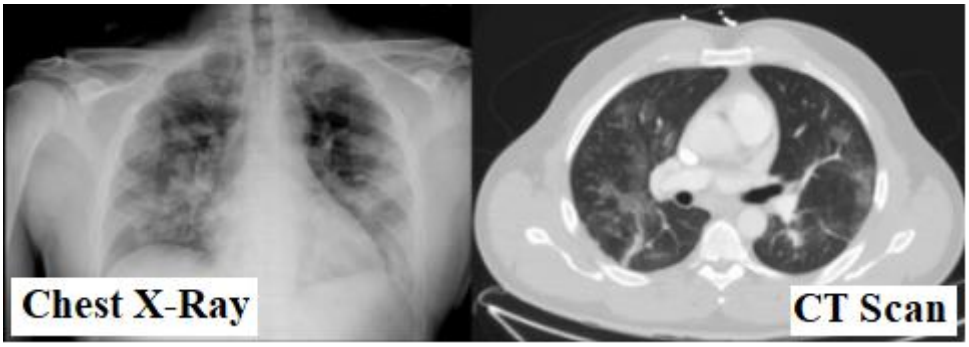
٢-٥-٥-التشخيص بالتصوير المقطعي المحوسب Computerized Tomography (CT):

يمكن إجراء فحوصات التصوير المقطعي المحوسب على الصدر البطن لفحص الأورام. وقد يطلب إجراء فحص بالأشعة المقطعية للصدر عندما لا تظهر الأشعة السينية x-ray شذوذاً أو لا تعطي معلومات كافية حول حجم الورم وموقعه [23].



الشكل (٢-٣) التصوير المقطعي المحوسب للرئتين [23].

وتتمثل إحدى ميزات التصوير المقطعي المحوسب في أنها أكثر حساسية من الأشعة السينية للصدر في اكتشاف عقيدات الرئة، يبين الشكل الفرق بين شكل الصور الشعاعية للرئتين، وتموضع البنية الداخلية، واتجاه الأنسجة التشريحية الخاصة بكل تقنية.



الشكل (٢-٤) الفرق بين التصوير بالأشعة السينية والتصوير المقطعي المحوسب للرئتين [24].

٢-٥-٦-التصوير بالرنين المغناطيسي Magnetic resonance imaging (MRI):

يعد التصوير بالرنين المغناطيسي مناسباً عندما تكون التفاصيل الدقيقة حول موقع الورم مطلوبة، وتستخدم تقنية التصوير بالرنين المغناطيسي موجات راديوية ومغناطيسية وجهاز كمبيوتر لإنتاج صور لهياكل الجسم. كما هو الحال مع التصوير المقطعي المحوسب، إذ يتم وضع المريض على سرير متحرك يتم إدخاله في ماسح التصوير بالرنين المغناطيسي، ولا توجد آثار جانبية لفحص التصوير بالرنين المغناطيسي، وتكون دقة الصورة التي ينتجها التصوير بالرنين المغناطيسي عالية للغاية، إذ يمكنها الكشف عن التغيرات الطفيفة في الهياكل داخل الجسم، والسلبية الوحيدة لهذا النوع من التصوير أنه لا يمكن استخدامه مع الأشخاص الذين يملكون صمامات قلبية اصطناعية أو أجهزة تنظيم ضربات القلب، أو أية أدوات معدنية مزروعة جراحياً؛ نظراً لخطر أن يحرك المغناطيس الأجزاء المعدنية لهذه الأجهزة [25].

٢-٥-٧-التصوير المقطعي بالإصدار البوزيتروني Positron emission tomography (PET):

هي تقنية تصوير متخصصة تستخدم عناصر مشعة قصيرة العمر لإنتاج صور ملونة ثلاثية الأبعاد لتلك المواد في الأنسجة داخل الجسم، وتقيس هذه التقنية نشاط الأنسجة وعملها، بينما تبحث فحوصات التصوير المقطعي المحوسب والتصوير بالرنين المغناطيسي في الهياكل التشريحية، كما يمكن أن تحدد فيما إذا كان نسيج الورم ينمو بشكل نشط أم لا. وبالتالي سوف يساعد في تحديد نوع الخلايا داخل ورم معين. ويتلقى المريض في هذه التقنية دواء مشعاً قصير العمر، ويتلقى تقريباً مقدار التعرض للإشعاع كصورتين شعاعية للصدر x-ray، ويقوم الدواء بإصدار البوزيترونات والتي تتفاعل لاحقاً مع الالكترونات بما يعرف بتفاعل الفناء الذي ينتج عنه زوج من فوتونات غاما (الزاوية بينهما ١٨٠)، وينطلقان باتجاه كاشف جهاز التصوير المقطعي بالإصدار البوزيتروني، إذ يسجل الكاشف هذه الأشعة، ويرسم خرائط للمنطقة، التي يوجد بها العقار المشع [26].

٢-٦-خلاصة:

تحدثنا في هذا الفصل عن سرطان الرئة والفروقات الشكلية بين أورام الرئة الحميدة والخبيثة، كما تم عرض طرق تشخيص سرطان الرئة المختلفة، وسنتحدث في الفصل القادم عن التقنيات العلمية والطرق المستخدمة في تشخيص الأورام بمساعدة الحاسوب.

الفصل الثالث

المبادئ النظرية للبحث

Theoretical Principles of Research

٣-١-مقدمة:

سنقدم في هذا الفصل شرحاً نظرياً للمفاهيم والخوارزميات والتقنيات المستخدمة في نظام الكشف المقترح عن سرطان للرئة باستخدام الحاسوب، وسنشرح مبدأ عمل كل منها بشكل مفصل وأهم ميزاتها وتطبيقاتها.

٣-٢-التشخيص بمساعدة الحاسب Computer – Aided Diagnosis (CAD):

يعرف التشخيص بمساعدة الحاسوب بأنه التشخيص الذي يقوم به الطبيب بالاعتماد على نتائج التحليل الآلي المحوسب للصور الطبية كرأي ثان في كشف الآفات والوصول إلى التشخيص الصحيح، وبطبيعة الحال فإن الطبيب هو من يقوم بإعطاء التشخيص النهائي.

ومع التطور السريع في مجال الطب وعلوم الحاسب، أصبح ما يسمى بالطبيب الالكتروني واقعاً ملموساً يمكن الوثوق به وعاملاً مساعداً في الكشف عن كثير من الأمراض وتشخيصها بشكل مبكر، ما يساعد على الشفاء المبكر أو حتى تجنب الكثير من هذه الأمراض. ويتنامى دور هذه النظم مع التقدم الكبير في طرق التصوير الطبي التي تنتج كميات هائلة من المعطيات التي تحتاج إلى من يفسرها ويستخلص منها معلومات تشخيص الأمراض بشكل سريع وفعال، وقد ينتج فحص تصوير طبقي CT لمنطقة الصدر أكثر من ٣٥٠ شريحة ثنائية البعد لكل مريض، وهذه البيانات تحتاج الكثير من الوقت والتركيز من طبيب الأشعة حتى يستطيع تفسيرها [27].

٣-٣-نظام تشخيص أو تحديد الأمراض بمساعدة الحاسب CAD:

هو عبارة عن نظام يساعد أطباء الأشعة في تفسير الصور الطبية بشكل دقيق وتحديد الأعراض المحتملة وتجنب التفسير الخاطئ أو إهمال بعض الآفات المرضية نتيجة التفسير الشخصي. وتعد الإشارة إلى أن نظم التشخيص تقدم حلاً أو رأياً إضافياً، ولا يمكن أن تحل بديلا عن طبيب الأشعة، الذي له الرأي الفصل في تحديد المرض.

وتستطيع نظم التشخيص تحليل الصور الطبية الناتجة عن عدة أنواع من أجهزة التصوير الطبي مثل التصوير المقطعي CT والرنين المغناطيسي MRI، والأمواج فوق الصوتية وغيرها من أجهزة التصوير الطبي. وهذه الصور قد تكون في صيغة ثنائية البعد D2 أو ثلاثية البعد D3.

وأحد أهم العوامل في نظم التشخيص هو الزمن اللازم لتشخيص أو تحديد المرض، حيث إن كلا من الطبيب والمريض لا يريدان أن ينتظرا ليقوم النظام بتشخيص الحالة. وبالتالي يحتاج نظام التشخيص إلى حواسيب ذات مواصفات عالية جداً وبإمكانات حسابية متفوقة لتخفيض الزمن اللازم للتشخيص إضافة للاستفادة من

التقنيات الحديثة في الكيان الصلب من المعالجة التفرعية التي تسهل تنفيذ المهام المعقدة وتختصر كثيراً من الزمن، وينبغي أن يترافق التطور في تقنيات العتاد الصلب بتطور مماثل في العتاد المرن من حيث البرامج والتطبيقات التي تقوم بعملية التشخيص.

وتستخدم نظم التشخيص بمساعدة الحاسب في الكشف المبكر عن كثير من الأمراض مثل سرطان الرئة، إذ خفض استخدام نظم التشخيص الكثير من حالات الوفيات الناتجة عن سرطان الرئة. كما تستخدم في الكشف عن العديد من أمراض القلب والأوعية الدموية، وفي الكشف عن الأورام التي قد تصيب مختلف أعضاء الجسم البشري، وغيرها من الأمراض [27].

٣-٤-معالجة الصور الرقمية Digital Image processing:

هي أحد فروع علم الحاسوب، تهتم بإجراء عمليات على الصور بهدف تحسينها طبقاً لمعايير محددة أو استخلص بعض المعلومات منها، يتألف نظام معالجة الصور التقليدي من خمسة مراحل متتالية وهي على الترتيب: استحصال الصورة، المعالجة الأولية، تجزئة الصورة، استخلاص السمات، تصنيف السمات.
وتشكل نظم معالجة الصورة مجال البحث الأساسي في الهندسة وتخصصات علوم الحاسب الكمبيوتر، ويمكن استخدام هذه التقنيات في المجالين الحيزي والترددي [28، 29].

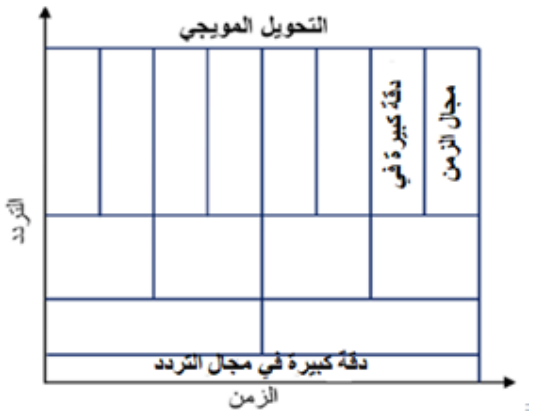
٣-٥-تحويل المويجات Wavelet Transform:

هي طريقة تحليلية رياضية تستخدم من أجل معالجة الإشارات والصور الرقمية للعديد من التطبيقات العملية، حيث يتم تحويل الإشارة (الصورة) من المجال الحيزي إلى مركبات ترددية مختلفة ضمن نافذة متغيرة Different Scale. ويمكن لنطاق التردد المنخفض أن يميز الخصائص الإجمالية للصورة (الشكل العام) في النافذة كبيرة الحجم. ويمكن لنطاق التردد العالي أن يميز التغيير المفاجئ للإشارة (بالنسبة للصورة فهي التفاصيل والحواف)، إذ يتم تغيير عرض النافذة للحصول على معلومات مختلفة التردد على طول الموجة، ليتم الحصول على ما يعرف بالمويجات التي يختلف ترددها حسب عرض النافذة المستخدم [30].

يعطى تابع تحويل المويجة بالشكل الآتي:

$$W(a,b) = \int_{-\infty}^{+\infty} f(x)\,\psi_{a,b}(x)dx \qquad (٣-١)$$

تمثل a, b معاملات التوسيع Scaling والانزياح Translation، بينما يمثل $\psi_{a,b}$ تابع التحليل. ويتضمن تابع المويجة فكرة الانتقال والتوسيع معا، ويستخدم لتوليد تفاصيل الصورة في مستويات دقة مختلفة، بينما يستخدم تابع التوسيع لتوليد تقريب approximation للصورة في مستويات دقة مختلفة.



الشكل (٣-١) التحويل المويجي [30].

وإحدى أهم خصائص التحويل الترددي أن عملية الجداء الالتفافي في المجال الحيزي مكافئة تماماً لعملية جداء عناصر الصورة الترددية في عناصر القناع في المجال الترددي، وبالتالي يعد الترشيح في المجال الترددي أسرع وأكثر كفاءة من المجال الحيزي، لأن القناع سيكون بحجم أصغر بكثير من الصورة، وسيتم تطبيقه على كل عنصر من عناصر الصورة على حدة. ومن ثم أخذ مجاورات القناع بالحسبان لتحديد قيمة عنصر واحد فقط، وبالتالي تعتبر هذه العملية بطيئة [31].

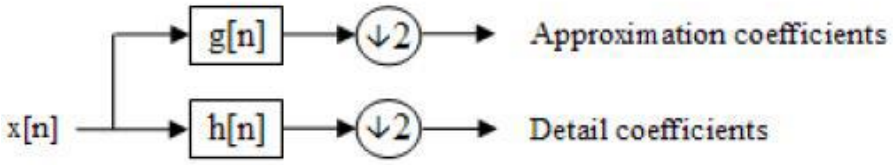
٣-٥-١-تحليل الصورة وإعادة تركيبها في التحويل المويجي:

يتم تطبيق هذا التحويل باستخدام المرشحات، حيث يتم تقسيم الإشارة x(n) إلى مرشح تمرير منخفض g[n] للحصول على المعلومات التقريبية للصورة Approximation Cofficients والتي تمثل مكونات الإشارة ذات التردد المنخفض (الشكل العام للصورة)، وعلى مرشح تمرير عالي h[n] للحصول على المعلومات التفصيلية للصورة Details Cofficients والتي تمثل مكونات الإشارة ذات التردد العالي (الحواف الأفقية والعمودية والقطرية)، ويسمح هذا التقسيم بتحليل الصورة ضمن حزم ترددية مختلفة مع دقة مختلفة.

تقليل ومضاعفة البيانات:

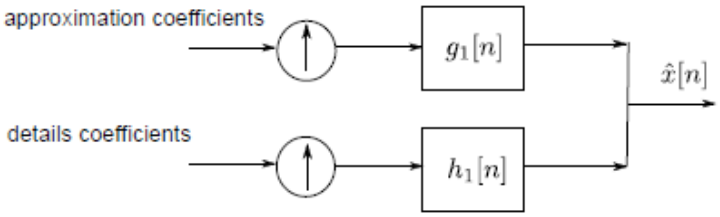
يتم فصل تردد الصورة المدخلة إلى حزم ترددية مساوية لعرض الحزمة بواسطة مرشحي التمرير المنخفض والعالي، إذ يحتوي خرج المرشحين على نصف التردد ولكن مساوياً لمجموع العينات لصورة الدخل، ويحتوي

الخرجان معاً على محتوى التردد نفسه للصورة المدخلة، وبذلك تتم مضاعفة كمية العينات، فمن أجل حل هذه المشكلة يتم استخدام ما يدعى بعملية تقليل العينات (أي مرشح نصف حزمة) والتي يرمز لها بالرمز (↓٢) أي يتم أخذ العينات بنسبة عامل اثنين [32].



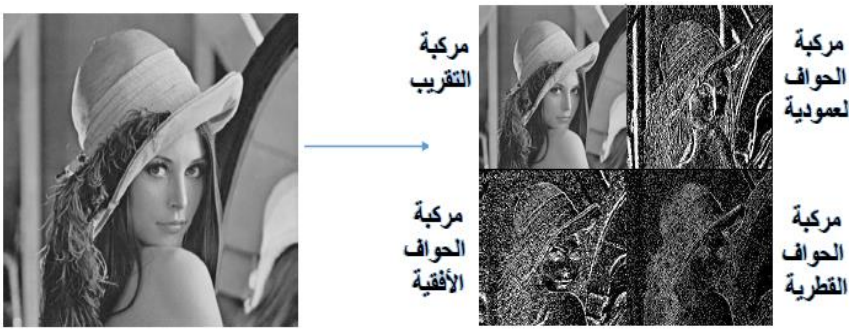
الشكل (٣-٢) مرحلة تفكيك الصورة إلى مركباتها الفرعية [32].

ويمكن العودة والحصول على الإشارة الأصلية من المعلومات التقريبية والتفصيلية والقيام بعملية زيادة البيانات إلى الضعف، وهذا ما يدعى بزيادة العينات (↑٢)، وفق الشكل الآتي:



الشكل (٣-٣) مرحلة التركيب والحصول على الصورة الأصلية [32].

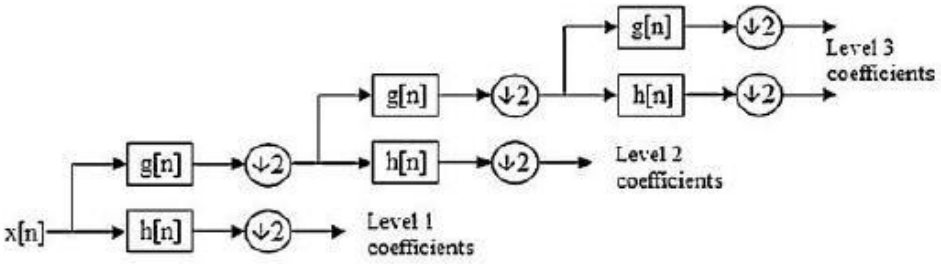
ويوضح الشكل الآتي تطبيق تقنية تحويل المويجات للحصول على المعلومات التفصيلية والتقريبية على صورة ما:



الشكل (٣-٤) نتيجة تطبيق تقنية تحويل المويجات على صورة ما [32].

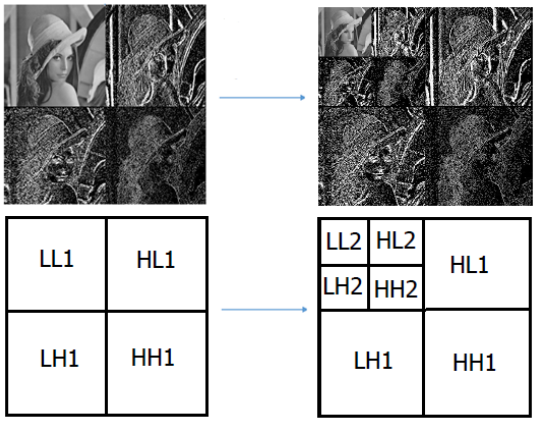
٣-٥-٢-تحويل الصورة متعدد المستويات (شجرة التحليل المويجي):

ينتج مرشح تمرير نصف الحزمة (المرتفع والمنخفض) في كل مرحلة تحليل إشارات ذات مجال ترددي مساوٍ لنصف المجال الترددي للصورة الأصلية، وهذا يضاعف الدقة الترددية للصورة، مما يعني إمكانية الحصول على المزيد من تفاصيل الصورة في كل عملية تحليل. ويتم الحصول على الصورة الأصلية، أي تركيب الصورة بالتجميع المتسلسل لكل المعاملات الناتجة سابقاً، (المعاملات التفصيلية والتقريبية) بدءاً من آخر عملية تحليل، وتستمر هذه العملية بنفس عدد مراحل التحليل حتى يتم الحصول على الصورة الأصلية [32].



الشكل (٣-٥) التحويل المويجي متعدد المستويات (ثلاث مراحل) [32].

ويوضح الشكل الآتي تطبيق تقنية تحويل المويجات متعدد المستويات للحصول على المعلومات التفصيلية والتقريبية على صورة ما:



الشكل (٣-٦) تقنية التحويل المويجي متعدد المستويات [32].

نحصل في المستوي الأول على مركبة التقريب LL1 ومركبة الحواف العمودية HL1 والأفقية LH1 والقطرية HH1. وعند المستوي الثاني تجري عملية التحلل على مركبة التقريب الناتجة من المستوي السابق، فنحصل على مركبة تقريب جديدة LL2 ومركبات الحواف العمودية HL2 والأفقية LH2 والقطرية HH2، وهكذا.

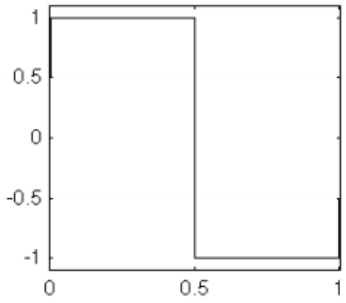
٣-٥-٣-أشهر أنواع دوال المويجة:

١- مويجة Haar Wavelet:

هي المثال الأبسط عن المويجات المتعامدة، وهي مويجة متقطعة تشبه التابع الخطوي. تدعى أيضاً مويجة (db1)، وتنتمي إلى عائلة مويجات Daubechies، وتملك تابع تدريج، ويعبر عنها بالتابع المويجي التالي:

$$\psi(X) = \begin{cases} 1 \ at \ 0 < X < 0.5 \\ -1 \ at \ 0.5 < X < 1 \\ 0 \ at \ 0 \end{cases} \qquad (٢-٣)$$

ويتكون التابع المويجي من نبضتين واحديتين متوضعتين بجانب بعضهما على المحور الزمني، إحداهما بعكس الأخرى كما موضح في الشكل (٣-٧) [33].

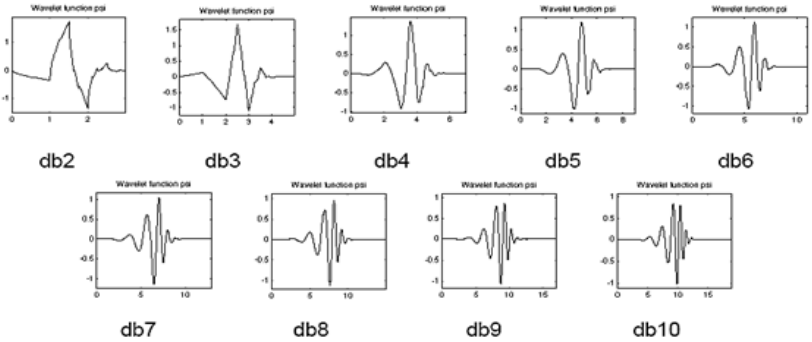


الشكل (٣-٧) التابع المويجي لمويجة Haar [33].

٢-مويجات Daubechies:

تم تطوير هذه المويجات من قبل Ingrid Daubechies، وهي من أشهر الباحثين وألمعهم في مجال المويجة. ويرمز لهذه العائلة بالرمز (dbn)، إذ تشير n إلى رتبة المويجة، وتظهر التوابع المويجية لها على الشكل الآتي (٣-٨):

تعد مويجات Daubechies مويجات غير متناظرة (وبشكل محدد عند قيم n الصغيرة)، وتزداد الانتظامية مع زيادة رتبة المويجة. كما أن التحليل الخاص بها متعامد.



الشكل (٣-٨) التوابع المويجية لمويجات Daubechies [34].

وليس من السهولة إيضاح الرتب الأعلى لدوال Daubechies بشكل تعبير تحليلي، إذ تمثل الرتبة في دوال Daubechies عدد لحظات الاختفاء، أو عدد اللحظات الصفرية لدالة المويجة. وهذا يتعلق بعدد التذبذبات للدالة المويجية. وإن العدد الأكبر للحظات الاختفاء، هو الأفضل لتحديد تردد تحليل الدالة إذ تقل الاعتمادية بين معاملات المويجة في مختلف المقاييس، مع زيادة رتبة المويجة [34].

٣-٥-٤-أهمية تحويل المويجات في الصورة:

عندما ننظر إلى صورة سنجد أنها مكونة من مجموعة من المناطق المتصلة ذات المادة المتشابهة والتدرجات الرمادية المختلفة. وإذا كانت مكونة من عناصر صغيرة وكبيرة وعناصر ذات تباين منخفض ومرتفع معاً، فإن التعامل مع الصورة **بدقات مختلفة** يعد من المزايا، وهذا هو الحافز من أجل المعالجة متعددة الدقة. ومن وجهة نظر رياضية، فإن الصورة هي مصفوفة ثنائية البعد، عناصرها تمثل الكثافة وإحصائيات متغيرة محلية. وهي تنتج مجموعة مختلفة من الملامح مثل الحواف والمناطق المتجانسة أو المتباينة.

وبإمكان تحويل المويجة أن يكشف هذه الخواص المميزة للصورة، والتي لا يمكن كشفها بأية طريقة أخرى، وهذا يلعب دوراً هاماً في استخراج معلومات الحافة وحفظها عند إزالة الضجيج من الصورة، في حين أن الطريقة التقليدية لإزالة الضجيج القائمة على تحويل فوريه، قد تسبب فقدان معلومات حواف الصورة لأنها غير قادرة على اكتشاف التغيرات المحلية المفاجئة.

٣-٥-٥-إزالة الضجيج باستخدام المويجات:

يزيل التحويل المويجي الضجيج بطريقة أكثر فعالية من الأساليب التقليدية المستخدمة. وتبدأ إزالة الضجيج بتطبيق التحويل المويجي على الصورة الضجيجية، ثم تطبيق حد العتبة على المعاملات الناتجة، وتطبيق معكوس التحويل للحصول على صورة تم تنقيتها مشابهة للصورة الأصلية. ويستند إزالة الضجيج بتطبيق العتبة على معاملات المويجة إلى ملاحظة أن الطاقة تتركز في الأبعاد الصغيرة لمتجهات المويجة في الصورة. وهذه المتجهات تكون كبيرة مقارنة مع المتجهات لأي صورة أخرى (مثل الضجيج)، والتي تتوزع طاقتها على عدد كبير من المعاملات، وبالتعويض بأصفار محل المعاملات الصغيرة، نستطيع تقليل الضجيج وإزالته مع الحفاظ على المعاملات المهمة التي تمثل الصورة الأصلية [35].

خطوات إزالة الضجيج باستخدام تحويل المويجات:

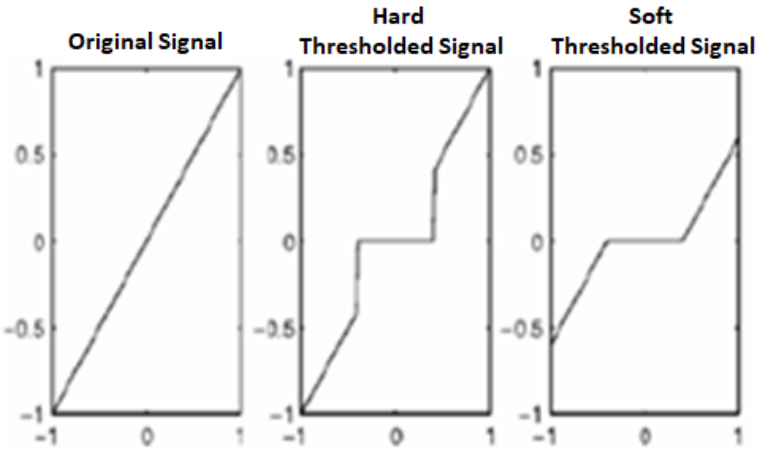
يضم الإجراء العام لإزالة الضجيج من الصورة ثلاث خطوات أساسية [36]:

١-**التحليل Decomposition**: اختيار المويجة، واختيار المستوى N، وحساب التحليل المويجي لإشارة S في المستوى N.

٢-**التعتيب Thresholding**: تطبق العتبة على معاملات التفاصيل، لكل مستوى من ١ إلى N، ويحدد نوع حد العتبة فقد يكون مرناً أو قاسياً.

العتبة الناعمة والقاسية Hard and Soft Thresholding:

تعد العتبة القاسية من أبسط الطرائق في التطبيق، أما العتبة المرنة، فتمتلك خصائص رياضية. كما هو مبين في الشكل (٣-٩):



الشكل (٣-٩) العتبة القاسية والمرنة [37].

لنفرض أن t تمثل العتبة، فتمثيل العتبة القاسية على الإشارة X كما موضح المعادلة:

$$X = \begin{cases} Xif|X| > t \\ 0|X| < t \end{cases} \quad (٣-٣)$$

أما العتبة المرنة فموضحة في المعادلة:

$$X = \begin{cases} sign(X)(|X| - t)if|X| > t \\ 0|X| < t \end{cases} \quad (٤-٣)$$

وتوصف المعالجة بالعتبة القاسية بوضع أصفار محل القيم التي فيها القيمة المطلقة أقل من العتبة t. أما العتبة المرنة، فيتم وضع أصفار محل القيم التي فيها القيمة المطلقة أقل من العتبة t، ثم تقريب المعاملات القريبة من الصفر إلى الصفر [37].

٣-**إعادة البناء Reconstruction**: إعادة بناء المويجة، يتم باستخدام معلومات التفاصيل الأصلية في المستوى N، وتعديل معلومات التفاصيل من ١ إلى N.

٣-٦-المعالجة باستخدام العمليات المورفولوجية Morphological operations:

يمكن وصف العمليات المورفولوجية بأنها مجموعة من إجراءات معالجة الصور التي تقوم باستخلاص المعلومات الضرورية من الصور، وكذلك التخلص من الضجيج، وحذف المعلومات غير الضرورية مع مراعاة الأشكال والتداخلات. وتقوم بتطبيق عناصر التشكيل على صورة الدخل مع مراعاة الهدف النهائي لإنتاج صورة بالحجم نفسه، ويتم ذلك عن طريق اختيار حجم المنطقة وحالتها مع مجاوراتها [38].

وعنصر التشكيل هو عبارة عن مصفوفة تتكون من مجموعة من الأصفار والواحدات. ويتم تعيين قيمة كل عنصر في صورة الخرج وفقاً لمقارنة العنصر المقابل في صورة الدخل مع مجاوراتها. ومن العمليات المورفولوجية الأساسية والأكثر شيوعاً هي التمدد dilation، والتعرية erosion.

١-التمدد dilation:

هو إضافة عناصر Pixels إلى حدود الكائنات في الصورة بناء على شكل نصف قطر عنصر البناء، بينما تتمثل التعرية في إزالة العناصر الموجودة على حدود الكائن. ويعتمد عدد العناصر التي تمت إضافتها أو إزالتها على حجم وشكل عنصر التشكيل المستخدم لمعالجة تلك الصورة.

٢-التعرية erosion:

هي عملية تقوم بتمديد مكمل الصورة، وللقيام بعملية تمدد الصورة نقوم بتضييق مكملها. وبمعنى آخر، يمكن استخدام التمدد لتوسيع مقدمة الصورة وتقليص الخلفية، بينما تستخدم التعرية لتقليص الصورة الأمامية وتوسيع الخلفية.

٣-٧-تجزئة الصورة Image Segmentation:

هي تقسيم الصورة إلى مناطق أو فئات، تتوافق مع الكائنات أو الأجزاء المختلفة منها، إذ يتم تخصيص كل عنصر في الصورة إلى واحدة من عدد من هذه الفئات. وعادة ما يكون التقسيم الجيد هو التقسيم الذي يحوي على ما يأتي:

١.عناصر الصورة pixels في الفئة نفسها، والتي لها التدرج الرمادي grayscale نفسه لمتغيرات متعددة القيم، والتي تشكل منطقة متصلة.

٢.عناصر الصورة المتجاورة والموجودة في فئات مختلفة، لها قيم مختلفة.

وغالباً ما يكون التقسيم هو الخطوة الحاسمة في تحليل الصورة، لأنه إذا تم إجراء التجزئة بشكل جيد، فسيتم تسهيل جميع المراحل الأخرى في تحليل الصور. لكن غالباً ما يكون النجاح جزئياً فقط عند استخدام خوارزميات التجزئة الآلية. ومع ذلك يفيد التدخل اليدوي عادة للتغلب على هذه المشاكل. وفي هذه المرحلة يجب أن يكون الحاسب قد قام بالفعل بمعظم العمل. ويمكن تطبيق خوارزميات التجزئة على الصور الأصلية، أو بعد تطبيق التحويلات والمرشحات عليها. كما يمكن استخدام طرائق التشكيل الرياضية لتحسين النتائج بعد عملية التجزئة، وبالتالي سيتم استخدام نتائج التجزئة من أجل استخراج المعلومات ذات الأهمية من الصور [39].

هناك ثلاث طرق عامة للتجزئة:

هي التجزئة باستخدام العتبة Thresholding، والتجزئة القائمة على الحافة edge، والتجزئة القائمة على المنطقة region.

٣-٧-١-التجزئة القائمة على الحواف Edge based Thresholding:

يتم تطبيق مرشح الحافة على الصورة، وتصنف العناصر على أنها حافة أو غير حافة اعتماداً على خرج المرشح، ويتم تخصيص العناصر غير المفصولة بحافة للفئة نفسها.

٣-٧-٢- التجزئة باستخدام العتبة Thresholding Segmentation:

العتبة Thresholding هي تقنية مهمة في تطبيقات تجزئة الصور، والفكرة الأساسية من العتبة هي تحديد قيمة عتبة المستوى الرمادي المثلى لفصل الكائنات المرغوبة من خلفية الصورة، بناء على توزيعها في المستوى الرمادي. وتجزئة الصورة عن طريق العتبة Thresholding Segmentation هي طريقة بسيطة ولكنها فعالة جداً لتقسيم الصور التي تحتوي على كائنات فاتحة مع خلفية داكنة، كما هو الحال في مثالنا. وتعتمد تقنية العتبة على مناطق مساحة الصورة، أي على خصائص الصورة، إذ تقوم بتحويل صورة التدرجات الرمادية إلى صورة ثنائية، أي أنها تختار الحد المناسب T لتقسيم عناصر الصورة إلى عدة مناطق، وفصل الكائنات عن الخلفية. ويتم ذلك بتخصيص القيمة 1 إلى جميع عناصر الصورة التي تزيد قيمتها عن العتبة، وتخصيص القيمة 0 إلى العناصر التي تقل قيمتها عن العتبة [40].

٣-٧-٢-١-أهم طرق التعتيب الأكثر شيوعاً:

١-العتبة البسيطة Simple Thresholding:

هي إحدى الطرق المستخدمة على نطاق واسع لتجزئة الصور، وتعد مفيدة جداً في التمييز بين المقدمة والخلفية عن طريق اختيار قيمة عتبة مناسبة T، لتحويل صورة المستوى الرمادي إلى صورة ثنائية، إذ تحتوي الصورة الثنائية على المعلومات المطلوبة حول موضع الكائنات المهمة وشكلها (المقدمة). وتتمثل ميزة الحصول على صورة ثنائية في أنه يقلل من تعقيد البيانات، ويبسط عملية التعرف والتصنيف.

والطريقة الأكثر شيوعاً لتحويل الصورة ذات المستوى الرمادي إلى صورة ثنائية هي تحديد قيمة عتبة واحدة T. وسيتم بعدها تصنيف جميع قيم المستوى الرمادي الأدنى من هذه القيمة على أنها سوداء (0)، وستكون القيم الأعلى منها هي بيضاء (1). وتصبح مشكلة التجزئة هي اختيار القيمة المناسبة للعتبة T.

$$g(x,y) = \begin{cases} 1 & if \quad f(x,y) \le T \\ 0 & otherwise \end{cases} \qquad (٥-٣)$$

٢-عتبة أوتسو Otso Thresholding:

هي طريقة تجزئة عالمية تعتمد على المنطقة الرمادية gray region لتحديد العتبة بشكل آلي، وتتطلب طريقة Otsu حساب الرسم البياني histogram للمستوى الرمادي قبل التنفيذ. وتقلل هذه العتبة من تعقيد البيانات، وتجعل عملية التعرف، والتصنيف أكثر بساطة [41].

خطوات Otsu Thresholding:

١-حساب العدد الكلي لعناصر Pixels الصورة P_{all}.

٢-تحديد قيمة عتبة معينة ولتكن t، التي سوف تقسم عناصر الصورة إلى مجموعتين (في مثالنا المقدمة والخلفية).

٣-حساب عدد عناصر المقدمة Foreground بعد تطبيق العتبة t، وهي $P_{FG}(\mathrm{t})$.

حساب عدد عناصر الخلفية Background بعد تطبيق العتبة t، وهي $P_{BG}(\mathrm{t})$.

٤-حساب الأوزان في المقدمة $W_{fg}(\mathrm{t})$، والأوزان في الخلفية $W_{bg}(\mathrm{t})$ وفق القانون:

$$W_{fg}(\mathrm{t}) = \frac{P_{FG}(\mathrm{t})}{P_{all}} \ , \ W_{bg}(\mathrm{t}) = \frac{P_{BG}(\mathrm{t})}{P_{all}} \qquad (٦-٣)$$

٥-حساب تباين قيم عناصر المقدمة ${\sigma_{fg}}^2$، وتباين قيم عناصر الخلفية ${\sigma_{bg}}^2$ وفق القانون:

$${\sigma_{fg}}^2=\frac{\sum(x_i-\overline{x_{fg}})}{N-1}\ ,\quad {\sigma_{bg}}^2=\frac{\sum(x_i-\overline{x_{bg}})}{N-1} \qquad (٣-٧)$$

٦-حساب التباين الكلي الموزون وفق القانون:

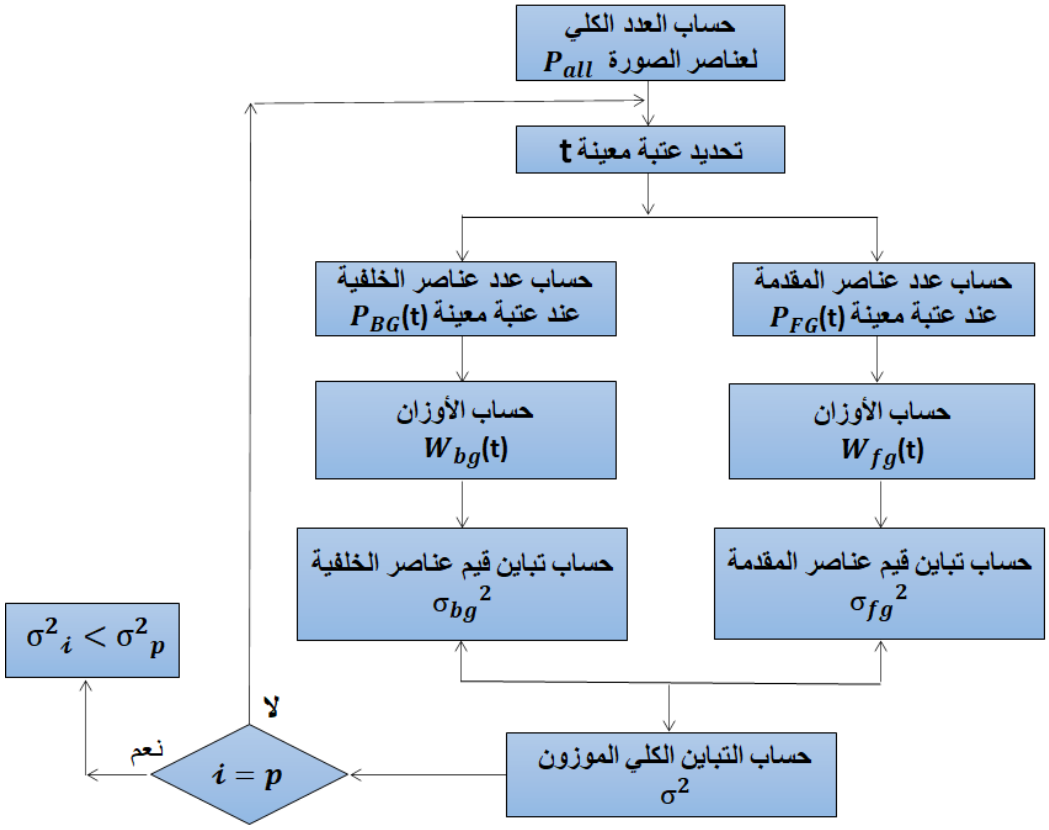
$$\sigma^2 = W_{fg}(\mathrm{t})\ * {\sigma_{fg}}^2 + W_{bg}(\mathrm{t}) * {\sigma_{bg}}^2 \qquad (٣-٨)$$

٧- إعادة تنفيذ الخطوات من ٢ إلى ٦، حتى يتم الوصول إلى أدنى قيمة للتباين الكلي الموزون، (أي $\boldsymbol{{\sigma^2}_i < {\sigma^2}_p}$)، والتي سيتم اعتماد قيمة العتبة المقابلة عندها.

إذ تمثل $\boldsymbol{{\sigma^2}_i}$: قيمة التباين الكلي الموزون عند تكرار معين $\boldsymbol{i}$.

و $\boldsymbol{{\sigma^2}_p}$: قيمة التباين الكلي الموزون عند آخر تكرار تم المرور به.

و $\boldsymbol{p}$: عدد قيم العتبات المحتملة، ويتم المرور عبر جميع قيم العتبة المحتملة على عناصر الصورة، من أجل الحصول على صورة ثنائية، وهي ستحتوي على المعلومات المطلوبة حول موضع الكائنات المهمة وشكلها.



الشكل (٣-١٠) خطوات خوارزمية عتبة Otsu [42].

٣-٧-٣-التجزئة القائمة على المنطقة Region based Thresholding:

تعمل هذه الخوارزميات بشكل تكراري من خلال تجميع عناصر الصورة المتجاورة معاً، التي تملك قيماً متشابهة، وتقسم عناصر الصورة التي تختلف في القيمة. ومن أهم الخوارزميات المستخدمة فيها:

٣-٧-٣-١-عنقدة البيانات Data Clustering:

هي عملية وضع البيانات في مجموعات متماثلة، إذ تكون البيانات في المجموعة الواحدة متشابهة بالنسبة لعدة خصائص محددة، ومختلفة عن البيانات في المجموعات الأخرى بالنسبة لهذه الخصائص (بالنسبة للصورة اللون، السوية الرمادية). وعند تطبيق هذا التعريف على الصورة، فإن طرائق العنقدة المختلفة، يمكن أن تستخدم في عمليات تجزيء الصورة لتقسيم عناصرها ضمن مجموعات بحيث تكون عناصرها ضمن المجموعة الواحدة متشابهة فيما بينها بالنسبة إلى عدد من الخصائص مثل اللون أو السوية الرمادية. وتقسم العنقدة مجموعة من البيانات إلى عدة مجموعات، إذ أن التشابه بين النقاط ضمن مجموعة معينة أكبر من التشابه بين نقطتين ضمن مجموعتين مختلفتين. تستخدم العنقدة على نطاق واسع، وليس فقط لتنظيم البيانات وتصنيفها، وإنما هي مفيدة لضغط البيانات وبناء نموذج ترتيبها. وإذا كان بإمكاننا أن نجد مجموعات من البيانات، فإنه بالإمكان بناء نموذج للمشكلة على أساس تلك المجموعات [43].

٣-٧-٣-٢-خوارزميات عنقدة البيانات:

١-العنقدة باستخدام خوارزمية K-mean Clustering:

هي خوارزمية لجمع عدد من البيانات استناداً إلى خصائص هذه البيانات وسماتها، وتتم عمليه التجميع من خلال تقليل المسافات بين البيانات ومركز التجمع [44].

وتتم هذه الخوارزمية من خلال الخطوات الآتية:

١.حساب إحداثيات مراكز التجميع.

٢.حساب المسافة بين كل البيانات ومراكز التجميع.

٣.تجميع البيانات وتنظيمها في مجموعات بناء على أقل المسافات بين المركز ونقاط البيانات.

٤.إعادة تنفيذ الخطوات من ١ إلى ٣ حتى الوصول إلى حالة الثبات.

يعتمد أداء هذه الخوارزمية على المواقع الأولية لمراكز التجمع، ومن المستحسن تنفيذ هذه الخوارزمية عدة مرات مع اختلاف المراكز في كل مرة عن المرات السابقة.

٢-العنقدة الضبابية باستخدام خوارزمية Fuzzy C-Mean:

هي عبارة عن خوارزمية تكرارية تقوم بتقسيم البيانات ضمن مجموعات، وذلك بالاعتماد على مفهوم العضوية الضبابية، فبدلا من أن يتم إسناد كل عنصر صورة إلى مجموعة واحدة كما هو الحال في خوارزمية K-Mean، فإن كل عنصر صورة سوف يملك قيماً عضوية مختلفة بالنسبة إلى كل مجموعة من المجموعات، وبالتالي فإن العناصر التي تقع في منتصف المجموعة تملك قيماً عضوية أكبر من العناصر التي تقع على حواف المجموعة [46, 45].

وتحاول خوارزمية FCM أن تحدد المجموعات ضمن البيانات من خلال الحصول على أصغر قيمة للتابع الموضح ضمن المعادلة الآتية:

$$\mathrm{J}(U, C_1, C_2, \dots, C_C) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^{m} d_{ij}^{2} \qquad (٣-٩)$$

حيث:

J : الهدف، إذا كانت قيمته بعد أحد التكرارات (تكرار تنفيذ الخوارزمية) أقل منها في التكرار السابق فهذا يعني أن الخوارزمية قد اقتربت من تحقيق فصل جيد لعناصر الصورة ضمن المجموعات المختلفة.

$\boldsymbol{n}$: عدد نقاط البيانات.

$\boldsymbol{m}$: عامل يحدد درجة الضبابية وقيمته $1 < m < \infty$، وكلما اقتربت قيمته من الواحد تصبح عملية تجميع البيانات أقل عشوائية، بحيث تصبح نقطة البيانات منتمية إلى مجموعة واحدة فقط، وبالمقابل كلما اقتربت قيمة هذا العامل من اللانهاية، أصبحت عملية تجميع البيانات أكثر عشوائية بحيث تصبح كل نقطة بيانات منتمية إلى كل مجموعة من المجموعات.

وقد تم تنفيذ عديد من الدراسات حول اختبار قيمة هذا العامل، وتبين أن القيمة الأفضل لهذا العامل هي 2.

$\boldsymbol{U}$: جدول العضوية: وهو عبارة عن جدول مؤلف من قيم العضوية لكل عنصر من عناصر الصورة بالنسبة إلى المجموعات، أي أن u_{ij} يحدد درجة انتماء العنصر X_j إلى المجموعة ذات الدليل i.

$\boldsymbol{C}$: عدد المجموعات المستخدمة في الخوارزمية $1 < C < n$، يتم تحديده قبل تنفيذ الخوارزمية بناء على المعرفة المسبقة، حيث C هي مصفوفة مراكز المجموعات $\{C_1, C_2, \dots, C_C\}$.

$\boldsymbol{u_{ij}}$: درجة العضوية (الانتماء إلى المجموعة)، تتعلق بمقلوب المسافة بين عنصر البيانات ومركز المجموعة، وتعطى من خلال المعادلة الآتية:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{jk}} \right)} \qquad (٣-١٠)$$

حيث:

$$d_{ij} = \left|\left| X_j - C_j \right|\right| \qquad (٣-١١)$$

خطوات الخوارزمية:

١. إعطاء قيم عشوائية لعناصر مصفوفة العضوية $U^{(k=0)}$.

٢. حساب مراكز المجموعات من خلال المعادلة:

$$C_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}} \qquad (٣-١٢)$$

٣. حساب التباين بين نقاط البيانات ومراكز المجموعات باستخدام معادلة التابع، ونتوقف إذا كانت قيمة التحسن عن الدورة السابقة أصغر من عتبة التحسن ε.

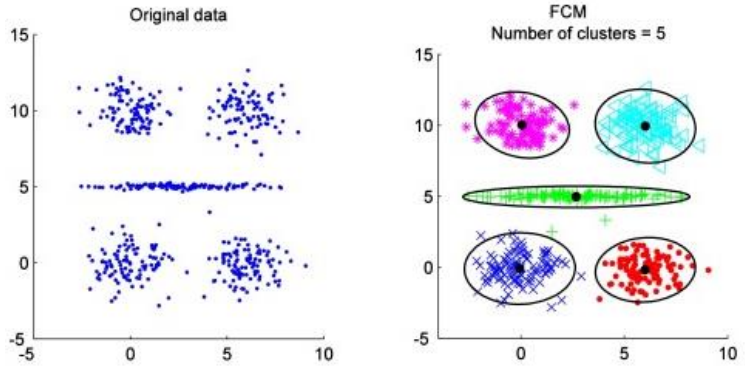
أي عندما: $\varepsilon \leq \left\| U^{(k)} - U^{(k-1)} \right\|$.

حيث ε: ثابت يمثل عتبة التحسن المقبول في النتيجة، ويتم تحديده من البداية.

٤. حساب $U^{(k)}$ الجديدة وتكرار الخطوات من ٢ إلى ٤.

وكلما اقترب فرق التحسن من قيمة الثابت، نكون قد اقتربنا من التوزيع الأمثل لمراكز المجموعات. وأصبحت النقاط البعيدة عن كل مركز أقل تأثيراً بسبب درجة العضوية التي تتناسب عكساً مع البعد عن المركز، وبالتالي فإن الفرق بين تنفيذ الخطوات يتناقص.

ويوضح الشكل مثالاً عن مجموعة من البيانات العشوائية، إذ تم استخدام خوارزمية FCM من أجل تجميعها ضمن عدة مجموعات.



الشكل (٣-١١) العناقيد الناتجة عن تطبيق خوارزمية FCM [45].

٣-٨-الذكاء الاصطناعي Artificial Intelligence:

يعرف العلماء الذكاء الاصطناعي بأنه نظام كمبيوتر قادر على أداء المهام التي تتطلب عادة محاكاة الذكاء البشري، مثل الإدراك البصري والتعرف على الكلام وصنع القرار والتعرف على النماذج والتصنيف [47].

ومن أهم فروع الذكاء الاصطناعي:

٣-٨-١-تعلم الآلة Machine Learning:

ويمكننا تعريفه بأنه العلم الذي يسمح للكمبيوتر بالتصرف، من دون أن يكون مبرمجاً مسبقاً للقيام بذلك التصرف بشكل صريح. أو بمعنى آخر يتعلم كيفية الاستجابة لأحداث معينة بالطريقة الصحيحة بشكل ذاتي من دون أن يتم تلقينه ذلك من قبل المبرمج. ويستخدم في العديد من التطبيقات كتنقيب البيانات وفي مجال الطب لأغراض التشخيص الطبي وفي مجال الاتصالات والاقتصاد وغيرها.

٣-٨-٢-أنواع تعلم الآلة:

١-التعلم بإشراف Supervised Learning:

يقصد به تدريب الخوارزمية على بيانات مصنفة Labeled، أي بيانات الدخل والخرج معاً. ومن خلال هذه العملية تبني الخوارزمية علاقة ما بين المدخلات والمخرجات، إذ تمكنها من التنبؤ بمخرجات بيانات جديدة. ومن أهم الخوارزميات المستخدمة في هذا النوع خوارزمية متجه الدعم الآلي SVM.

٢-التعلم من دون إشراف Unsupervised Learning:

يقصد به تدريب الخوارزمية على بيانات غير مصنفة Unlabeled، أي بيانات الدخل فقط. ومن خلال هذه العملية تبني الخوارزمية علاقة ما بين المدخلات، بحيث تمكنها من التنبؤ بمخرجات بيانات جديدة.

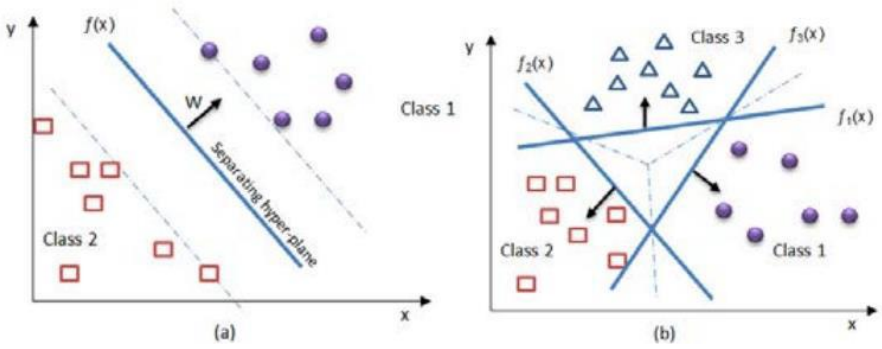
٣-التعلم المعزز Reinforcement Learning:

ويقصد به أن الخوارزمية تتعلم من تلقاء نفسها، ومن أخطائها.

٣-٨-٣-خوارزمية متجه دعم الآلة Support Vector Machine:

عبارة عن أسلوب يجمع بين النظرية الإحصائية والتعليم الموجه. وقد طورت من قبل الباحث فابنيك Vapnik عام ١٩٩٨. وتعتمد فكرة خوارزمية متجه دعم الآلة على البحث عن أفضل طريقة لتقسيم البيانات إلى مجموعتين بوضع مستوي فائق hyperplane بينهما، بغض النظر عن طبيعة البيانات سواء كانت قابلة للفصل الخطي أو لا، وهنا تكمن مصدر قوتها.

ويتمثل هدفها بالبحث عن أفضل مستوى فاصل لتصنيف البيانات، وذلك بجعل المسافة بين بيانات التدريب والمستوى الفاصل أكبر ما يمكن [48، 49].



الشكل (٣-١٢) آلية عمل SVM (a) للفصل بين صنفين، (b) للفصل بين أكثر من صنفين [48].

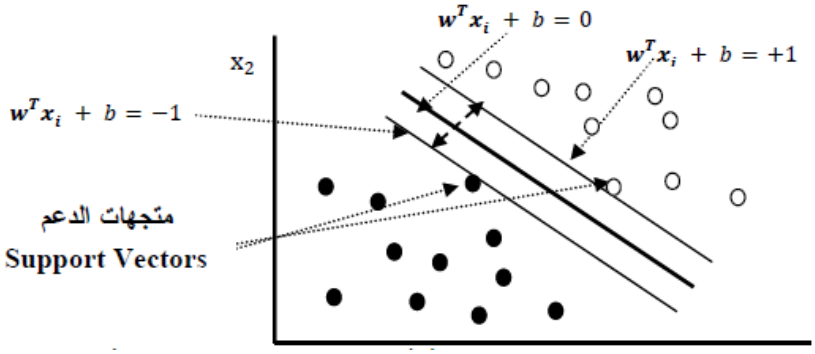
تصنف البيانات الأساسية في فضاء الإدخال وفق النموذج الرياضي الآتي:

$$W^T.x_i + b \geq +1 \quad for \ d_{i=} +1\ ,\ i = 1,2,\dots n \qquad (١٣-٣)$$

$$W^T.x_i + b \leq -1 \quad for \ d_{i=} -1\ ,\ i = 1,2,\dots n \qquad (١٤-٣)$$

إذ تمثل W متجه الأوزان، x متجه الإدخال، b قيمة التحيز، وd تمثل قيمة الإخراج.

كما يمكن ملاحظة معادلات الحدود في المستوى من الشكل الآتي:



الشكل (٣-١٣) متجهات الدعم الواقعة على الحد الفاصل بين البيانات في الفضاء [49].

١-المستوي الفائق Hyperplane:

من الممكن أن يكون المستوي الفائق ممثلاً في فضاء ثنائي البعد، أي عبارة عن خط مستقيم إذا كان عدد السمات features المدخلة ٢، أو في فضاء ثلاثي البعد (مستوٍ)، إذا كان عدد السمات المدخلة ٣، أو في فضاء مكون من n بعد إذا كان عدد المزايا المدخلة n.

ومن الشكل السابق نلاحظ أن معادلة المستوي الفاصل، تكتب بالشكل الآتي:

$$W^T.x_i + b = 0 \qquad (١٥-٣)$$

٢-متجهات الدعم Support Vectors:

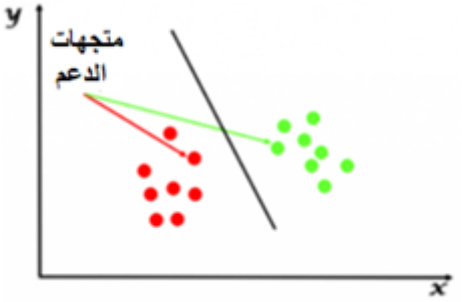
تعد متجهات الدعم هي مجموعة نقاط البيانات الأقرب على المستوي الفائق، وهي بمثابة نقاط ذات أهمية كبيرة في عملية تصنيف البيانات، لأنه يتم من خلالها تحديد المستوي الفائق الأفضل في عملية فصل البيانات. وبالتالي إزالتها أو تغيير موقعها يتطلب تحديد مستوى فائق آخر من جديد [49، 50].

ويمكن حساب المسافة بين النقاط في المستوي ومعادلة المستوي الفاصل من خلال العلاقة الآتية:

$$d(w,b,x_i) = \frac{|W^T.x_i+b|}{||w||} \qquad (١٦-٣)$$

وبعد عدد من الإجراءات والتحويلات الرياضية، يتم إيجاد قيم كل من متجه الأوزان المثالي W^* والتحيز المثالي b^*، ومن بعدها تحسب دالة التصنيف الآتية:

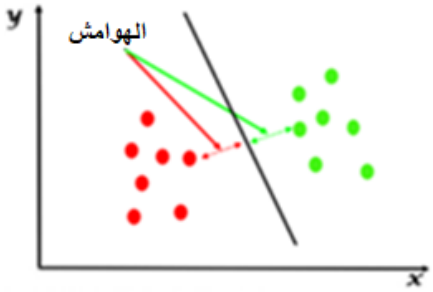
$$f(x) = I(W^*.x + b^*) \qquad (١٧-٣)$$



الشكل (٣-١٤) متجهات الدعم [50].

٣-الهامش Margin:

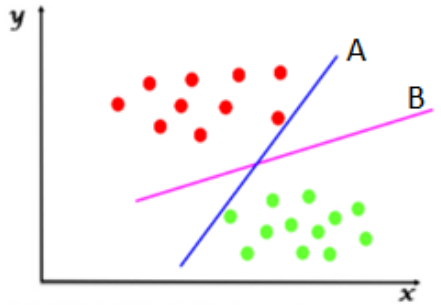
هو عبارة عن المسافة بين المستوي الفائق، وأقرب نقطة من نقاط مجموعة البيانات. كما هو موضح في الشكل (٣-١٥):



الشكل (٣-١٥) الهوامش [50].

وكلما كانت هذه المسافة كبيرة، زادت احتمالية تصنيف بيانات جديدة بشكل صحيح، كما هو واضح في الشكل (٣-١٦)، فنلاحظ أن المستوى الفائق B هو أفضل من المستوي الفائق A، وذلك لأن المسافة بينه

وبين أي نقطة من نقاط متجهات الدعم أكبر ما يمكن. وهذا ما يتيح إمكانية تصنيف أكبر عدد من النقاط بشكل صحيح. وبالتالي تكمن أهمية الهامش في الحصول على الموضع الأفضل للمستوي الفائق. وهذا ما كنا نقصده في الأعلى بعبارة المستوي الفائق الأفضل.



الشكل (٣-١٦) تقييم المستوي الفائق الأفضل واختباره [50].

٣-٨-٣-١-أنواع متجهات دعم الآلة:

١-متجه دعم الآلة الخطي Linear SVM:

عبارة عن مصنف classifier يتم استخدامه لفصل البيانات التي تتميز أنها قابلة للفصل خطيا، بواسطة مستقيم، مهمته فصل مجموعة البيانات إلى مجموعتين.

ويدعى هذا المصنف Linear SVM Classifier أي مصنف متجه دعم الآلة الخطي. وينص مبدأ عمله على إيجاد القيم المثالية لمعادلة الخط المستقيم التي تفصل هذه البيانات بأفضل شكل ممكن كما في الشكل (٣-١٧)، أي بأكبر مسافة للهامش ما بين أي نقطة من نقاط متجهات الدعم والمستوي الفائق [50، 51].

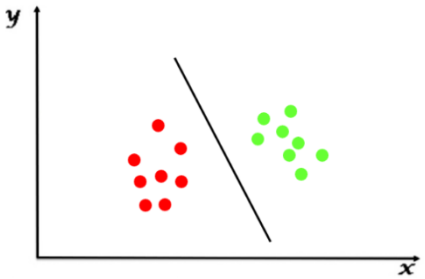
ويتم تمثيل المستوى الفاصل في خوارزمية SVM الخطية بالشكل التالي:

$$f(x) = W^T.X = 0 \qquad (٣-١٥)$$

وتعبر W عن متجه الأوزان، X بيانات التدريب التي يمكن تمثيلها رياضياً بالشكل الآتي:

$$\{(x_1, y_1), (x_2, y_2), \dots (x_n, y_n)\} \in R^n \qquad (٣-١٦)$$

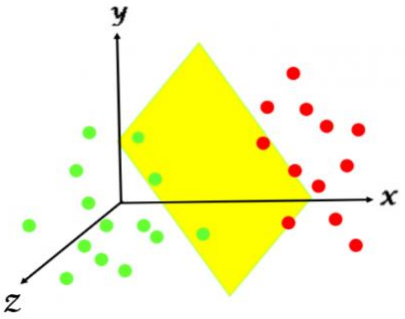
إذ x_n متجه السمات، y_n الصنف الخاص ب x_n فعلى سبيل المثال إذا كان $y_n = 1$، هذا يعني أن البيانات تنتمي إلى الصنف الأول.



الشكل (٣-١٧) فصل مجموعة البيانات خطيا (بواسطة خط مستقيم) [50].

٢-متجه دعم الآلة غير الخطي Non Linear SVM:

عبارة عن مصنف يتم استخدامه لفصل البيانات التي تتميز بأنها غير قابلة للفصل خطيا بواسطة مستقيم كما في الشكل (٣-١٨)، ويدعى هذا المصنف Non Linear SVM Classifier، أي مصنف متجه دعم الآلة غير الخطي. ويتم تطبيق طريقة النواة kernel لتحويل مجموعة البيانات من فضائها الحالي (الفضاء ثنائي البعد) إلى فضاء أعلى، وذلك بإضافة بعد جديد، ومن ثم اختبار قابلية فصل البيانات، ففي حال التحقق تبحث الخوارزمية عن المستوي الفاصل المناسب لعملية الفصل، وإلا تستمر الخوارزمية بإضافة أبعاد، ومن ثم اختبار البيانات.



الشكل (٣-١٨) فصل مجموعة بيانات غير قابلة للفصل الخطي بواسطة مستوٍ [50].

ومن أهم التوابع المستخدمة في طريقة النواة kernel:

Radial Basis Function (RBF) kernel, Polynomial kernel, Linear kernel.

٣-٨-٣-٢-إيجابيات خوارزمية متجه دعم الآلة وسلبياتها:

الإيجابيات:

١-تعمل بشكل فعال مع البيانات القابلة للفصل الخطي، أي التي تملك هامش فصل واضح.

٢-تعطي نتائج أكثر دقة في حال كان عدد سمات بيانات التدريب أكبر من عدد أمثلة التدريب.

٣-كفاءتها عالية لأنها تستخدم جزءاً من بيانات التدريب في اتخاذ القرار، ويقصد بذلك اختيار المستوي الفائق الأفضل اعتماداً على نقاط البيانات الواقعة على المتجهات الداعمة فقط، وليس اعتماداً على كامل بيانات التدريب.

السلبيات:

١-لا تعمل بشكل جيد عندما يكون لدينا مجموعة بيانات كبيرة، لأن المدة الزمنية اللازمة لتدريبها عالية.

٢-لا تعمل بشكل جيد للغاية عندما تحوي مجموعة البيانات أخطاء كثيرة، سواء كانت ضجيجاً في أثناء جمع البيانات أو متداخلة بشكل كامل، ولذلك تم حل هذه المشكلة من خلال اعتماد مرحلة المعالجة الأولية للصورة.

٣-٨-٣-٣-استخدامات خوارزمية متجه دعم الآلة:

تستخدم هذه الخوارزمية في تطبيقات تصنيف الصور وتجزئتها، وتصنيف النصوص، وتصنيف في تطبيقات معالجة اللغات الطبيعية، وتصنيف البريد الإلكتروني، فيما إذا كان مزيفاً أو حقيقياً، وكذلك تحليل الآراء وغيرها. كما تستخدم في التطبيقات البيولوجية وغيرها من العلوم. والتطبيق الأشهر استخدامها في التعرف على الأرقام المكتوبة بخط اليد في نظام البريد.

بما أن بيانات مسألتنا (تصنيف الأورام) متداخلة نوعاً ما فنحن بحاجة إلى طريقة لفصل بيانات غير خطية، لذلك اعتمدنا في بحثنا على خوارزمية SVM، بالإضافة إلى كفاءة الخوارزمية العالية في اتخاذ القرار بالاعتماد على جزء من بيانات التدريب وليس على كاملها، بينما استخدمنا التابع الخطي Radial Basis Function ('rbf') في طريقة النواة kernel لاعتماده على فضاء سمات التدريب الأساسية من دون أي تغير، كما في التوابع الأخرى ما يجعله الأكثر دقة، وهذا يناسب مسألتنا [52].

٣-٩-خلاصة:

أخذنا في هذا الفصل لمحة نظرية عن جميع التقنيات والخوارزميات المستخدمة في هذا البحث، إذ تم الحديث عن تقنيات معالجة الصور الرقمية، وتخفيض الضجيج في الصورة، وذكر خصائصها وتطبيقاتها. كما تم شرح أهم طرائق التجزئة الرقمية والفرق بينها، وأخيراً قدمنا شرحاً تفصيلياً عن خوارزمية الذكاء الاصطناعي المقترحة من أجل عملية تصنيف الأورام، وسنتحدث في الفصل القادم عن نظام الكشف المقترح في البحث وخطواته الأساسية.

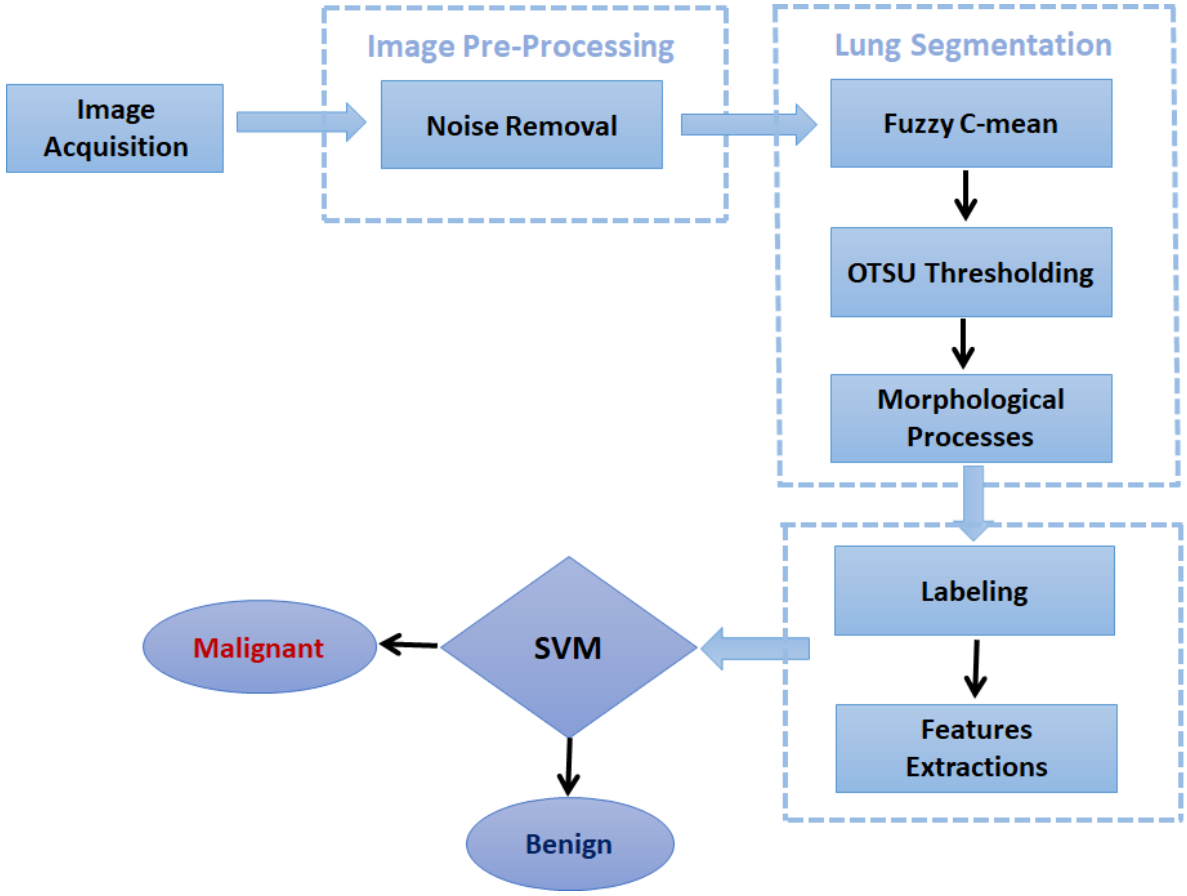
الفصل الرابع

الخوارزمية المقترحة للكشف عن العقيدات الرئوية

٤ – ١ –مقدمة:

سنتحدث في هذا الفصل عن الخوارزمية المقترحة للكشف عن العقيدات الرئوية، وسنشرح جميع مراحل النظام بالتفصيل، مع عرض تطبيق كل مرحلة من هذه المراحل على الصور الطبية المستخدمة.

ويقدم البحث مخبراً افتراضياً لتحليل صور الأشعة للرئتين باستخدام عمليات معالجة الصور الرقمية وتطوير خوارزمية تقوم بإجراء عمليات المعالجة الأساسية التي تتيح توصيف الصور، واستخراج خصائصها وسماتها واستخلاص المعلومات الهامة منها كخطوة أولى في عملية تشخيصها، إذ يتكون النظام المقترح للكشف الآلي عن العقيدات السرطانية من عدة خطوات مبينة على النحو الآتي:

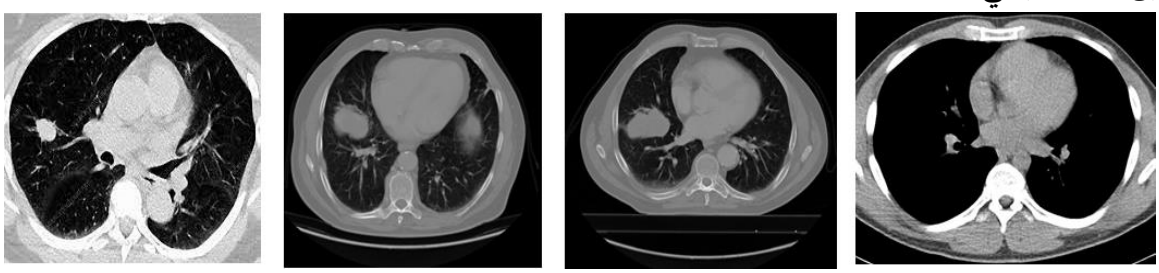


الشكل (٤–١) خطوات نظام الكشف عن العقيدات المقترح.

٤-٢-استحصال الصورة Image Acquisition:

تم الحصول على الصور المقطعية المحوسبة CT من موقع قاعدة بيانات Kaggle. وتتكون البيانات من 137 صورة على الشكل الآتي:
43 صورة ذات أورام (عقيدات حميدة)، 49 صورة ذات عقيدات خبيثة، و23 صورة تحوي أكثر من ورم في الصورة نفسها، بالإضافة إلى 22 صورة طبيعية لا تحوي أوراماً.

تم اختيار التصوير المقطعي المحوسب CT لما له من فوائد واسعة في تخطيط العلاج الإشعاعي. كما توفر كثافة إلكترونية للأنسجة ذات الأهمية، وتحقق دقة مكانية جيدة وتبايناً يسمح بالتحديد الدقيق للأهداف المرجوة، فضلاً عن أنها تحوي ضجيجاً منخفضاً ووضوحاً أفضل مقارنة بصور الأشعة السينية والتصوير بالرنين المغناطيسي.



الشكل (٤-٢) عينة بسيطة من قاعدة البيانات المستخدمة.

٤-٣-المعالجة الأولية للصورة:

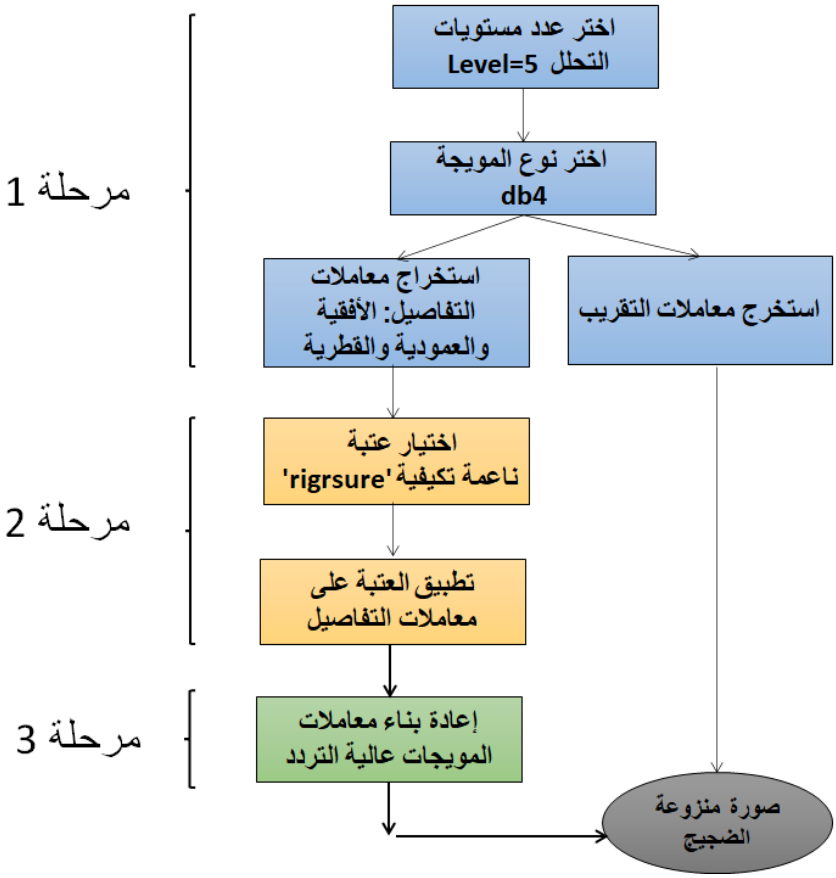
المعالجة الأولية هي مرحلة أساسية في اكتشاف عقيدات الرئة، والهدف الرئيسي من معالجة الصورة هو تحسين جودتها، وجعلها أكثر ملاءمة وسهولة لتطبيق معين. ومن أهم الخطوات الأساسية لتحسين جودة الصور هي مرحلة **إزالة الضجيج**.

٤-٣-١-إزالة الضجيج باستخدام تحويل المويجات:

بما أن الصور الطبية عبارة عن مجموعة من المناطق المتصلة ذات التدرجات الرمادية المتباينة، والتي تحتوي على عدة عناصر ومكونات صغيرة وكبيرة ذات ترددات عالية (الحواف) وأخرى ذات ترددات منخفضة، فإن تحويل المويجة يعد مناسباً لهذا النوع من الإشارات بسبب قدرته على تحليل الأجزاء المختلفة من الإشارة بمقاييس وترددات مختلفة، وهذا يساعدنا على القيام بعمليات تفيد في عمليات التعرف وحفظ الإشارة، وإزالة الضجيج، واستخلاص ميزات مفيدة [53].

٤-٣-٢-خطوات إزالة الضجيج باستخدام تحويل المويجات:

يبين المخطط الآتي مراحل إزالة الضجيج باستخدام تقنية التحويل المويجي [54، 55]:



الشكل (٤-٣) مراحل إزالة الضجيج باستخدام تقنية تحويل المويجات [54].

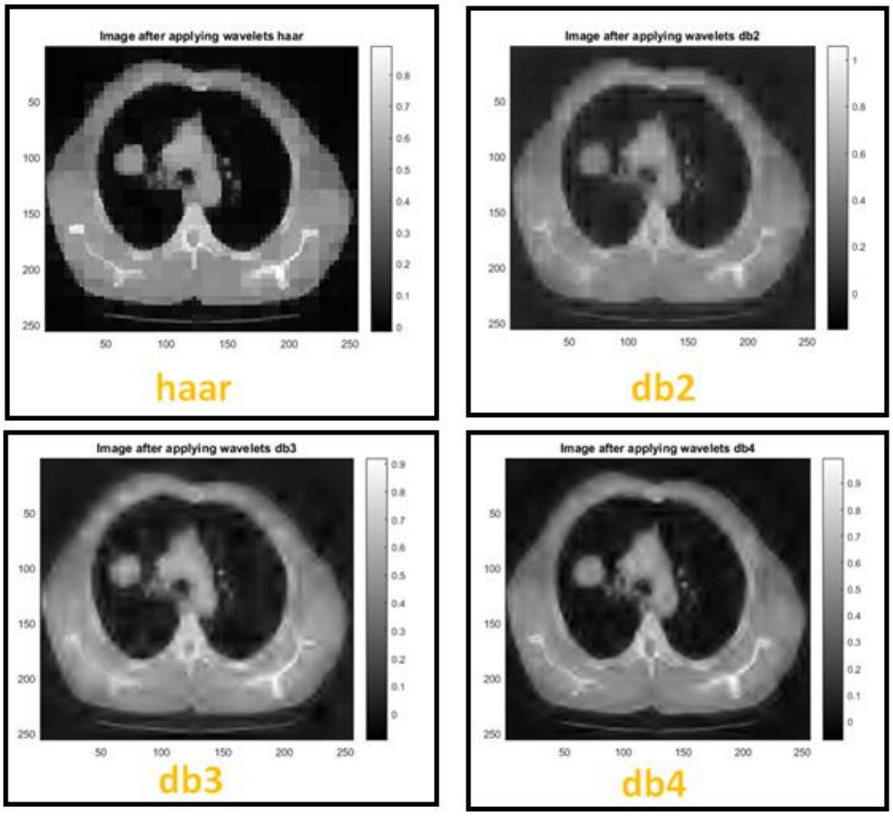
نختبر النتائج على الصورة الآتية بتغيير نوع المويجة المستخدمة: Haar (db1), db2, db3, db4، وعند عدة مستويات: Level = 1, 2, 3, 4, ...، ليتم اختيار المويجة الأفضل، والمستوى الأنسب لعملية التعتيب.



الشكل (٤-٤) الصورة المراد معالجتها.

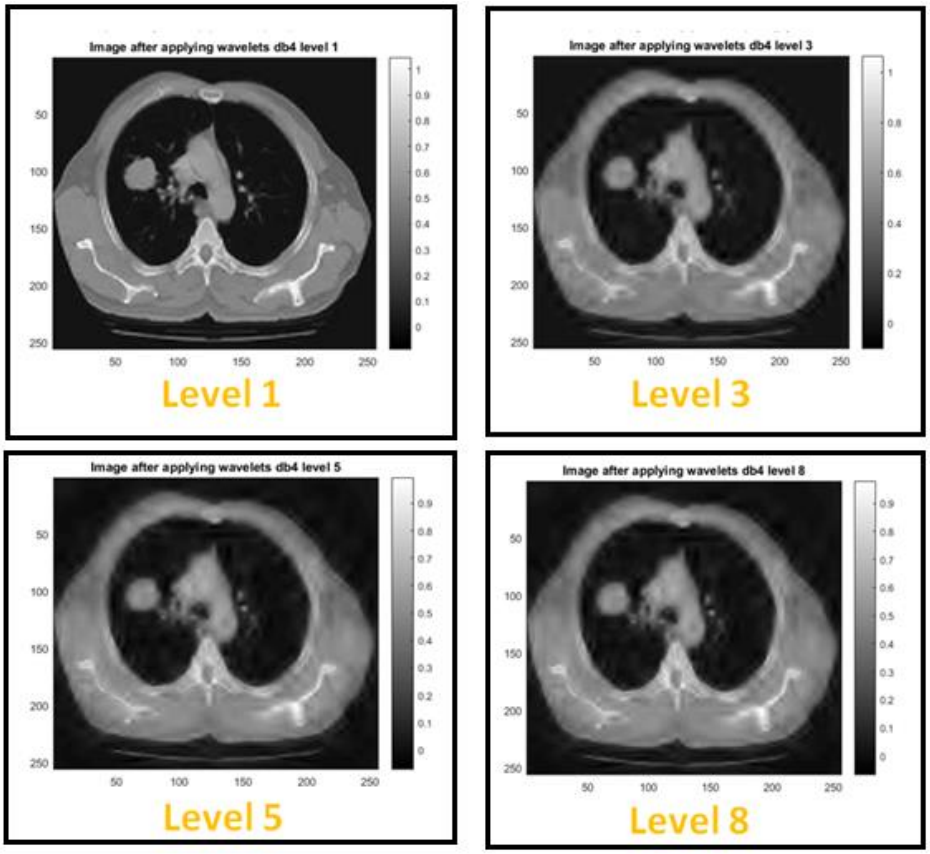
تعطي المويجة db4 نتائج أفضل وأكثر جودة من بقية المويجات في إزالة الضجيج والشوائب (الحويصلات الرئوية) والتباين بين السويات الرمادية للمناطق عالية الكثافة التي تحتوي على العظام والعضلات الدهنية مع المحافظة على المعلومات الهامة فقط في الصورة (الحواف) والشكل الخارجي للأورام بدون تشويه.

وثم نقوم بتطبيق هذه المويجة عند عدة مستويات على الصورة نفسها، للحصول على أفضل نتيجة للتحويل المويجي.



الشكل (٤-٥) ناتج تطبيق عدة أنواع للمويجات على إحدى عينات قاعدة البينات.

بعد إجراء عملية التحلل باستخدام المويجة db4 عند عدة مستويات Level = 1, 2, 3, 4, 5, 6, 7, 8، تم الحصول على صورة مضغوطة وناعمة بدءاً من المستوى Level 5 وثبات النتائج عند هذا المستوى، لذلك نقوم بالتوقف عنده، واعتماد نوع المويجة db4، والمستوى الخامس، على بقية صور قاعدة البيانات كافةً، وبالتالي سوف تساعد طريقة التحويل المويجي على تقسيم الورم بسهولة أكبر.



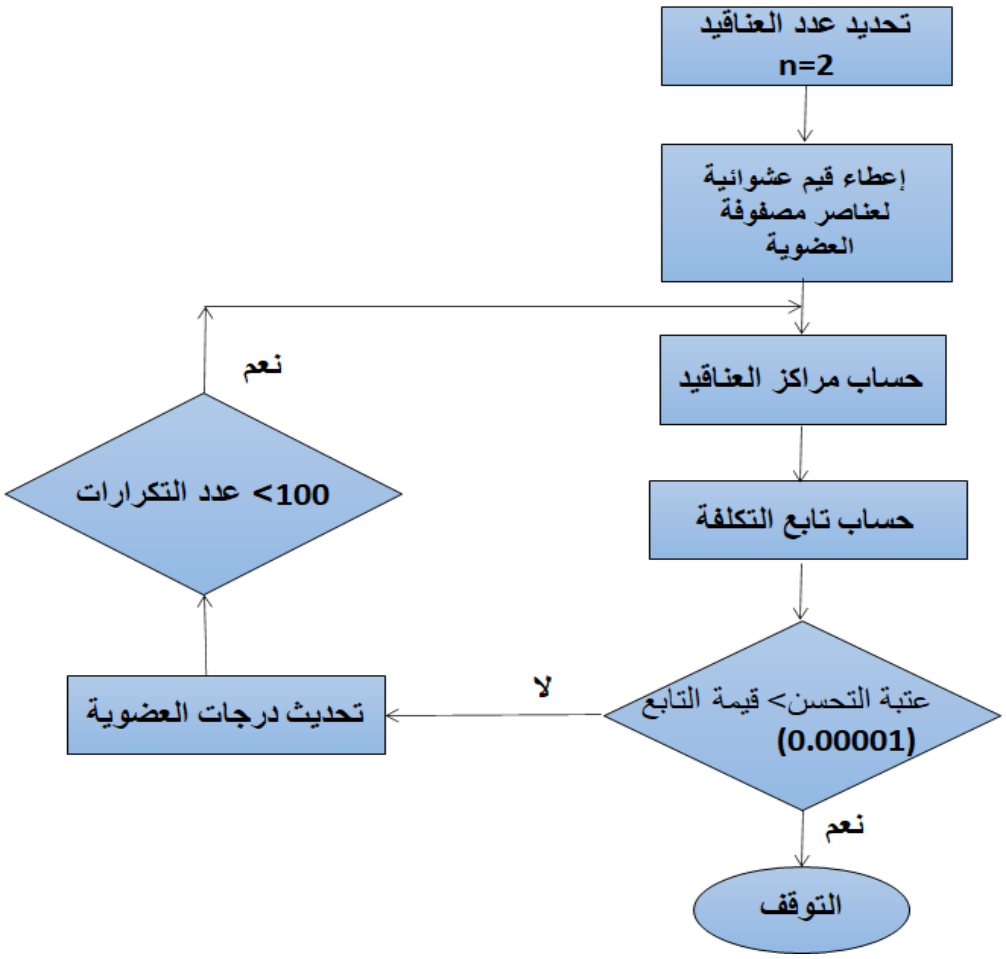
الشكل (٤-٦) ناتج إجراء عملية التحليل المويجي باستخدام نوع المويجة db4 عند عدة مستويات.

٤-٤-مرحلة تجزئة الصورة:

الهدف الرئيسي من التجزئة هو تبسيط أو تغيير تمثيل الصورة إلى شيء أكثر جدوى، وأسهل في التحليل، ويستخدم عادة لتحديد مواقع الكائنات والحدود بشكل أدق والكشف عن التشوهات (الأورام).

وتتضمن صور CT المناطق الورمية التي يجب استخلاصها من أجل تحليلها وإعطاء التشخيص المناسب لها، وتعد هذه العملية عملية صعبة نظراً لطبيعة هذه الأورام من حيث الشكل. وبالنظر إلى مجموعة عناصر الصورة على أنها مجموعة من البيانات التي تشترك مع بعضها بخصائص معينة (اللون، قيم السويات الرمادية)، يمكن استخدام خوارزمية عنقدة البيانات من أجل تجميع عناصر المجموعة الواحدة مع بعضها، وتختلف مع عناصر المجموعات الأخرى [56].

ويبين المخطط الآتي مراحل تطبيق خوارزمية FCM المقترحة في البحث:



الشكل (٤-٧) خطوات خوارزمية FCM المقترحة [57].

وتعد خوارزمية FCM أهم خوارزميات العنقدة التي تستخدم في معالجة الصور الطبية، ولا تعتمد على قيم عتبات ثابتة للسويات الرمادية، بل يتم التعامل مع كل حالة على حدة، ما يعطي نتائج جيدة نتيجة وجود تباين في قيم السويات الرمادية للصور الطبية ناتج عن عوامل عدة مثل بارامترات عملية التصوير والمواصفات الجسمية للمريض.

ولتطبيق تابع FCM على الصورة السابقة، تم اختيار عدد العناقيد **n=2** هما: العنقود التي سيحتوي على العناصر ذات الأهمية (المنطقة الورمية) والعنقود الأقل أهمية (خلفية الصورة السوداء) التي سوف تزال لاحقاً، وتم تحديد القيمة الأصغرية للتحسن (الافتراضية) **Improvement Value = 0.00001**. أما العدد الأعظمي للتكرارات **Iteration = 100**. ويبين الجدول (٤-١) نتائج تنفيذ التابع FCM على الصورة السابقة.

الجدول (٤-١) قيم تابع التكلفة وقيم التحسن خلال تكرارات التابع FCM:

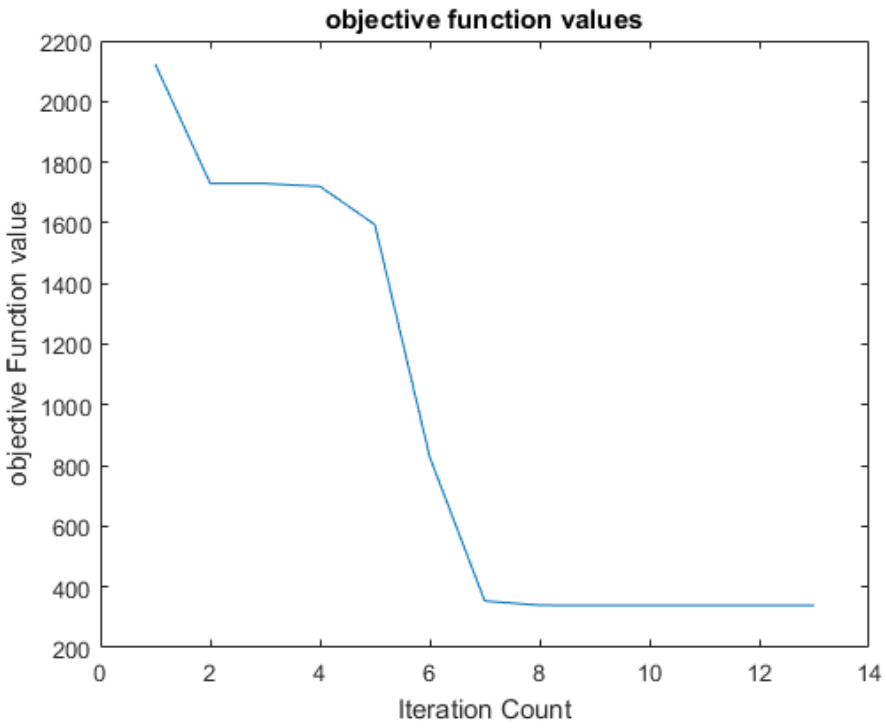
رقم التكرار Iteration count	قيمة تابع التكلفة Cost function	قيمة التحسن Improvement value
1	2123.460937	
2	1730.150396	393.310541
3	1729.540012	0.610384
4	1720.126132	9.41388
5	1594.124422	126.00171
6	827.446603	766.677819
7	353.703395	473.743208
8	339.652961	14.050434
9	339.340147	0.312814
10	339.326041	0.014106
11	339.325336	0.000705
12	339.325300	0.000036
13	339.325298	**0.000002**

واستغرق تنفيذ التابع زمناً قدره 0.389728 ثانية، وبطبيعة الحال سيتناسب زمن التنفيذ طرداً مع عدد التكرارات وعكساً مع سرعة المعالج المستخدم.

فقد بلغت قيمة التحسن في التكرار رقم 13 = Iteration count، وهي قيمة أصغر من القيمة الأصغرية المسموح بها (الفرق بين قيمتي التابع في تكرارين متتاليين):

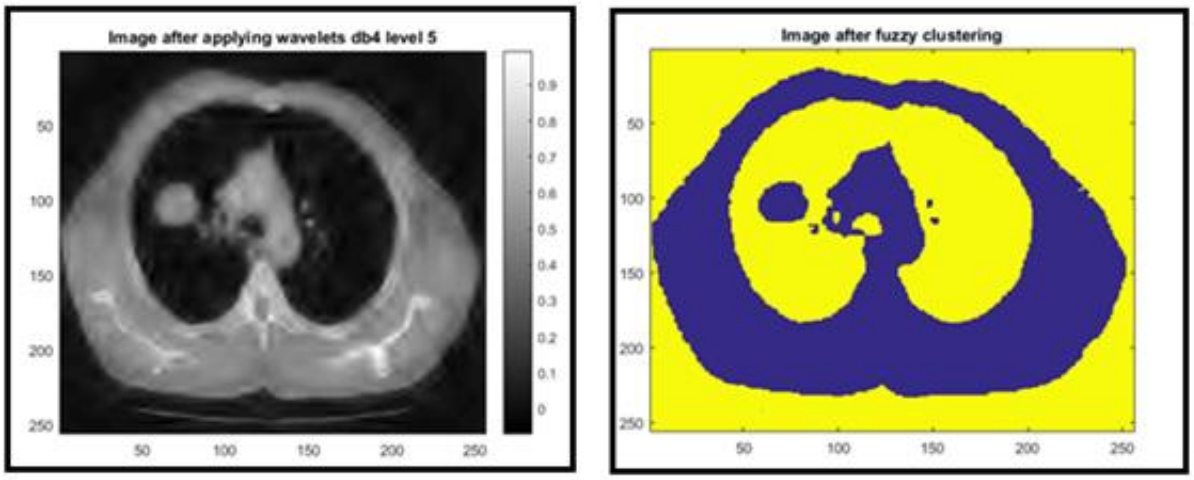
0.000002 = (339.325300-339.325298)، وبالتالي سوف تتوقف الخوارزمية عند هذا التكرار.

ويبين الشكل (٤-٨) مخططاً بيانياً لقيم تابع التكلفة خلال تكرار تنفيذ خطوات التابع FCM:



الشكل (٤-٨) مخطط بياني يبين قيم تابع التكلفة خلال تكرار تنفيذ خطوات تابع FCM.

ويبين الشكل الآتي ناتج تطبيق خوارزمية العنقدة FCM على الصورة:



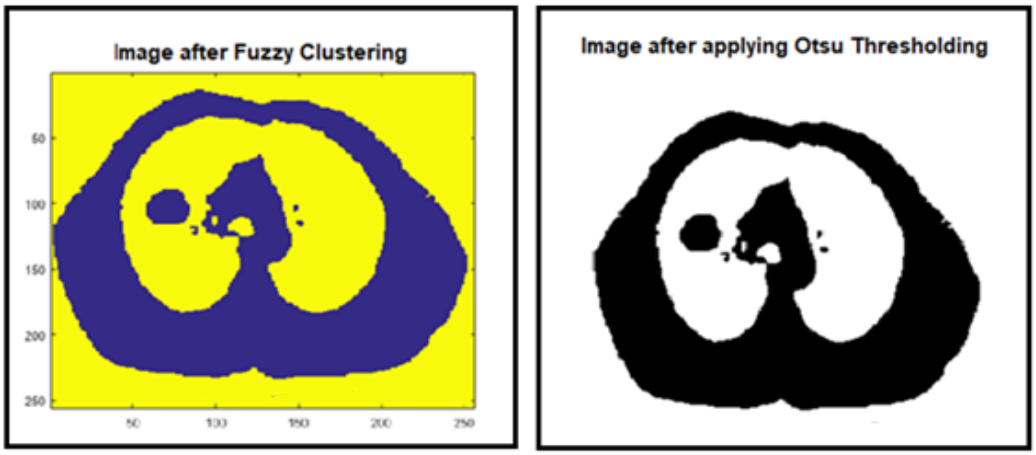
الشكل (٤-٩) ناتج تطبيق العنقدة FCM على الصورة.

٤-٤-١-التعتيب باستخدام Otsu Thresholding:

هي طريقة تعتيب إجمالية قائمة على العنقدة، وهي تحول صورة المستوى الرمادي إلى صورة ثنائية، سوف يتم استخدامها لاستخراج الكائنات ذات الأهمية (الأورام) من خلفية الصورة. وتعالج هذه الطريقة الرسم البياني للصورة، وتعمل على تقسيم الكائنات عن طريق تقليل التباين داخل العنقود الواحد، وزيادة التباين بين العناقيد المختلفة، كما أنها تقلل من تعقيد البيانات، وتجعل عملية التعرف والتصنيف أكثر بساطة.

وتم استخدام **عتبة Otsu** نظراً لبساطتها وبسبب الاختلاف في شدة المستويات الرمادية للصور الشعاعية للرئة التي تتطلب قيما مختلفة للعتبة. ولذلك تم تحديد العتبة تلقائياً لمختلف الصور باستخدام هذه العتبة.

ويبين الشكل الآتي ناتج تطبيق عتبة Otsu على الصورة الناتجة عن العنقدة:



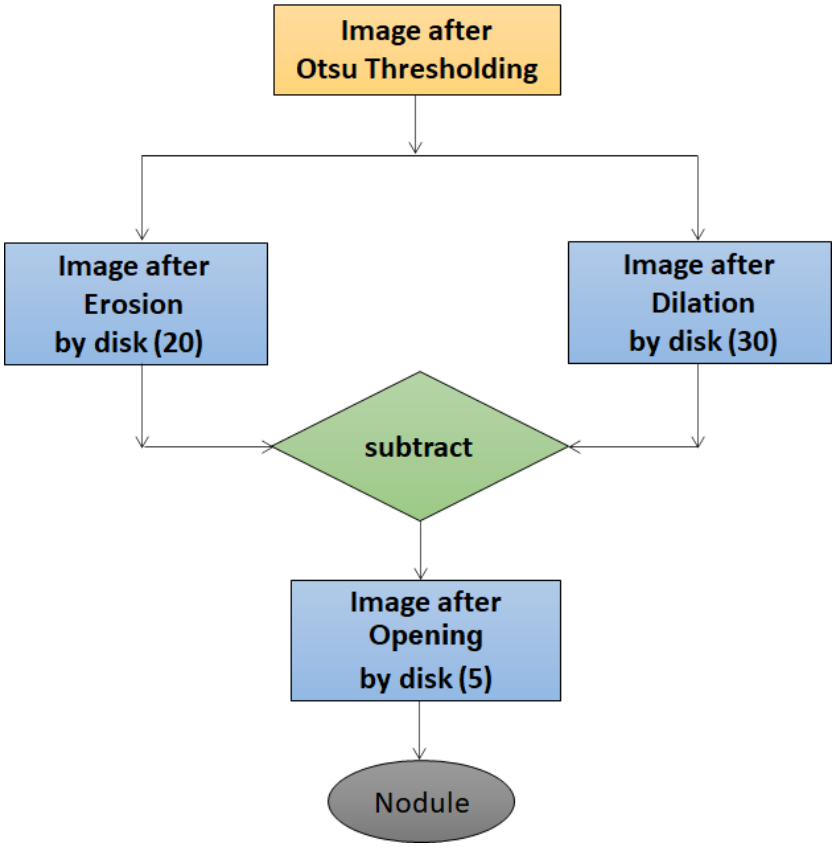
الشكل (٤-١٠) ناتج تطبيق عتبة Otsu على الصورة.

ونلاحظ أنه تم الحصول على صورة ثنائية بعد استخدام هذه التقنية، إذ تم اكتشاف حواف الصورة بشكل صحيح، وهي تسمح بفصل الرئتين والأجزاء الأخرى في الصورة.

٤-٤-٢-استخلاص الأورام من الصور المجزأة باستخدام العمليات المورفولوجية:

عادة ما تخضع الصور للعمليات المورفولوجية لإزالة المكونات المتصلة (الكائنات) غير المرغوبة في الصورة، إذ يتم تطبيق معايير عدة لمحاولة التخلص قدر الإمكان من هذه المكونات، مع الأخذ بالحسبان أن تطبيق هذه المعايير يجب أن يكون دقيقاً من أجل عدم التأثير في دقة النتائج كأن يؤدي تطبيق أحد هذه المعايير على سبيل المثال إلى حذف منطقة مهمة (ورمية) من الصورة إلى جنب مناطق متصلة غير مرغوبة [58].

إذ تم في هذه البحث استخدام عمليتي التمدد والتعرية لعزل وإزالة المنطقة الورمية من الصورة لاسترجاعها لاحقاً، بينما يتم الحفاظ على الأجزاء الأخرى فقط في الصورة. وفي التقنيات اللاحقة سيتم طرح هذه الصورة من الصورة التي بها الورم بالإضافة إلى أجزاء أخرى. وسنحصل بالنتيجة على الورم على شكل صورة مجزأة.

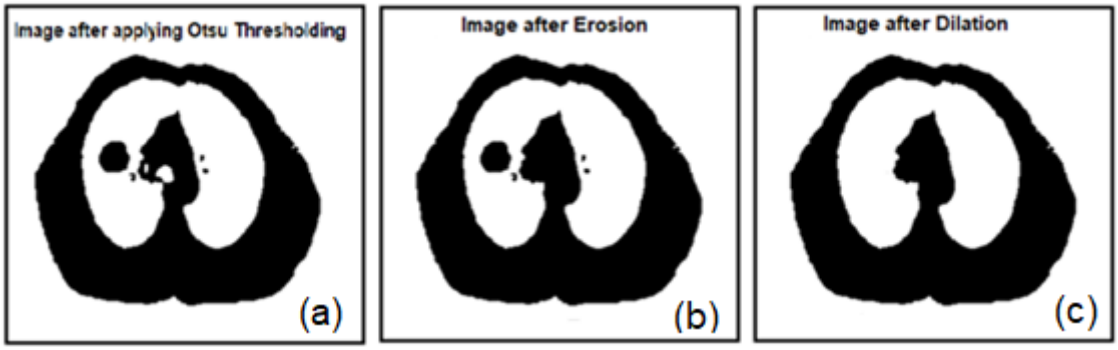


الشكل (٤-١١) مراحل استخلاص المنطقة الورمية باستخدام العمليات المورفولوجية.

ويمكن تلخيص خطوات استخلاص الأورام من الصور المجزأة وفق الآتي:

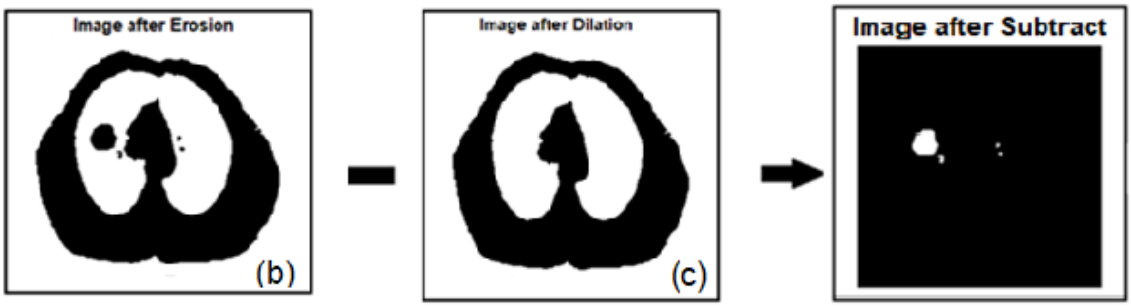
١-تطبيق عملية تعرية **erosion** باستخدام عنصر تشكيل ذي حجم 20 بكسل من نوع disk، لإزالة المناطق البيضاء ذات المساحات الصغيرة غير المرغوب بها، إذ يتم حذف كل عنصر ذات قيمة أقل من القيمة المحددة التي تمثل شرطاً للإزالة.

٢-تطبيق عملية تمدد **dilation** على الصورة الناتجة باستخدام عنصر تشكيل ذي حجم 30 بكسل من نوع disk، لإزالة المنطقة السوداء التي تمثل الورم المراد استخلاصه، إذ أنه بعد تحليل مساحات المناطق الورمية في العديد من الصور الشعاعية للرئة، تم استنتاج قيمة مساحة معيارية أصغرية تجريبياً هي 30 بيكسل والتي تقابل مساحة أكبر موجود حوالي 2000 بكسل.



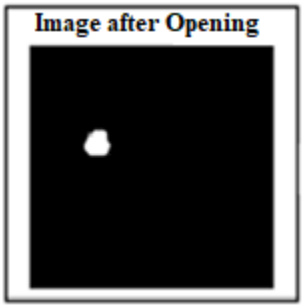
الشكل (٤-١٢) تطبيق عمليتي (b) التعرية و(c) التمدد على الصورة المجزأة (a).

٣-إجراء عملية طرح **subtract** بين الصورتين الناتجتين عن الخطوة ١ و٢، إذ تكون قيم البكسلات الناتجة هي تلك القيم الخاصة بالصورة الأولى مطروحاً منها قيم البكسل المقابلة من الصورة الثانية.



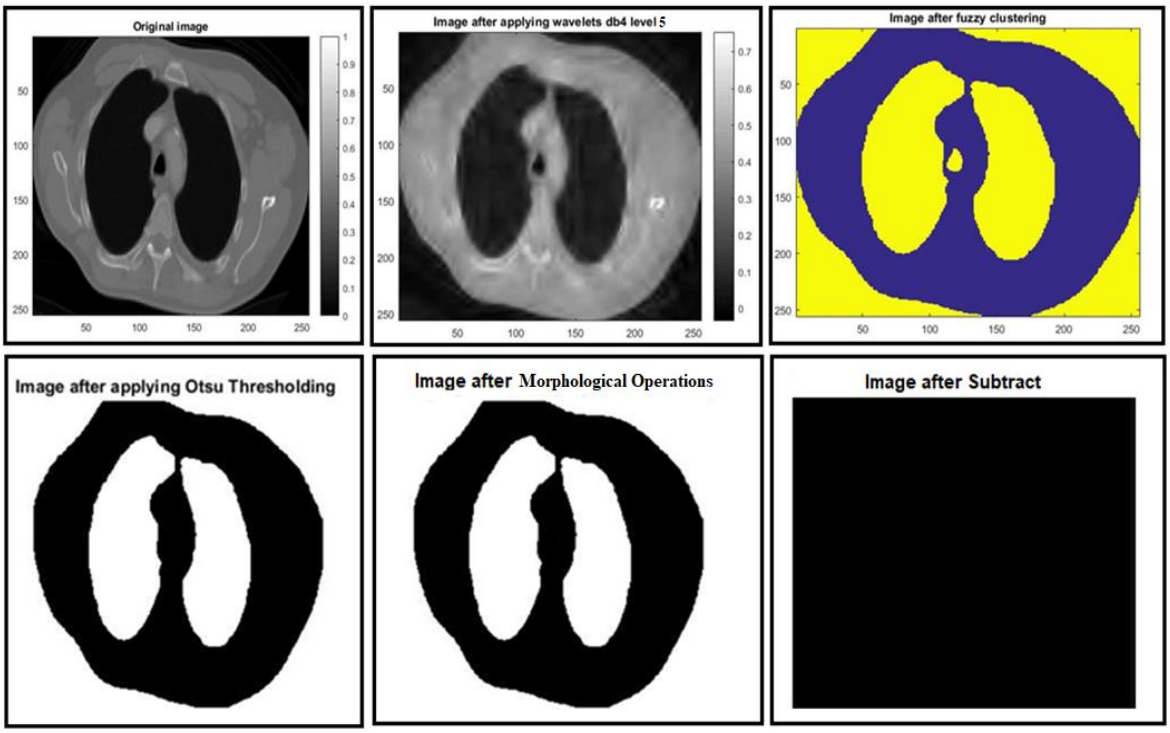
الشكل (٤-١٣) إجراء عملية الطرح من أجل استخلاص الورم.

٤-إجراء عملية فتح **Opening** باستخدام عنصر تشكيل ذي حجم 5 بكسل من نوع disk على الصورة لإزالة الأجزاء الصغيرة المتبقية والتي لا تمثل جزءاً من الورم، مع المحافظة على حجم الورم دون تغيير.



الشكل (٤-١٤) ناتج تطبيق عملية الفتح على الصورة المجزأة.

وبعد تطبيق هذه التقنية يمكننا الحصول على صورة توضح المنطقة المجزأة ذات الاهتمام (منطقة الورم) بالنسبة إلى الصور غير الطبيعية (الورمية)، بينما نحصل على صورة سوداء كاملة بالنسبة للصورة الطبيعية التي لا تحوي أية أورام. كما هو موضح في الشكل الآتي:

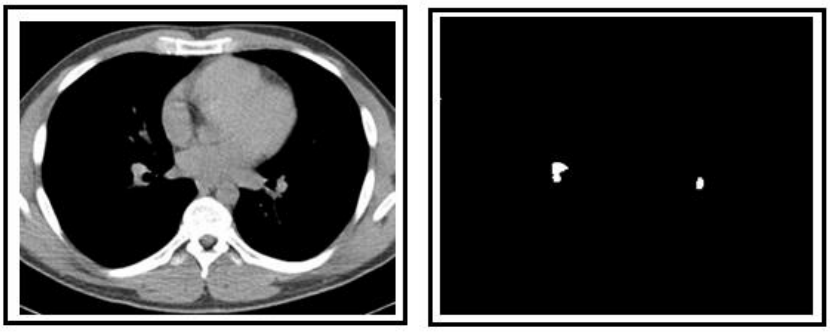


الشكل (٤-١٥) تطبيق الخوارزمية المقترحة على صورة طبيعية (من دون ورم).

٤-٤-٣-خوارزمية العنونة Labeling:

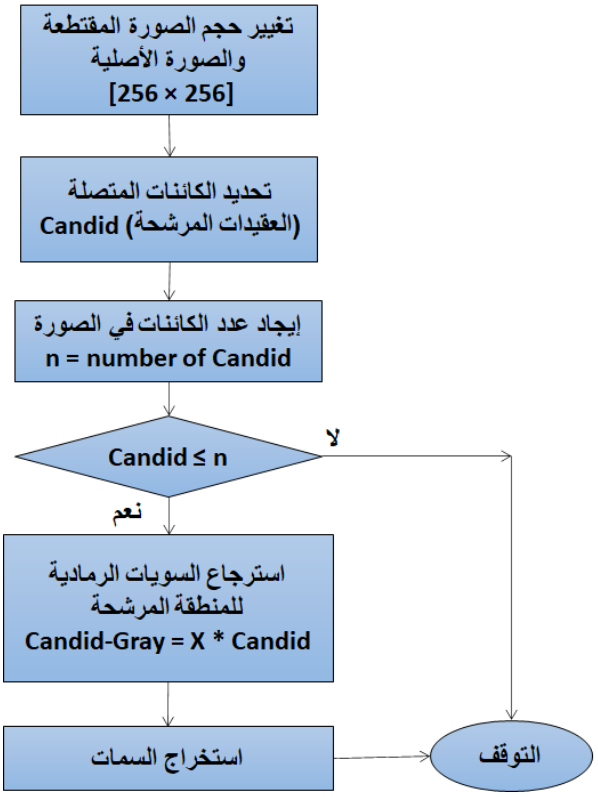
لكي يصبح النظام قادراً على اكتشاف الصور التي تحتوي على أكثر من منطقة مشبوهة وتصنيفها، يجب تحديد كل ورم على حدة، واستخلاص السمات الخاصة بكل منطقة، ومن ثم تشخيصها.

لذلك تم اللجوء إلى تقنية عنونة الصورة التي تسمح بتحديد المكونات المتصلة Candids وإيجاد عددها في الصورة الثنائية اعتماداً على مفهوم الاتصالية الرباعية أو الثمانية بين البيكسلات المتصلة أو المتجاورة، إذ تم اعتماد الاتصالية الثمانية 8-Connectivity في بحثنا، لأنها تعطي نتائج محيطية أكثر دقة مع المحافظة على حواف الصورة. ويبين الشكل الآتي استخلاص منطقتين ورميتين مشبوهتين من صورة واحدة:



الشكل (٤-١٦) تطبيق الخوارزمية المقترحة على صورة تحوي ورمين معاً.

٤-٥-استخراج السمات Feature Extraction:

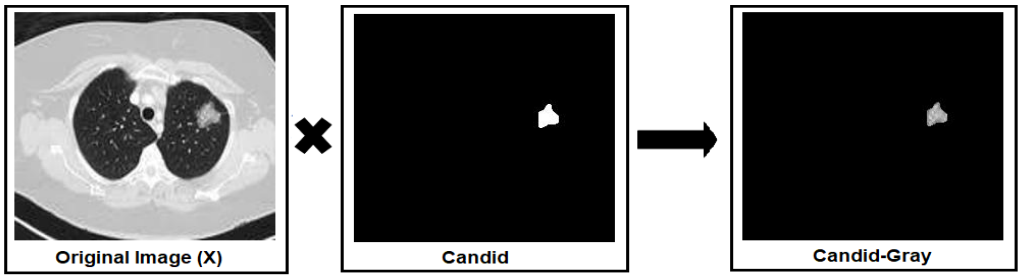


الشكل (٤-١٧) مرحلة استخراج السمات.

للقيام بعملية تحليل المناطق المريبة المستخرجة وتشخيصها، تم استخراج عدد من سمات المنطقة والواصفات الشكلية والرياضية، كالمساحة ودرجة الاستدارة وغيرها، وذلك اعتماداً على نوع السمة بما يتوافق مع طبيعة الصورة إما ثنائية Binary (تحتوي على اللونين الأسود 1 والأبيض 0)، أو ذات تدرجات رمادية Grayscale (تحتوي الأبيض والأسود مع تدرجات رمادية ذات قيم من 0 إلى 255)، والتي ستتم عملية التصنيف بناء عليها. ويجب مراعاة خصائص بعض السمات التي لا يمكن الحصول عليها من الصور الثنائية، وإنما تتطلب خصائصها استرجاع السويات الرمادية للمنطقة المرشحة. وتم تحقيق هذه الخطوة من خلال ضرب الصورة المقتطعة بالصورة الأصلية،

$$\textbf{Candid-Gray} = \textbf{X} * \textbf{Candid}$$

وتطبق هذه العملية على كل كائن Candid من الكائنات [59].



الشكل (٤-١٨) استرجاع السويات الرمادية للمنطقة المرشحة.

ونلاحظ من الشكل (٤-١٨) أنه تم الحصول على صورة الورم بمستوياته الرمادية الأصلية Candid-Gray مع المحافظة على حجمه وأبعاده كما هي في الصورة الثنائية Candid، وقد ساعد على ذلك توحيد حجم كل من الصورة المقتطعة والأصلية لكي تتطابق إحداثيات الورم الموجود في كل من الصورتين.

ويبين الجدول التالي أنواع السمات المستخدمة في تصنيف الأورام المكتشفة، بالإضافة إلى وصفها وتحديد خصائص المنطقة المستخرجة منه:

الجدول (٤-٢) السمات المستخدمة في تصنيف الأورام المكتشفة [60، 61، 62]:

	السمة	الوصف	مكان الاستخراج
1	**المساحة Area**	عدد عناصر الصورة في المنطقة الورمية	الصورة الثنائية BW
2	**المحيط Perimeter**	عدد عناصر الصورة المحيطية للمنطقة الورمية	الصورة الثنائية BW
3	**الانضغاط Compactness**	$Periemeter^2/4\pi \times Area$	الصورة الثنائية BW
4	**المحور الأعظمي Major Axis Length**	طول المحور الأعظمي للقطع الناقص	الصورة الثنائية BW
5	**المحور الأصغري Minor Axis Length**	طول المحور الأصغري للقطع الناقص	الصورة الثنائية BW
6	**الشذوذية Eccentricity**	معدل المسافة بين محوري القطع الناقص	الصورة الثنائية BW
7	**الاستدارة Circularity**	مقياس استدارة الكائنات	الصورة الثنائية BW
8	**القطر الأصغري Equiv Diameter**	قطر أصغر دائرة تحتوي المنطقة	الصورة الثنائية BW
9	**التباين Contrast**	مقياس الشدة بين العنصر ومجاوره	الصورة الرمادية Gray
10	**الانحراف المعياري Standard Deviation**	الجذر التربيعي للتباين	الصورة الرمادية Gray
11	**الترابط Correlation**	مقياس مدى الارتباط الخطي للعناصر المتجاورة	الصورة الرمادية Gray
12	**التجانس Homogeneity**	مقياس مدى قرب توزع العناصر	الصورة الرمادية Gray
13	**الطاقة Energy**	مقياس تناسق وتوزع القيم العظمى المشكلة للسطح	الصورة الرمادية Gray
14	**المتوسط Mean**	معدل كثافة العناصر في المنطقة الورمية	الصورة الرمادية Gray

وتبين الجداول الآتية قيم متجهات السمات المستخرجة لبعض الأورام في صور قاعدة البيانات، مع إضافة سمة الصنف الخاصة بكل ورم، إذ يشير الرقم 0 إلى أن الورم حميد، بينما يشير الرقم 1 إلى خباثة الورم.

الجدول (٤-٣) قيم متجهات السمات الثنائية لبعض الأورام في صور قاعدة البيانات:

الورم	المساحة	المحيط	الانضغاط	المحور الأعظمي	المحور الأصغري	الشذوذية	الاستدارة	القطر الأصغري	الصنف
1	653	93.156	13.289	35.764	23.424	0.794	0.945	65.166	0
2	256	54.956	11.797	20.097	16.319	0.774	1.065	41.736	0
3	293	68.470	16	25.869	14.877	0.793	0.785	52.541	0
4	196	52.186	13.894	17.442	14.738	0.767	0.904	43.812	0
5	358	68.916	13.266	24.919	18.545	0.794	0.947	52.654	0
6	394	78.254	15.542	25.320	20.455	0.789	0.808	57.389	0
7	363	77.956	16.741	23.544	21.332	0.775	0.750	55.941	1
8	910	160.692	28.375	55.076	25.922	0.7802	0.442	79.516	1
9	1105	189.596	32.530	58.674	28.743	0.7865	0.386	83.278	1
10	783	179.772	41.274	56.408	27.060	0.7982	0.304	80.091	1
11	284	71.770	18.137	24.002	17.371	0.7764	0.692	54.754	1
12	158	61.962	24.299	24.637	9.782	0.7802	0.517	47.801	1

الجدول (٤-٤) قيم متجهات السمات الرمادية لبعض الأورام في صور قاعدة البيانات:

رقم	التباين	الانحراف المعياري	الترابط	التجانس	الطاقة	المتوسط	الصنف
1	10835.699	104.094	0.910	0.993	0.848	0.514	0
2	176.853	13.298	0.878	0.989	0.801	0.527	0
3	1072.221	32.744	0.845	0.977	0.870	0.519	0
4	500.636	22.374	0.828	0.980	0.807	0.532	0
5	6026.894	77.633	0.845	0.978	0.842	0.524	0
6	1028.917	32.076	0.871	0.989	0.836	0.518	0
7	5556.437	74.541	0.885	0.980	0.848	0.547	1
8	381.447	19.530	0.857	0.991	0.837	0.511	1
9	2509.951	50.099	0.882	0.986	0.811	0.521	1
10	2213.035	47.042	0.853	0.983	0.802	0.525	1
11	871.697	29.524	0.861	0.976	0.818	0.518	1
12	2478.083	49.780	0.843	0.985	0.830	0.512	1

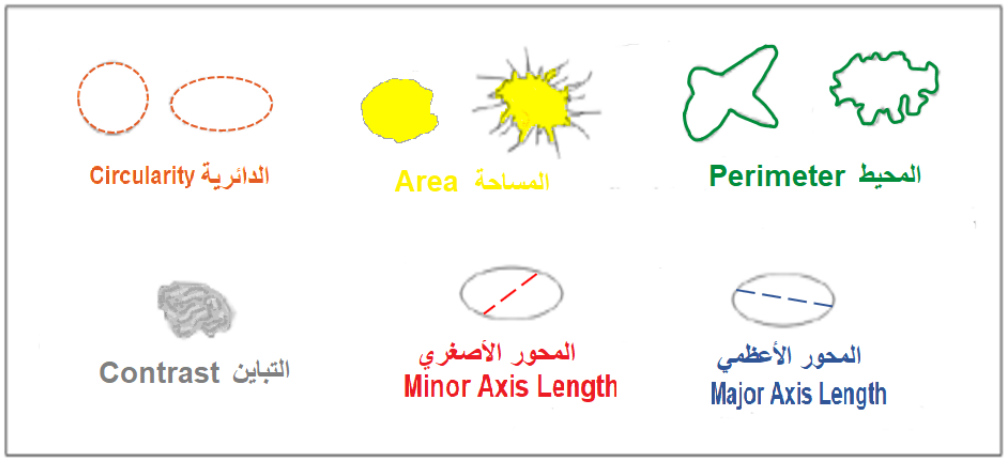
٤-٥-١-تقليل متجه السمات المستخرجة:

يمكن تحسين أداء خوارزمية التصنيف المقترحة عن طريق اختيار مجموعة السمات الأكثر أهمية لتمثيل العقيدات، لأن البيانات متعددة السمات تتضمن معلومات زائدة غير مفيدة، بل تؤثر على عملية التصنيف، إذ تزيد زمن التصنيف، وتخفض الدقة، كما تزيد مشكلة تخزين البيانات الزائدة [63].

بعد القيام باستخراج السمات المذكورة وتجريبها في الجدولين السابقين ودراسة قيمها. تم تقليل أبعاد شعاع السمات المستخدم كدخل للمصنف، وذلك من خلال حذف بعض السمات، إذ جرى حذف سمتي المتوسط والانحراف المعياري، لأن المتوسط يعطي قيمة عامة غير مفيدة لتحديد امتداد وتباين البكسلات في الصورة. كما أن الانحراف المعياري يعرف على أنه الجذر التربيعي للتباين، وبالتالي يكفي تحديد قيمة التباين كي يتحدد الانحراف المعياري. وكذلك تم حذف سمة القطر الأصغري، لأنه يعيد أصغر قطر للدائرة التي لها مساحة المنطقة نفسها. كما أن سمة طول المحور الأعظمي والأصغري تعطيان تقديراً أكثر شمولياً منه، وتم حذف سمة الانضغاط لأنه يحسب من قانون المحيط والمساحة، وبالتالي بإمكاننا الاستعاضة عنه باعتماد السمات الأكثر أهمية منه.

كما نلاحظ من الجدولين السابقين أن كلاً من سمة الطاقة والتجانس والترابط والشذوذية، أظهرت قيماً متماثلة بالنسبة إلى نوعي الأورام، وبالتالي تعد غير قادرة على التمييز بينها، لذلك يتم حذفها والاستغناء عنها.

وبحذف السمات السابقة، يكون قد تم تقليل أبعاد متجه السمات المستخدم من 14 سمة إلى 6 سمات (المساحة، المحيط، المحور الأعظمي، المحور الأصغري، الاستدارة، التباين). وبذلك تم ضبط متجه السمات ليتوافق مع العمل الأمثل لمصنف متجه دعم الآلة SVM، كونه يظهر دقة منخفضة إذا كانت أبعاد متجه الدخل كبيرة. وتكون السمات النهائية التي سيتم استخدامها كدخل للتصنيف:



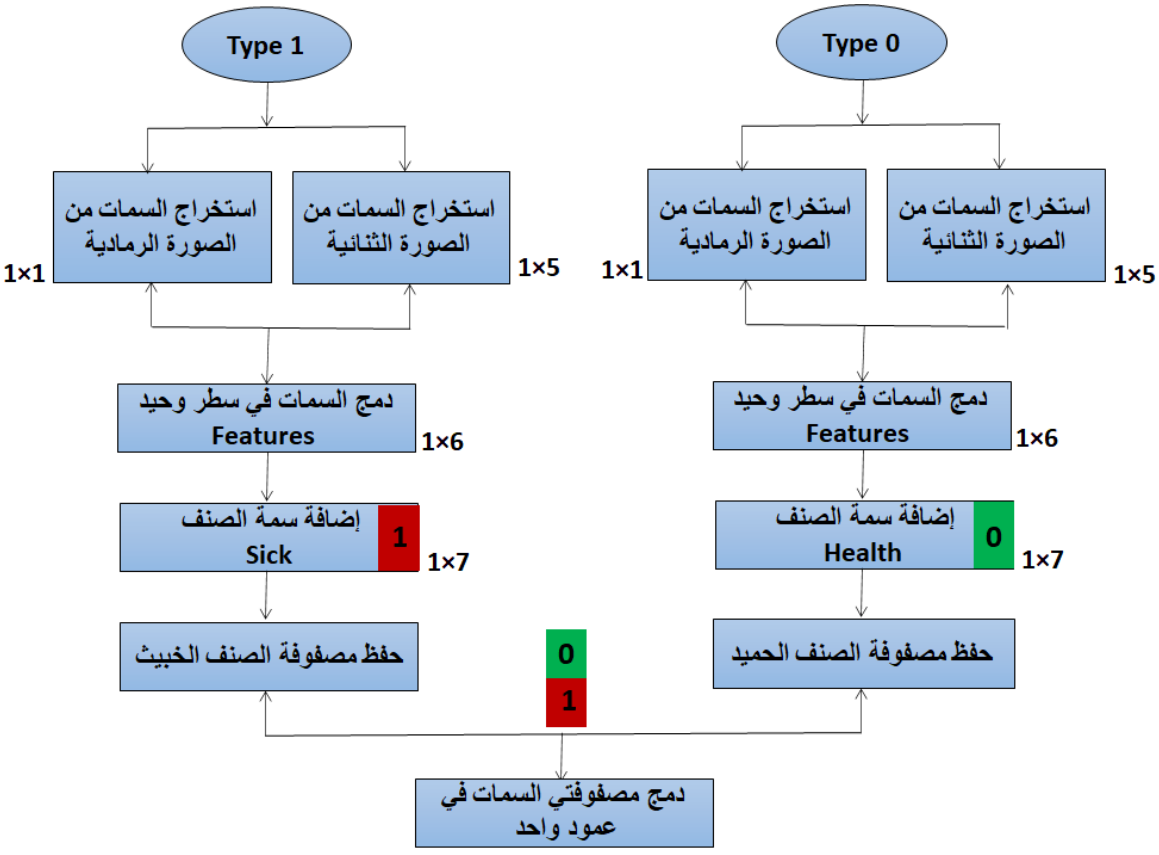
الشكل (٤-١٩) السمات المستخدمة في تصنيف الأورام المكتشفة.

٤-٦-مرحلة التصنيف Classification:

التصنيف هو الخطوة الأخيرة لتحديد ما إذا كانت العقيدات سرطانية أم لا، ويتم تحقيق هذه الخطوة باستخدام إحدى خوارزميات الذكاء الاصطناعي. ومن أهمها متجه دعم الآلة SVM الذي يعد خوارزمية تصنيف قوية جداً ذات كلفة حسابية أقل. ويعمل بشكل ممتاز مع هامش فصل واضح، كما يفضل استخدامه عندما يكون هناك صنفان فقط من البيانات، كما في مثالنا [64، 65].

٤-٦-١-خطوات التصنيف باستخدام مصنف متجه دعم الآلة SVM:

بعد تحديد السمات المثلى التي تعد أكثر أهمية لتصنيف الأورام الحميدة والخبيثة، يتم دمج كل من هذه السمات في سطر وحيد، وإضافة سمة الصنف تجهيزاً لتدريب المصنف واختباره. ويبين الشكل الآتي خطوات استخراج السمات الخاصة بكل صنف:



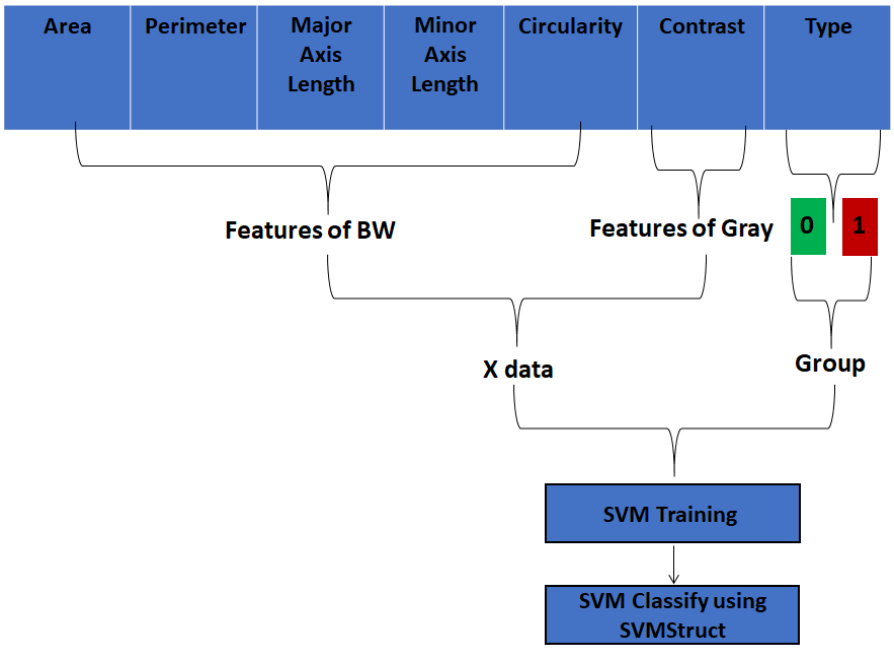
الشكل (٤-٢٠) خطوات استخراج السمات المثلى تمهيداً لتدريب المصنف.

١-تدريب المصنف SVM Training:

يتم دمج المصفوفتين كأعمدة وتغذيتها إلى المصنف، بحيث يتوافق كل سطر مع كائن Candid، وكل عمود يتوافق مع سمة Feature، ومن ثم تدريب الشبكة حتى تصبح قادرة على التصنيف. وعند انتهاء التدريب نحصل على بنية المصنف SVM Struct، التي تحتوي على جميع معلومات المصنف المدرب.

٢-التصنيف SVM Classify:

من أجل كل دخل جديد تتم مقارنة متجه السمات الخاصة به مع بنية المصنف، وبذلك يجري تصنيفه إلى الصنف الموافق [66، 67].



الشكل (٤-٢١) خطوات التصنيف باستخدام خوارزمية SVM.

٤-٧-خلاصة:

قدمنا في هذا الفصل شرحاً تفصيلياً وبالرسوم التوضيحية عن خطوات ومراحل النظام المقترح، وتطبيقه على قاعدة البيانات المستخدمة، وسنتحدث في الفصل القادم عن التقييم العملي لأداء النظام، ومناقشة نتائج العمل. كما سنقوم بذكر أهم الآفاق المستقبلية والمقترحات التي من الممكن العمل على إنجازها في أبحاث قادمة كتطوير لفكرة هذا البحث.

الفصل الخامس

تقييم الأداء

Performance Evaluation

٥-١-مقدمة:

سنناقش في هذا الفصل نموذج اختبار النظام المطور، وتقييم أداء خوارزميتي التجزئة والتصنيف من خلال حساب بارامترات الدقة والاستدعاء والإحكام، ومقارنة هذه النتائج بالدراسات السابقة.

٥-٢-نموذج تقييم الأداء:

من أجل تقييم النموذج المقترح، وتحديد دقة الكشف، تم استخدام مصفوفة الارتباك Confusion Matrix التي تعد واحدة من أكثر بارامترات تقييم الأداء دقة [68]، وتتكون مصفوفة الارتباك من البارامترات الآتية:

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

الشكل (٥-١) بارامترات مصفوفة الارتباك [68].

١- **TP–True Positive**: أي عدد العينات التي نجح فيها النموذج في التعرف على العينات الصحيحة، وأشار إليها أنها عينات صحيحة.

٢- **TN – True Negative**: أي عدد العينات التي نجح فيها النموذج في التعرف على العينات الخاطئة، وأشار إليها أنها عينات خاطئة.

٣- **FP – False Positive**: أي عدد العينات التي فشل فيها النموذج في التعرف على العينات الصحيحة، وأشار إليها أنها عينات خاطئة.

٤- **FN – False Negative**: أي عدد العينات التي فشل فيها النموذج في التعرف على العينات الخاطئة، وأشار إليها أنها عينات صحيحة.

لذلك فإن النموذج الفعال سوف يجعل قيمة TP وTN أكبر ما يمكن، بالإضافة إلى جعل قيمة FP وFN أصغر ما يمكن.

ومن خلال مصفوفة الارتباك يمكن الوصول إلى أربعة معايير تقيم أداء النموذج، هي:

١- الدقة Accuracy:

تعبر عن نسبة نجاح النموذج في التعرف على العينات الصحيحة والخاطئة، أي هو نسبة عدد التوقعات الصحيحة على العدد الكلي للتوقعات، ويمكن حساب هذه القيمة من خلال المعادلة:

$$Accuacy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (١-٥)$$

٢- الاستدعاء Recall أو الحساسية Sensitivity:

تعبر عن نسبة الحالات الصحيحة، ونجح النموذج في التعرف بأنها صحيحة على عدد الحالات الكلي للحالات الصحيحة وتعرف عليها النموذج بأنها صحيحة، بالإضافة إلى عدد الحالات الخاطئة التي تعرف عليها النموذج بأنها خاطئة. ويمكن حساب هذه القيمة من خلال المعادلة:

$$Recall = \frac{TP}{TP+FN} \qquad (٢-٥)$$

٣-الإحكام Precision:

تعبر عن نسبة الحالات الصحيحة، ونجح النموذج في التعرف بأنها صحيحة على عدد الحالات الكلي للحالات الصحيحة، وتعرف النموذج عليها بأنها صحيحة، بالإضافة إلى عدد الحالات الصحيحة، وفشل النموذج في التعرف عليها، أي أنها تشير إلى عدد الحالات التي نجح فيها النموذج في التعرف على العينات الإيجابية (وهي بالفعل إيجابية)، ويستخدم Precision لتقليل الإيجابيات الكاذبة FP. ويمكن حساب هذه القيمة من خلال المعادلة الآتية:

$$Precision = \frac{TP}{TP+FP} \qquad (٣-٥)$$

٤-التوافقية بين الدقة والاستدعاء $F1-Score$:

يمكن حساب هذه القيمة من المعادلة الآتية:

$$F1 = 2 * \frac{precision*recall}{precision+recall} \qquad (٤-٥)$$

وسوف يتم تطبيق هذه المعايير لتقييم أداء كل من خوارزمية التجزئة والتصنيف.

٥-٢-١-تقييم أداء خوارزمية التجزئة:

تنتج خوارزميات تجزئة الصور مجموعة من الأقنعة التي تصف كل كائن (ورم) تم اكتشافه في الصورة، ويجب أن يتطابق قناع المنطقة الورمية المجزأة مع الدقة المكانية لصورة الدخل (الارتفاع والعرض)، ومع عدد المناطق (الكائنات) المحتملة التي يمكن اكتشافها [69].

يتم تقييم خوارزمية تجزئة الصورة أولاً على مستوى الصورة، ثم يتم التحقق من صحتها على مستوى قاعدة البيانات جميعها، أي حساب المتوسط الكلي لدقة جمع الصور.

٥-٢-١-١-تقييم أداء خوارزمية التجزئة على مستوى الصورة:

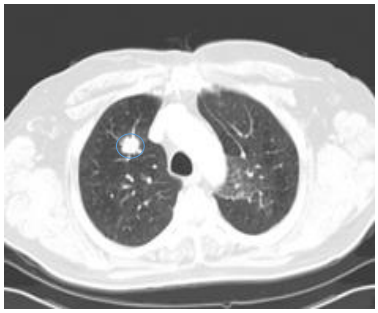
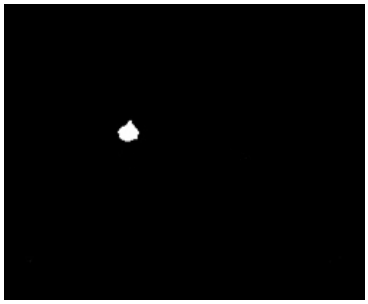
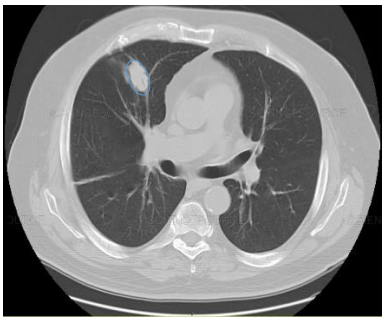
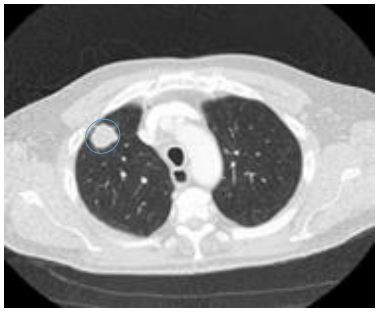
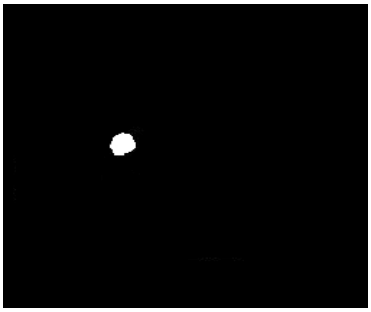
لمعرفة صحة الخوارزمية يجب الاستعانة بما يعرف بالحقيقة المرجعية Ground Truth (صور مجزأة يدوياً من قبل الطبيب)، وذلك للمقارنة بين ناتج عملية التجزئة والحقيقة المرجعية لكل صورة.

إذ يتم تمثيل بارامترات مصفوفة الارتباك على مستوى الصورة على الشكل الآتي:

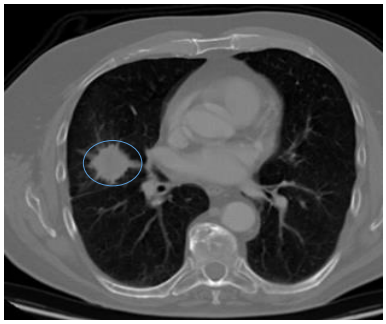
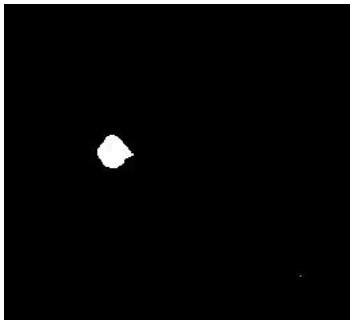
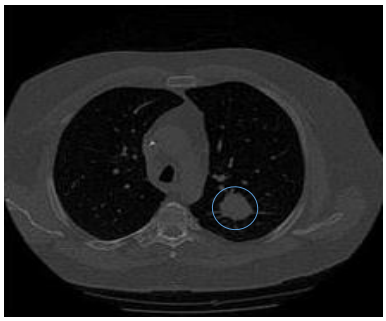
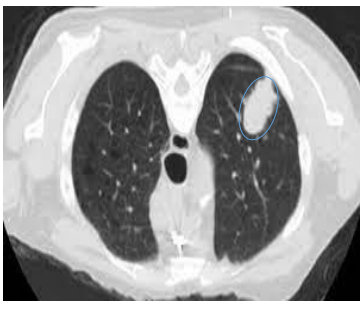
١. **TP** عدد عناصر الصورة (البكسلات) التي تنتمي للورم وتم تصنيفها كذلك.
٢. **TN:** عدد عناصر الصورة التي لا تنتمي للورم وتم تصنيفها كذلك.
٣. **FP:** عدد عناصر الصورة التي تم تصنيفها على أنها ورم وهي ليست ورماً.
٤. **FN:** عدد عناصر الصورة التي تم تصنيفها على أنها سليمة وهي في الحقيقة تنتمي إلى الورم.

وتوضح الجداول الآتية نتيجة تطبيق خوارزمية التجزئة على عدة صور من قاعدة البيانات.

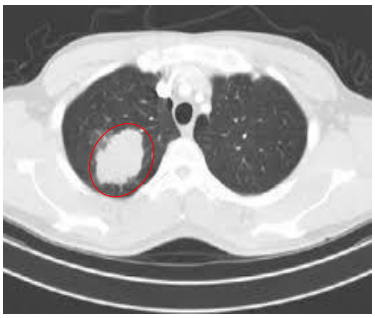
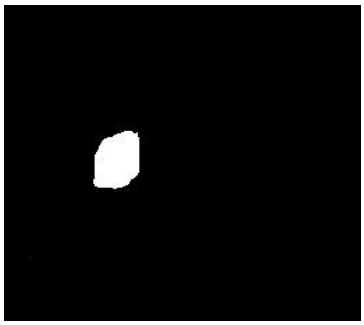
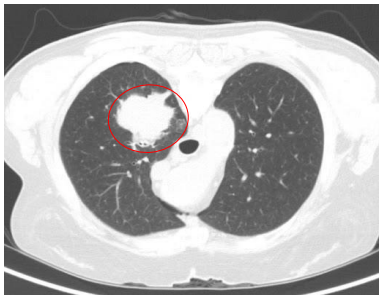
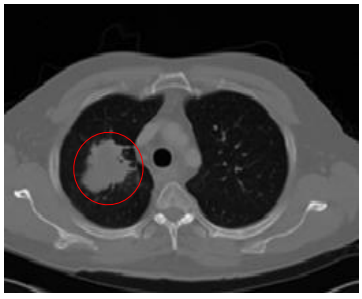
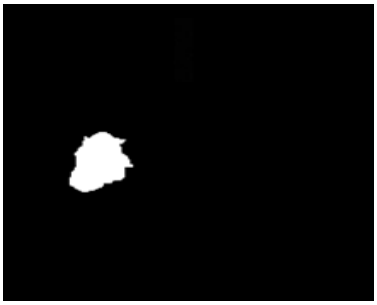
الجدول (٥-١) صور المناطق المجزأة ذات الورم الحميد:

الرقم	الصورة المقطعية المحوسبة CT	المنطقة الورمية المكتشفة
1		
2		
3		
4		

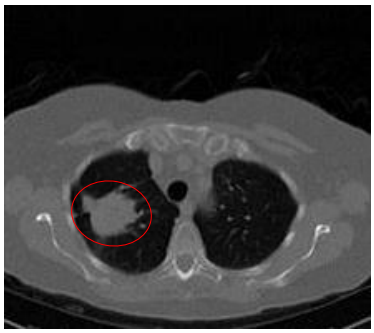
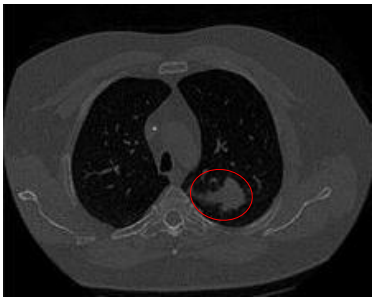
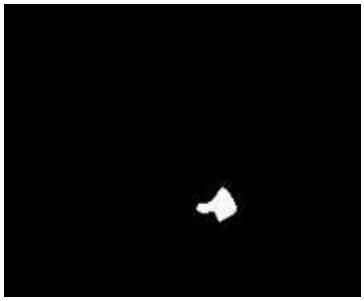
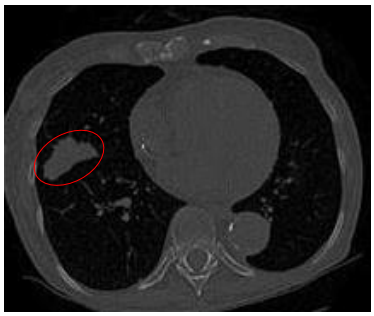
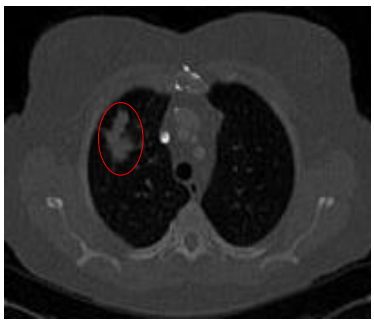
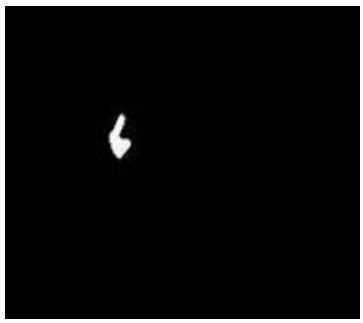
تابع الجدول (٥-١) صور المناطق المجزأة ذات الورم الحميد:

الرقم	الصورة المقطعية المحوسبة CT	المنطقة الورمية المكتشفة
5		
6		
7		
8		

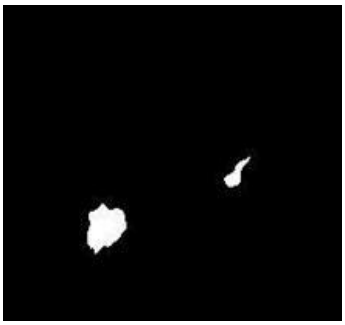
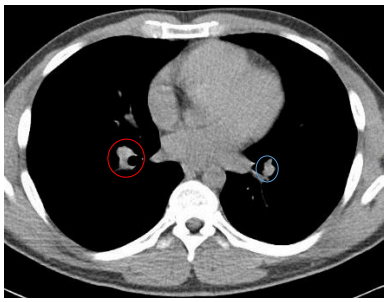
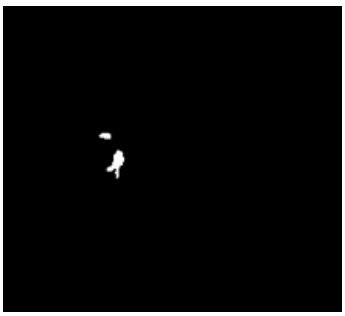
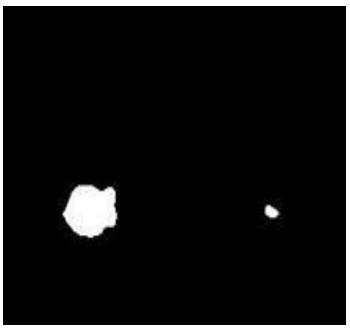
الجدول (٥-٢) صور المناطق المجزأة ذات الورم الخبيث:

الرقم	الصورة المقطعية المحوسبة CT	المنطقة الورمية المكتشفة
9		
10		
11		
12		

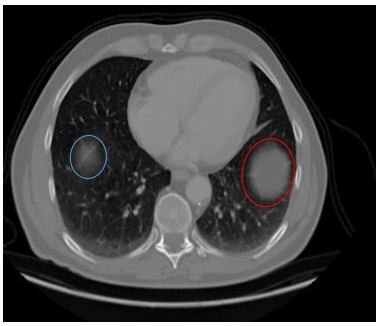
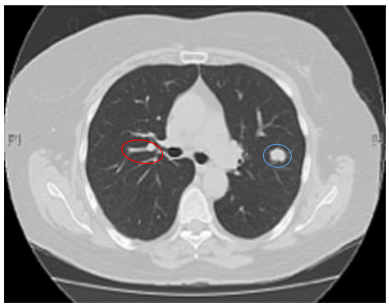
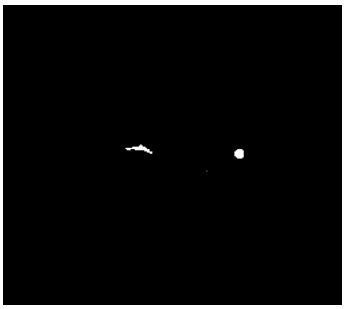
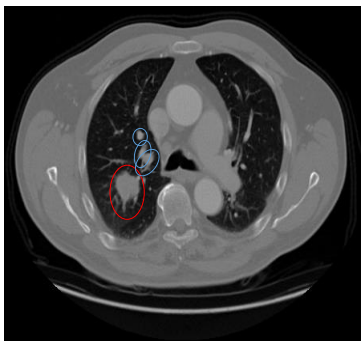
تابع الجدول (٥-٢) صور المناطق المجزأة ذات الورم الخبيث:

الرقم	الصورة المقطعية المحوسبة CT	المنطقة الورمية المكتشفة
13		
14		
15		
16		

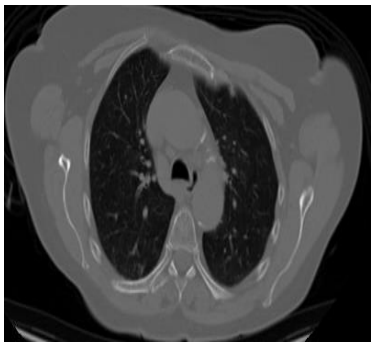
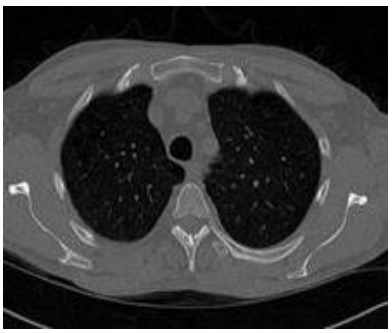
الجدول (٥-٣) صور المناطق المجزأة ذات الأورام المتعددة:

الرقم	الصورة المقطعية المحوسبة CT	المنطقة الورمية المكتشفة
17		
18		
19		
20		

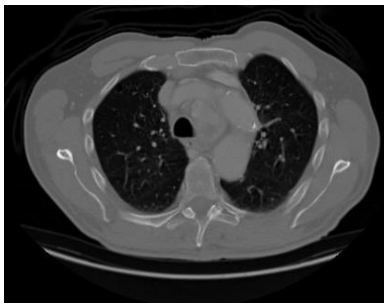
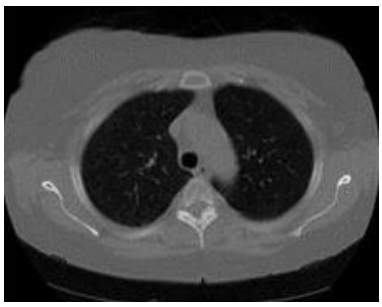
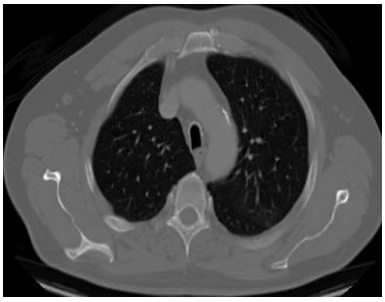
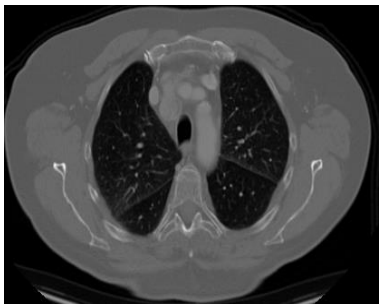
تابع الجدول (٥-٣) صور المناطق المجزأة ذات الأورام المتعددة:

الرقم	الصورة المقطعية المحوسبة CT	المنطقة الورمية المكتشفة
21		
22		
23		
24		

الجدول (٥-٤) صور المناطق السليمة بدون أورام:

الرقم	الصورة المقطعية المحوسبة CT	المنطقة الورمية المكتشفة
25		
26		
27		
28		

تابع الجدول (٥-٤) صور المناطق السليمة بدون أورام:

الرقم	الصورة المقطعية المحوسبة CT	المنطقة الورمية المكتشفة
29		
30		
31		
32		

وتوضح الجداول الآتية قيم عناصر مصفوفة الارتباك لكل صورة ورمية/سليمة مجزأة، مع قيم معايير تقييم نموذج التجزئة، ومن خلال هذا الجدول، نلاحظ نجاح النموذج المقترح في استخراج معظم الصور، لأن قيم TP أكبر ما يمكن، وقيم FP & FN أصغر ما يمكن.

إذ يمكننا حساب قيم معايير التقييم على الشكل الآتي:

١-الدقة Accuracy:

يمكن أن يوفر هذا المعيار في حالة تقييم نموذج تجزئة الصورة نتائج غير صحيحة عندما تكون الأصناف غير متوازنة، أي عندما يكون تمثيل صنف ما صغيراً في الصورة (المنطقة الورمية **TP:** ١٠٠-٢٢٠٠ بكسل)، بينما يشكل الصنف الآخر (الخلفية السوداء/عدم وجود ورم **TN:** ٦٥٥٣٦ بكسل) الجزء الأكبر من الصورة. وهنا سنحصل على دقة تجزئة فائقة مهما كان حجم وشكل الورم الناتج غير مشابه للدخل، لذلك سيتم اللجوء إلى المعايير الأخرى في التعامل مع هذه المشكلة [70].

٢-الاستدعاء Recall:

يمثل عدد العناصر الورمية المجزأة بشكل صحيح في الصورة.

٣-الإحكام Precision:

يمثل عدد العناصر الورمية المجزأة من الصورة والتي تتطابق مع الحقيقة المرجعية.

٤-التوافقية بين الاستدعاء والإحكام $F1$ (معامل النرد Dice Coefficient):

وهو المعيار الأهم لتقييم خوارزميات التجزئة، إذ يقوم بحساب مقياس التشابه Similarity Measure بين **الخرج** (الصورة المجزأة) **والهدف** (الحقيقة المرجعية)، أي يقيس عدد البكسلات المشتركة بين قناع الخرج وقناع الدخل، مقسوماً على العدد الإجمالي للبكسلات الموجودة في القناعين [69، 70].

الجدول (٥-٥) قيم عناصر مصفوفة الارتباك لصور المناطق المجزأة ذات الورم الحميد، مع قيم معايير تقييم نموذج التجزئة:

رقم الصورة	نوع الورم	مساحة الورم TP	FP	FN	الاستدعاء Recall	الإحكام Precision	التوافقية F1(Dice)
1	حميد	471	34	11	0.977	0.932	0.953
2	حميد	202	18	12	0.943	0.918	0.929
3	حميد	346	41	0	1	0.894	0.944
4	حميد	256	7	0	1	0.973	0.986
5	حميد	587	26	0	1	0.957	0.978
6	حميد	504	0	32	0.940	1	0.969
7	حميد	289	0	16	0.947	1	0.972
8	حميد	686	–	–	–	–	0

الجدول (٥-٦) قيم عناصر مصفوفة الارتباك لصور المناطق المجزأة ذات الورم الخبيث، مع قيم معايير تقييم نموذج التجزئة:

رقم الصورة	نوع الورم	مساحة الورم TP	FP	FN	الاستدعاء Recall	الإحكام Precision	التوافقية F1(Dice)
9	خبيث	1260	45	13	0.989	0.965	0.976
10	خبيث	1394	0	49	0.966	1	0.982
11	خبيث	2231	76	0	1	0.967	0.983
12	خبيث	1443	61	32	0.978	0.959	0.967
13	خبيث	795	19	27	0.967	0.976	0.970
14	خبيث	513	14	38	0.971	0.973	0.950
15	خبيث	610	0	26	0.959	1	0.979
16	خبيث	367	14	18	0.953	0.963	0.957

الجدول (٥-٧) قيم عناصر مصفوفة الارتباك لصور المناطق المجزأة ذات الأورام المتعددة، مع قيم معايير تقييم نموذج التجزئة:

رقم الصورة	نوع الورم	مساحة الورم TP	FP	FN	الاستدعاء Recall	الإحكام Precision	التوافقية F1(Dice)
17	خبيث	734	17	14	0.981	0.977	0.978
	خبيث	254	0	56	0.819	1	0.900
18	خبيث	162	0	29	0.848	1	0.917
	حميد	87	0	43	0.669	1	0.801
19	خبيث	226	8	16	0.933	0.965	0.948
	حميد	84	0	17	0.831	1	0.907
20	خبيث	1127	13	10	0.991	0.988	0.989
	حميد	119	0	32	0.788	1	0.881
21	خبيث	920	33	0	1	0.965	0.982
	حميد	232	0	41	0.849	1	0.918
22	خبيث	128	0	20	0.864	1	0.927
	حميد	105	0	22	0.826	1	0.904
23	حميد	285	6	8	0.972	0.979	0.974
	خبيث	133	0	4	0.970	1	0.984
24	خبيث	425	21	15	0.965	0.952	0.957
	حميد	98	10	0	1	0.907	0.951
	حميد	154	–	–	–	–	0
	حميد	103	–	–	–	–	0

٥-٢-١-٢-تقييم أداء خوارزمية التجزئة على مستوى قاعدة البيانات:

تتراوح قيمة معامل النرد بين (0–100%) 0–1، حيث يشير 0 إلى عدم وجود تداخل بين قناع الصورة المجزأة وقناع الدخل المطلوب، ويدل الرقم 1 على التجزئة المتداخلة تماماً.
ونطبق الطريقة المقترحة على جميع صور قاعدة البيانات، إذ يتم اعتبار التجزئة على أنها جيدة مع قيمة عتبة Dice > 0.8 [71].

ويصبح تمثيل بارامترات مصفوفة الارتباك على مستوى قاعدة البيانات على الشكل الآتي:

١. **TP:** عدد الصور التي تملك قيمة معامل النرد أكبر من العتبة المحددة 0.8.
٢. **FP:** عدد الصور التي تملك قيمة معامل النرد أصغر من العتبة المحددة 0.8.
٣. **FN:** عدد الصور التي تملك قيمة معامل النرد معدوماً، أي أن الخوارزمية لم تكتشف الورم الموجود في الصورة.

ويوضح الجدول الآتي قيم عناصر مصفوفة الارتباك لجميع صور قاعدة البيانات بعد حساب قيمة معامل النرد، ومن خلال هذا الجدول، نلاحظ نجاح النموذج المقترح في استخراج معظم الصور، لأن قيم TP أكبر من قيم FN & FP.

الجدول (٥–٨) قيم عناصر مصفوفة الارتباك لقاعدة البيانات جميعها:

	FN	FP	TP	العدد الكلي
قاعدة البيانات	**8**	**2**	**127**	**137**

ويوضح الجدول الآتي قيم معايير تقييم نموذج التجزئة. ومن خلال هذا الجدول، نلاحظ نجاح النموذج المقترح في استخراج معظم الصور، إذ أن قيم الاستدعاء والإحكام والتوافقية قريبة من الواحد.

الجدول (٥–٩) قيم محددات تقييم أداء خوارزمية التجزئة:

Parameter	Value
Recall %	94.07%
Precision %	98.44%
$\boldsymbol{F1-Score}$ %	96.20%

نجد من النتائج السابقة أن النظام أظهر فاعلية في استخراج الأورام على اختلاف الحالات المقدمة له، فقد تمكن من عزل المناطق الورمية في الصور التي تتضمن أنسجة وحويصلات رئوية على الرغم من أن لها السويات الرمادية نفسها للمناطق الورمية كما في الصور (3، 6، 17، 20، 22). كما نجح في استخراج الأورام وتمييزها في الصور التي كانت فيها الأورام قريبة من الجدران والأنسجة الداخلية للرئة كما في الصور (4، 7، 9، 13).

ونجح النظام في اكتشاف الأورام في الصور غير الواضحة التي تحوي شوائب عالية كما في الصور (1، 2، 5، 23)، وفي الصور التي تحوي أوراماً غير واضحة ومبهمة الحدود كما في الصور (11، 14، 16، 21). وقد نجح النظام في عدم استخراج المناطق المشبوهة التي لا تعد تكتلات ورمية كما في الصورة (19) إذ نلحظ وجود كتلة في الرئة اليمنى ولم يتم عزلها واعتبارها كتلة ورمية. كما نجح أيضاً في عدم استخراج أي منطقة في الصور السليمة التي لا تحتوي على مناطق ورمية كما في الصور (25، 26، 27، 28، 29، 30، 31، 32)، وقد أظهر النظام قدرة على استخراج حدود المناطق الورمية وحوافها بدقة جيدة كما في الصور (10، 12، 15، 18).

كما نلحظ من الجداول السابقة وجود تجزئة خاطئة في الصورة رقم (8)، إذ لم يتم التعرف على الورم الموجود في الصورة، وذلك يعود لالتصاق الورم بشكل كبير بجدران الرئة اليمنى، مما أدى إلى صعوبة عزله واعتباره من أنسجة الرئة. وكذلك في الصورة رقم (24) التي تم فيها دمج ورمين مع بعضهما عن طريق الخطأ، بسبب صغر المسافة بينهما وامتلاكهما قيم سويات رمادية متقاربة، مما أدى إلى حدوث مشكلة التداخل.

٥-٢-٢-تقييم أداء خوارزمية التصنيف:

بعد الانتهاء من تقييم مرحلة التجزئة، سوف يتم حذف الصور المجزأة بشكل خاطئ (10 صور)، وحذف الصور الطبيعية (السليمة) كونها لا تحوي أوراماً (22 صورة)، وبالتالي لن تدخل في عملية التصنيف.

أي سيتم الاعتماد على الصور الورمية المجزأة بشكل صحيح من قاعدة البيانات، والتي بلغ عددها 105 صورة، من أجل بناء الشبكة وتدريبها واختبارها، وقد وزعت الصور النهائية على الشكل الآتي:

مجوعة بيانات التدريب:

81 صورة لتدريب الشبكة العصبية عليها، فقد تم اختيار عينات تحتوي على عدد كاف من نوعي الأورام لضمان تدريب الشبكة بشكل جيد.

وبما أن هناك عدة صور تحوي أكثر من ورم في الصورة نفسها، فإن عدد الكائنات المرشحة Candid التي تمثل الأورام سوف يزيد عن 81 ورماً، وسيتم التعامل مع كل ورم (كائن مرشح) على أنه صورة بذاتها. وهذا يؤدي إلى زيادة عدد بيانات التدريب الموزعة على الشكل الآتي:

30 صورة ذات أورام حميدة، 35 صورة ذات أورام خبيثة، 16 صورة ذات أورام متعددة.

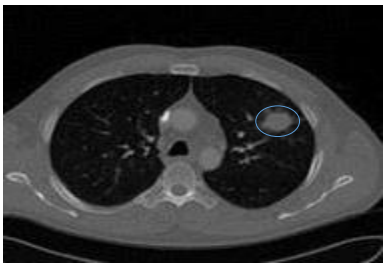
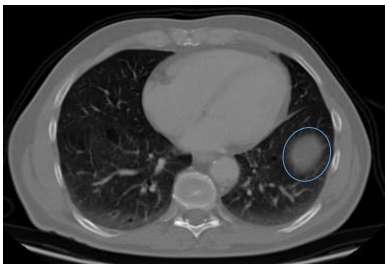
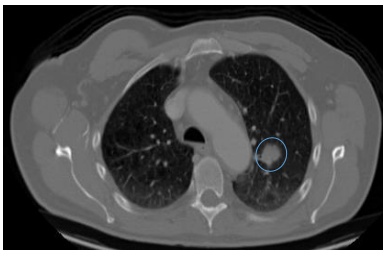
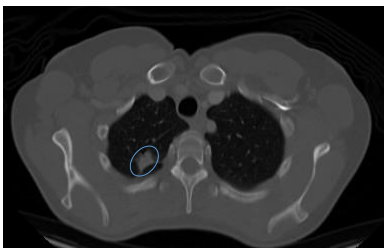
مجموعة بيانات الاختبار:

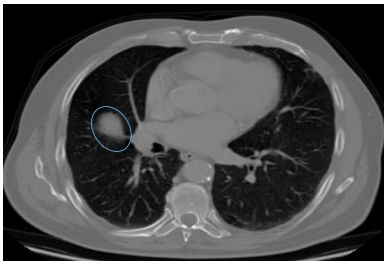
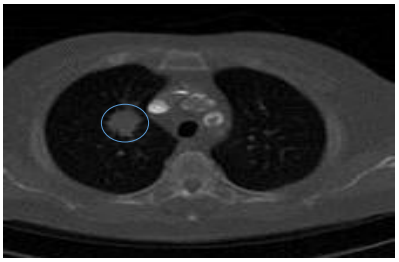
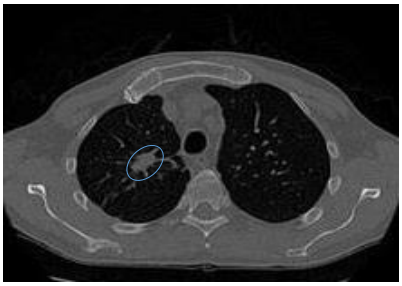
24 عينة (تحوي 29 ورماً) غير مدرب عليها الشبكة بغاية تقييم الأداء، موزعة على الشكل الآتي:

10 صور ذات أورام حميدة، 9 صور ذات أورام خبيثة، 5 صور ذات أورام متعددة تتضمن (8 أورام خبيثة، 2 أورام حميدة).

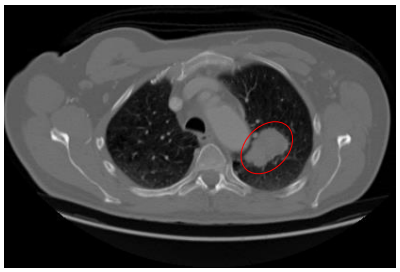
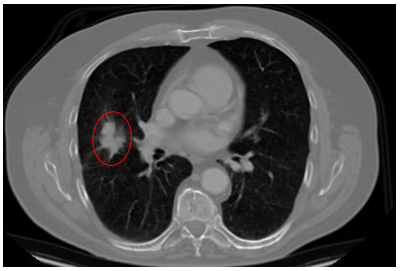
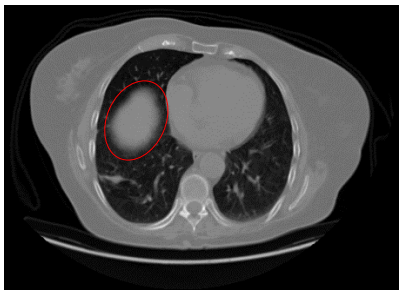
ويوضح الجدول التالي نتائج عملية التصنيف على مجموعة بيانات الاختبار، مع توضيح التشخيص الآلي والطبي لكل منها.

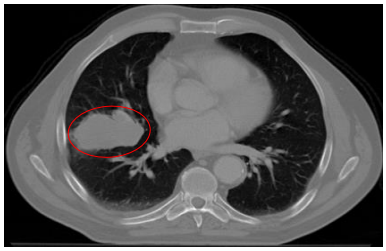
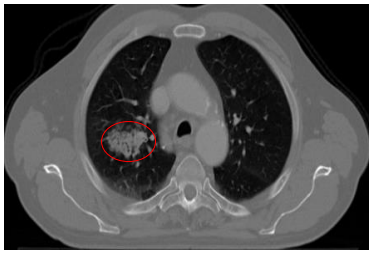
الجدول (٥-١٠) صور مجموعة بيانات الاختبار ذات الورم الحميد، مع التشخيص الطبي والآلي:

الرقم	الصورة	التشخيص الطبي	التشخيص الآلي
1		ورم حميد في الجهة اليمنى **TN**	ورم حميد في الجهة اليمنى **TN**
2		ورم حميد في الجهة اليمنى **TN**	ورم حميد في الجهة اليمنى **TN**
3		ورم حميد في الجهة اليمنى **TN**	ورم حميد في الجهة اليمنى **TN**
4		ورم حميد في الجهة اليمنى **TN**	ورم حميد في الجهة اليمنى **TN**
5		ورم حميد في الجهة اليسرى **TN**	ورم حميد في الجهة اليسرى **TN**

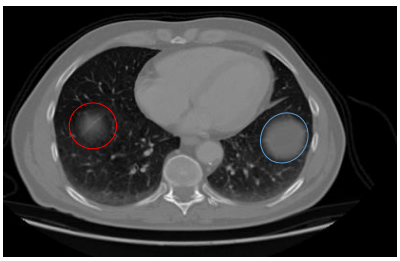
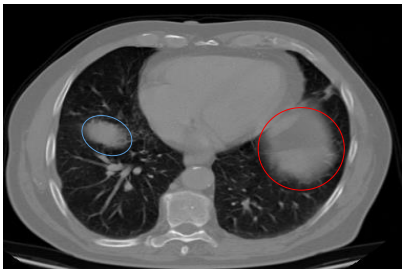
6		ورم حميد في الجهة اليسرى **TN**	ورم حميد في الجهة اليسرى **TN**
7		ورم حميد في الجهة اليسرى **TN**	ورم حميد في الجهة اليسرى **TN**
8		ورم حميد في الجهة اليسرى **TN**	ورم حميد في الجهة اليسرى **TN**
9		ورم حميد في الجهة اليسرى **TN**	ورم حميد في الجهة اليسرى **TN**
10		ورم حميد في الجهة اليسرى **TN**	ورم حميد في الجهة اليسرى **TN**

الجدول (٥-١١) صور مجموعة بيانات الاختبار ذات الورم الخبيث، مع التشخيص الطبي والآلي:

الرقم	الصورة	التشخيص الطبي	التشخيص الآلي
1		ورم خبيث في الجهة اليمنى **TP**	ورم خبيث في الجهة اليمنى **TP**
2		ورم خبيث في الجهة اليسرى **TP**	ورم حميد في الجهة اليسرى **FP**
3		ورم خبيث في الجهة اليسرى **TP**	ورم خبيث في الجهة اليسرى **TP**
4		ورم خبيث في الجهة اليسرى **TP**	ورم خبيث في الجهة اليسرى **TP**

5		ورم خبيث في الجهة اليسرى **TP**	ورم خبيث في الجهة اليسرى **TP**
6		ورم خبيث في الجهة اليسرى **TP**	ورم خبيث في الجهة اليسرى **TP**
7		ورم خبيث في الجهة اليسرى **TP**	ورم خبيث في الجهة اليسرى **TP**
8		ورم خبيث في الجهة اليسرى **TP**	ورم خبيث في الجهة اليسرى **TP**
9		ورم خبيث في الجهة اليسرى **TP**	ورم حميد في الجهة اليسرى **FP**

الجدول (٥-١٢) صور مجموعة بيانات الاختبار ذات الأورام المتعددة، مع التشخيص الطبي والآلي:

الرقم	الصورة	التشخيص الطبي	التشخيص الآلي
1		ورم خبيث في الجهة اليمنى **TP** ورم خبيث في الجهة اليسرى **TP**	ورم خبيث في الجهة اليمنى **TP** ورم خبيث في الجهة اليسرى **TP**
2		ورم خبيث في الجهة اليمنى **TP** ورم حميد في الجهة اليسرى **TN**	ورم خبيث في الجهة اليمنى **TP** ورم حميد في الجهة اليسرى **TN**
3		ورم خبيث في الجهة اليمنى **TP** ورم خبيث في الجهة اليسرى **TP**	ورم خبيث في الجهة اليمنى **TP** ورم خبيث في الجهة اليسرى **TP**
4		ورم خبيث في الجهة اليمنى **TP** ورم خبيث في الجهة اليسرى **TP**	ورم خبيث في الجهة اليمنى **TP** ورم خبيث في الجهة اليسرى **TP**
5		ورم خبيث في الجهة اليمنى **TP** ورم حميد في الجهة اليسرى **TN**	ورم خبيث في الجهة اليمنى **TP** ورم حميد في الجهة اليسرى **TN**

يوضح الجدول الآتي (٥-١٣) قيم عناصر مصفوفة الارتباك لكل صنف من الأصناف (الأورام) في قاعدة البيانات، ومن خلال هذا الجدول نلاحظ نجاح النموذج المقترح في تصنيف معظم الأورام، حيث إن قيم TP & TN أكبر من قيم FP & FN:

وتبين النتائج التجريبية تطبيق نظام اكتشاف وتصنيف أورام الرئة المقترح بعد الاعتماد على مجموعة بيانات الاختبار المؤلفة من 24 حالة عشوائية، (تحوي 29 ورماً)، لم يتم تدريب الشبكة عليها ومقارنتها مع الخرج المطلوب بغاية تقييم الأداء.

الجدول (٥-١٣) قيم عناصر مصفوفة الارتباك لصنفي الأورام:

قاعدة البيانات	FN	FP	TN	TP	العدد الكلي
أورام خبيثة	–	**2**	–	**15**	**17**
أورام حميدة	**0**	–	**12**	–	**12**
العدد الكلي	**0**	**2**	**12**	**15**	**29**

ويتم تمثيل البارامترات في هذه الحالة على الشكل الآتي:

١. **TP:** أي عدد الأورام الخبيثة التي تم تصنيفها على أنها خبيثة.
٢. **TN:** أي عدد الأورام الحميدة التي تم تصنيفها على أنها حميدة.
٣. **FP:** أي عدد الأورام الخبيثة التي تم تصنيفها على أنها حميدة.
٤. **FN:** أي عدد الصور الحميدة التي تم تصنيفها على أنها خبيثة.

ونستنتج أيضاً من قيم مصفوفة الارتباك أن النموذج ناجح في التعرف على الأورام وتصنيفها، لأن قيم معايير الدقة، والاستدعاء، والإحكام، والتوافقية تساوي أو قريبة من الواحد.

الجدول (٥-١٤) قيم محددات تقييم أداء خوارزمية التصنيف:

Parameter	Value
Accuracy %	93.1%
Recall %	100%
Precision %	88.23%
$\boldsymbol{F1-Score}$ %	93.74%

ويبين الجدول الآتي مقارنة النظام المقترح مع الأبحاث السابقة.

الجدول (٥-١٥) مقارنة بين الأبحاث المنشورة بين عامي 2016 و2020 مع النظام المقترح:

البحث	المعالجة المسبقة	طريقة التجزئة	السمات المستخرجة	عدد السمات	المصنف	الدقة %
[1] 2020	×	×	Variance - Life Time	2	RF KNN	82.98
[2] 2019	×	×	Contrast - Correlation - Variance - ASM - Inverse Difference Moment - Entropy	7	NN KNN	89.7
[3] 2019	Median Filter	K – Means	Volume - Area - Standard Deviation - Circularity - Compactness - Major & Minor Axis Length – Eccentricity	8	DT	95
[4] 2019	Laplacian Filter	×	-	-	CNN	82.43
[5] 2019	Adaptive Histogram Equalization	×	Variance - Mean - Standard Deviation - Kurtosis - Skewness - Energy - Entropy - Contrast - Homogeneity – Correlation	10	NN	94.56
[6] 2018	Gaussian Filter	Region Growing	Variance - Lobulation - Compactness - Sphericity - Energy – Entropy	6	RF	84.62
[7] 2018	Median - Gaussian - Bilateral - Filter	Global Thresholding	Contrast - Dissimilarity - Correlation - ASM Entropy - Homogeneity - Skewness - Standard Deviation - Entropy - Rout Mean Square	10	MLP	88.55
[8] 2018	Median - Gaussian Filter	Watershed	Area - Perimeter - Eccentricity - Centroid - Diameter - Pixel Mean Intensity	6	SVM	92
[9] 2017	×	×	-	-	Inception v3	85
[10] 2017	×	Intensity Thresholding	-	-	3D CNN	67.7
[11] 2017	×	Histogram Thresholding	-	-	NN	77.8
[12] 2016	×	Intensity Thresholding	Area - Perimeter - Equivalent Diameter - Major & Minor Axis Length - Circularity - Roughness - Sphericity - Spiculation - Lobulation - Volume - Entropy - Energy - Contrast - Inverse Difference Moment	15	SVM	86.36
[13] 2016	×	Rule Based Thresholding	Mean - Variance - Energy - Entropy - Kurtosis - Skewness - Contrast - Smoothness - Correlation - Eccentricity - Circularity - Fractal Dimension	12	SVM	78.08
النظام المقترح 2022	Wavelet Transform	Fuzzy C-Mean & Otsu Thresholding	Area - Perimeter - Major Axis Length - Contrast-Minor Axis Length - Circularity	6	SVM	93.1

٥-٣-النتائج والمناقشة:

تم اقتراح نظام جديد لتجزئة صور العقيدات الرئوية وتصنيفها، وقد جرى تطبيق النظام واختباره على مجموعة من الصور المقطعية المحوسبة، إذ تم في البداية إزالة الضجيج من هذه الصور، ثم استخراج المناطق المشبوهة من خلال استخدام خوارزمية تجزئة هجينة بدءاً بعنقدة عناصر الصورة حسب السويات الرمادية، وذلك باستخدام خوارزمية FCM، بالإضافة إلى تطبيق عتبة Otsu، وقد تم اعتماد هذه العتبة بسبب الاختلاف في كثافة المستويات الرمادية للمناطق الورمية التي تتطلب عتبات مختلفة للصور المختلفة. ثم تطبيق بعض العمليات المورفولوجية كعمليتي التمدد والتعرية لإزالة المناطق غير المرغوبة، وبعدها الحصول على الورم المجزأ من خلال عملية طرح الصورتين الناتجتين من الخطوتين السابقتين. وقد تم التوصل إلى نتائج جيدة، تمثلت في استخراج المناطق الورمية كافة.

وجرى بعد ذلك استخراج السمات الأساسية التي تقوم بتمييز الأورام الخبيثة عن الأورام الحميدة، فقد بلغ عدد السمات المستخرجة لكل منطقة ورمية 14 سمة، وبعدها تم تقليل أبعاد متجه السمات المستخرجة، واختيار أكثر 6 سمات مثلى قادرة على التمييز بين نوعي الأورام لتشكل متجه دخل متناسب مع أداء المصنف.

وتم تدريب هذه الخوارزمية على 81 صورة من التصوير المقطعي المحوسب، ومن ثم تطبيق النظام على 24 صورة من قاعدة البيانات، لم يتم تدريب الشبكة عليها. وقد قام النظام بتصنيف الأورام المكتشفة في هذه الصور. كما أظهر النظام المقترح كفاءة عالية من الناحية التجريبية في تجزئة الأورام في الصورة المقطعية، إذ تم تقييم أداء خوارزمية التجزئة وخوارزمية التصنيف. وأثبت عند المقارنة أنهما تفوقا على العديد من الأبحاث والدراسات السابقة والمذكورة في الجدول (٥-١٥) من حيث الدقة التي بلغت 93.1، إلا بالنسبة للدراسة [3] التي حققت نسبة 95 ويتوقع ذلك بسبب اعتمادها على سمات ثلاثية الحجم 3D تعطي خصائص ومواصفات ورمية أكثر دقة، والدراسة [5] التي حققت دقة 94.56 أعلى قليلاً بسبب استخدامها لعدد أكبر من بيانات التدريب.

٥-٤-الاستنتاجات:

- أظهر النظام أهمية دمج نظامي التجزئة والتصنيف، مقارنة بالأنظمة التي تعتمد على التجزئة فقط.
- أظهر النظام كفاءة فيما يأتي:
  - إزالة الضجيج والتشوهات والتباين في الكثافة والسويات الرمادية بين الأنسجة المختلفة من دون التأثير على حجم الورم وحدوده.

- استخراج المناطق الورمية على اختلاف الحالات المقدمة له، إذ تمكن من عزل المناطق الورمية في الصور التي تتضمن أوردة وشرايين دموية وأنسجة رئوية.
- اكتشاف الأورام في الصور غير الواضحة التي تحوي نسبة ضجيج عالية.
- استخراج الأورام في الصور التي تحوي أوراماً غير واضحة ومبهمة الحدود.
- استخراج حدود المناطق الورمية وحوافها بدقة جيدة، ولم يتم استخراج أية مناطق إضافية.
- عدم استخراج أي منطقة في الصور السليمة التي لا تحتوي على مناطق ورمية.

- كما أظهر النظام كفاءة في تقليل السمات الأكثر أهمية فقط وتحديدها، والتي تتوافق طبقاً لعمل المصنف من دون التأثير في دقة التصنيف.
- قام النظام بتصنيف الأورام بدقة عالية من خلال عدد محدود من السمات وبأسرع زمن ممكن.
- أظهر النظام فاعلية في تحقيق عملية التشخيص الطبي من خلال إنجاز عملية التشخيص بشكل آلي، ومن دون تدخل المستخدم في أي مرحلة من مراحل العمل.

٥-٥-المقترحات والتوصيات:

١.العمل على دمج الخوارزميات المستخدمة مع خوارزميات الأمثلة، من أجل تحسين عملية العنقدة.

٢.تطبيق تقنيات تحسين تباين الصورة بعد خطوة إزالة الضجيج، من أجل الحصول على صورة محسنة جاهزة للدخول إلى مرحلة التجزئة، ما يعطي نتائج أكثر دقة.

٣.استخراج سمات ذات بعد أكبر (ثلاثية البعد) كالحجم volume والكروية sphericity، ودراسة أثر هذه السمات في عملية ودقة التصنيف.

٤. تطبيق طرق اختيار السمات المثلى بعد عملية الاستخراج، مثل طرق الغلاف Wrapper والمرشحات Filter والمضمنة Embedded، بالإضافة إلى اعتماد معايير الملاءمة بين السمات، كقياسات المسافة وكمية المعلومات وغيرها، والتي تؤدي إلى خفض زمن التصنيف، وتقليل مشكلة تخزين البيانات الزائدة.

٥.زيادة حجم العمل المستقبلي من خلال تصنيف المراحل المختلفة للسرطان وتحديدها، وبيان درجة تقدمه.

٥-٦-خلاصة:

تم في هذا الفصل تقييم النظام المقترح وعرض نتائج استخراج الصور وتصنيف الأورام، كما تم ذكر أهم الآفاق المستقبلية والمقترحات التي من الممكن العمل على إنجازها في أبحاث قادمة كتطوير لفكرة هذا البحث.

المراجع:

[1] Naseem, R., Alimgeer, K. S., & Bashir, T. (2017). Recent trends in computer aided diagnosis of lung nodules in thorax CT Scans, IEEE.

[2] Dostert, C., Van Bruggen, R., Steele, C., Mossman, B. T., & Tschopp, J. (2008). Innate immune activation through Nalp3 inflammasome sensing of asbestos and silica. Science, 320(5876), pp. 674-677.

[3] Dabade, S., Chaudhari, S., Jadhav, S., & Nichal, A. (2017). A review paper on computer aided system for lung cancer detection, In International Conference on BigData, IoT and Data Science (BID).

[4] Naseem, R., Alimgeer, K. S., & Bashir, T. (2017). Recent trends in Computer Aided diagnosis of lung nodules in thorax CT scans. In International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), pp. 1-12, Karachi, Pakistan. IEEE.

[5] Wang, Q. and Vallejo, M. (2020). Fluorescence Lifetime Endomicroscopic Image-based ex-vivo Human Lung Cancer Differentiation Using Machine Learning. In Proceedings of the International Conference on Computer Vision. IEEE.

[6] Reddy, U. J. and Reddy, B. V. R. (2019). Categorization & Recognition of Lung Tumor using Machine Learning Representations, Current Medical Imaging, 15, pp. 405-413.

[7] Ahmed, S.R. and Barazanchi, I. (2019). Lung cancer classification using data mining and supervised learning algorithms on multi-dimensional data set. In Periodicals of Engineering and Natural Sciences, Vol. 7, No. 2, pp. 438-447. DOI:http://pen.ius.edu.ba.

[8] Ucar, M. and Ucar, E. (2019). Computer-Aided Detection of Lung Nodules in Chest X-Rays using Deep Convolutional Neural Networks, Sakarya University Journal, Vol. 2, No. 1.

[9] Lakshmanaprabu, S. K. and Mohanty, S. N. (2019). Optimal deep learning model for classification of lung cancer on CT images, Future Generation Computer Systems.

[10] Gong, J., Liu, J., Wang, L., & Zheng, B. (2018). Automatic detection of pulmonary nodules in CT images by incorporating 3D tensor filtering with local image feature analysis. Physica Medica 46, PP.124–133. DOI: https://doi.org/10.1016/j.ejmp.2018.01.019. PMID: 29519398.

[11] Singh, G. A. and Gupta, P. K. (2018). Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans, The Natural Computing Applications Forum 2018, DOI: https://doi.org/10.1007/s00521-018-3518-x.

[12] Baker, D. and Kilpatrick, J. (2017). Predicting Lung Cancer Incidence from CT Imagery. In International Conference on Computing, Communication & Automation. IEEE.
[13] Fan, L., Xia, Z., & Zhang, X. (2017). Lung Nodule Detection Based on 3D Convolutional Neural Networks. Computers in Biology and Medicine, 43(4), PP. 287–300.
[14] Wu, Q. and Zhao, W. (2017). Small-Cell Lung Cancer Detection Using a Supervised Machine Learning Algorithm, Current Medical Imaging Reviews, 14(1), PP. 108–117.
[15] Dhara, A., Mukhopadhyay, S., Dutta, A., & Garg, M. (2016). A Combination of shape and texture features for classification of pulmonary nodules in lung CT images. J Digit Imaging 2016, 29(4), PP. 466-475.
[16] Rendon-Gonzalez, E. and Ponomaryov, V. (2016). Automatic Lung nodule segmentation and classification in CT images based on SVM. 9Th International Kharkiv Symposium on Physics and Engineering of Microwaves, Millimeter and Submillimeter Waves (MSMW).
[17] Kraft, M. (2011). Approach to the patient with respiratory disease. In: Goldman L, Schafer AI, (Eds.), Goldman's Cecil Medicine (pp. 362-356). Philadelphia, PA: Saunders Elsevier.
[18] Sayar, A., Turna, A., Kılıcgun, A., Solak, O., Urerurer, N., & Gurses, A. (2004). Prognostic significance of surgical-pathologic multiple-station N1 disease in nonsmall cell carcinoma of the lung. European journal of cardio-thoracic surgery, 25(3), pp. 434-438.
[19] Schild, S. E., Hillman, S. L., Tan, A. D., Ross, H. J., McGinnis, W. L., & North Central Cancer Treatment Group. (2017). Long-Terms Results of a Trial of Concurrent Chemotherapy and Escalating Doses of Radiation for Unresectable Non–Small Cell Lung Cancer. Journal of Thoracic Oncology. 25(31), pp. 5030-5041.
[20] ناصر، فيصل-جركس، عامر. (٢٠٠٥). الأشعة التشخيصية والإيكوغرافي. سورية. مديرية الكتب والمطبوعات.
[21] Vansteenkiste, J. F., Stroobants, S. G., Dupont, P. J., De Leyn, P. R., Verbeken, E. K., & Demedts, M. G. (1999). Prognostic importance of the standardized uptake value on 18F-fluoro-2-deoxy-glucose–positron emission tomography scan in non–small-cell lung cancer: an analysis of 125 cases. Journal of Clinical Oncology, 17(10), pp. 3201- 3206.
[22] Strauss, G. M., Herndon, J. E., Maddaus, M. A., Johnstone, D. W., Johnson, E. A., & Vokes, E. E. (2008). Adjuvant paclitaxel plus carboplatin compared with observation in stage IB non–small-cell lung cancer: CALGB 9633 with the Cancer and Leukemia Group B, Radiation Therapy Oncology Group, and North Central

Cancer Treatment Group Study Groups. Journal of Clinical Oncology, 26(31), pp. 5043-5051.
[23] Miller, V. A., Kris, M. G., Shah, N., Patel, J., Gomez, J., & Tyson, L. (2004). Bronchioloalveolar pathologic subtype and smoking history predict sensitivity to gefitinib in advanced non–small-cell lung cancer. Journal of Clinical Oncology, 22(6), pp. 1103-1109.
[24] Wilk, K. (2021). The Cancer Imaging Archive posts COVID-19 imaging data to benefit community. National Institutes of Health (NCI).
[25] Tsubamoto, M., Kuriyama, K., Kido, S., Arisawa, J., Kohno, N., Johkoh, T., & Kuroda, C. (2002). Detection of Lung Cancer on Chest Radiographs: Analysis on the Basis of Size and Extent of Ground-Glass Opacity at Thin-Section CT. Radiology, 224(1), pp. 139-144.
[26] Quekel, L. G. B. A., Goei, R., Kessels, A. G. H., & van Engelshoven, J. M. A. (2001). Detection of lung cancer on the chest radiograph: impact of previous films, clinical information, double reading, and dual reading. Journal of clinical epidemiology, 54(11), pp. 1146-1150.
[27] Van Ginneken, B., Romeny, B. T. H., & Viergever, M. A. (2001). Computer-aided diagnosis in chest radiography: a survey. IEEE Transactions on medical imaging, 20(12), pp. 1228-1241.
[28] Qidwai, U. and Chen, C. H. (2009). Digital image processing: an algorithmic approach with MATLAB. New York: CRC press.
[29] Pratap, G. P. and Chauhan, R. P. (2016). Detection of Lung cancer cells using image processing techniques. In Proceedings of IEEE international conference on power electronics, intelligent control and energy systems (ICPEICES), pp. 1-6.
[30] Korfiatis, P., Skiadopoulos, S., Sakellaropoulos, P., & Costaridou, L. (2014). Combining 2D wavelet edge highlighting and 3D thresholding for lung segmentation in thin-slice CT. The British journal of radiology, 3(6), pp. 1-13.
[31] Assefa, M., Faye, I., Malik, A. S., & Shoaib, M. (2013). Lung nodule detection using multi-resolution analysis. In Proceedings of the International Conference on Complex Medical Engineering (pp. 457-461). Japan: The Graduate School of Natural Science and Technology.
[32] Arulmurugan, R. and Anandakumar, H. (2018). Early detection of lung cancer using wavelet feature descriptor and feed forward back propagation neural networks classifier. In Computational vision and bio-inspired computing (pp. 103–110). Springer: New York, US.
[33] Chau, F., Linag, Y., Gao, J., & Shao, X. (2004)**. Chemometrics: from basics to wavelet transform.** Wiley, United States of America, p:316.
[34] Misiti, M., Misiti, y., Oppenheim, G., & Poggi, J. (2007). **Wavelets and their Applications.** ISTE, Great Britain and the United States, p:352.

[35] Merry, R. (2005). **Wavelet Theory and Applications A Literature Study**. Eindhoven University of Technology, Department of Mechanical
Engineering, Control Systems Technology Group.
[36] Sumithra, A., Member, M., IACSIT, B., & Thanushkodi, K. (2009). Performance Evaluation of Different Thresholding Methods in Time Adaptive Wavelet Based Speech Enhancement. International Journal of Engineering and Technology (IACSIT), Vol. 1, No. 5, Pp. 439-447.
[37] Nagabushanam, M. and Ramachandran, S. (2019). Design And Implementation of Parallel and Pipelined Distributive Arithmetic Based Discrete Wavelet Transform IP Core. European Journal of Scientific Research, Vol. 35, No. 3, pp. 378-392.
[38] Gonzalez, C. and Woods, R. E. (2018). Digital Image Processing 3rd edition.
[39] Bhargavi, K. and Jyothi, S. (2014). A Survey on Threshold Based Segmentation Technique in Image Processing. International Journal of Innovative Research and Development, 3(12), pp. 76-85.
[40] John, J. and Mini, M. (2016). Multilevel Thresholding based segmentation and feature extraction for pulmonary nodule detection. Procedia Technol 2016, 24, pp. 957-963.
[41] Qu, Z. and Zhang, L. (2010). Research on image segmentation based on the improved Otsu algorithm. In Proceedings of the 2nd International Conference on Intelligent Human-Machine Systems and Cybernetics (pp. 228-231). Nanjing: China.
[42] Otsu, N. (1979). Thresholds selection method form grey-level histograms. IEEE Trans. On Systems, Man and Cybernetics, 9(1).
[43] Sangamithraa, P. and Govindaraju, S. (2016). Lung tumor detection and classification using EK-Mean clustering. In International Conference on Wireless Communications, Signal Processing and Networking.
[44] Sahu, S. P., Agrawal, P., Londhe, N. D., & Verma, S. (2019). Lung segmentation of CT images using fuzzy C-means for the detection of cancer in early stages. In Advances in data and information sciences (pp. 167–176). Springer: New York, US.
[45] Liu, H., Geng, F., & Zhang, C. (2017). A fast weak-supervised pulmonary nodule segmentation method based on modified self-adaptive FCM algorithm. Soft Comp 2017, 12(22), pp. 3983-3995.
[46] Nithya, E. and Kumar, S. (2016). Segmentation of lung nodule in CT data using active contour model and fuzzy C-means clustering. Alexandria Eng J, 55, pp. 2583-2588.
[47] Suzuki, K. (2014). Pixel-based machine learning in computer-aided diagnosis of lung and colon cancer. In Machine Learning in Healthcare Informatics. Springer, pp. 81–112.

[48] Min, S., Lee, J., & Han, I. (2006). Hybrid genetic algorithms and support vector machines for bankruptcy prediction. Expert Systems with Applications, 31(3), pp. 652-660.
[49] Steinwart, I. and Christmann, A. (2018). Support vector machines. Springer Science & Business Media.
[50] Chen, F. and Li, C. (2010). Combination of feature selection approaches with SVM in credit scoring. Expert Systems with Applications, 37(7), p. 4902-4909.
[51] Bellotti, T. and Crook, J. (2019). Support vector machines for credit scoring and discovery of significant features. Expert Systems with Applications, 36(2), p. 3302-3308.
[52] Danenas, P. and Garsva, G. (2015). Selection of Support Vector Machines based classifiers for credit risk domain. Expert Systems with Applications, 42(6), p. 3194-3204.
[53] Zhou, S., Cheng, Y., & Tamura, S. (2014). Automated lung segmentation and smoothing techniques for inclusion of juxtapleural nodules and pulmonary vessels on chest CT images. Biomedical Signal Processing and Control, 13(3), PP. 62-70.
[54] Kumar, B. and Vidhya, R. (2011). Signal Denoising with Soft Threshold by Using Chui-Lian Multiwavelet. International Journal of Electronics & Communication Technology, Vol. 2, No. 1, P. 38-42.
[55] Aggarwal, R. and Singh, j. (2011). Noise Reduction of Speech Signal Using Wavelet Transform with Modified Universal Threshold. International Journal of Computer Applications, V. 20, No. 5, P. 14-19.
[56] Zhang, J., Zhang, W., Chen, C., Guan, Y., & Wang, C. (2010). Computed diagnosis system for lung tumor detection based on PET/CT images. In Proceedings of the 3rd International Conference on Biomedical Engineering and Informatics (pp. 166-170). Yantai University: China.
[57] The MATLAB Team. (2011). fuzzy C-Means Clustering Tutorial (Fuzzy Logic Toolbox). MATLAB, Product Help, Version 7.8.0347 (R2011a).
[58] Mahmud, M. A., Karim, A. Z., & Rahman, M. M. (2015). Brain Tumor Detection using MR Images through Pixel based Methodology. Global Journal of Computer Science and Technology, 15(4), PP. 302-315.
[59] Huang, P., Park, S., Yan, R., & Lee, J. (2017). Added value of computer-aided CT image features for early lung cancer diagnosis with small pulmonary nodules. Radiology, 286, PP. 286–295.
[60] Theresa, M. and Bharathi, V. S. (2017). Computer aided diagnostic (CAD) for feature extraction of lungs in chest radiograph using different transform features. Biomed. Res, PP. 1-8.
[61] Xie, Y., Zhang, J., Xia, Y., Fulham, M., & Zhang, Y. (2018). Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT. Information Fusion, 42, pp. 102-110.

[62] Gindi, A. M., Al Attiatalla, T. A., & Sami, M.M. (2014). A Comparative Study for Comparing Two Feature Extraction Methods and Two Classifiers in Classification of Early stage Lung Cancer Diagnosis of chest x-ray images. Journal of American Science, 10(6), PP. 13-22.

[63] Cao, P., Liu, X., Yang, J., & Zhao, D. (2017). A multi-kernel based framework for heterogeneous feature selection and over-sampling for computer-aided detection of pulmonary nodules. Pattern Recognition, 64, pp. 327-346.

[64] Akram, S., Javed, M. Y., Akram, M. U., Qamar, U., & Hassan, A. (2016). Pulmonary nodules detection and classification using hybrid features from computerized tomographic images. Journal of Medical Imaging and Health Informatics, 6(1), PP. 252–259.

[65] Dandıl. E. (2018). A Computer-Aided Pipeline for Automatic Lung Cancer Classification on Computed Tomography Scans. Journal of Healthcare Engineering, p:12.

[66] CarvalhoFilho, A. O., Silva, A. C., de Paiva, A. C., & Gattass, M. (2016). Computer-aided diagnosis system for lung nodules based on computed tomography using shape analysis, a genetic algorithm and SVM. Med. Biol. Eng. Comput, pp. 1-8.

[67] Kaucha D. P. (2017). Early Detection of Lung Cancer using SVM Classifier in Biomedical Image Processing. In International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Chennai, pg. 3143-3148.

[68] Li, C., Zhu, G., Wu, X., & Wang, Y. (2018). False-positive Reduction on Lung Nodules Detection in Chest Radiographs by Ensemble of Convolutional Neural Networks. IEEE Access, 6(99), pp. 16060–16067.

[69 https://towardsdatascience.com/metrics-to-evaluate-your semantic segmentation-model-6bcb99639aa2.

[70] https://towardsdatascience.com/how-accurate-is-image-segmentation-dd448f896388.

[71] https://www.jeremyjordan.me/evaluating-image-segmentation-models/.

Abstract:

The entry of the computer into many areas led to the development of new technologies in many fields such as the medical field, which helped doctors in detecting and diagnosing diseases accurately and credibly. In addition to the doctor's experience, the accuracy of the computer leads to a high diagnostic capacity, and contributes greatly to the success of surgeries and saving many lives. Lung cancer is ranked first in the list of the most common cancers on the planet, but making a therapeutic decision such as removing pulmonary nodules requires prior knowledge of the dimensions and specifications of these nodules, so it was necessary to work on developing new automated methods to extract these tumours and determine their characteristics.

The research aims to develop a new algorithm to detect and classify lung tumours in CT images based on digital image processing techniques and artificial intelligence algorithms. The proposed algorithm includes four basic stages, begins with the pre-processing of the images, which is performed using the wavelet transform technique to reduce the noise in the images, then the improved images are segmented in the second stage to separate the different parts, which increases the ability to divide the tumor into the later stages, the next stage depends on applying the Fuzzy C-Mean Clustering algorithm, and selecting an otsu threshold that ensures correct extraction of all objects in the images since the tumor has different gray levels in each image, in this stage it is essential to apply the needed Morphological processes (erosion and dilation) to remove some unwanted objects from the threshold images, the final part in this stage is subtracting the two images resulting from the processes of erosion and dilation to detect the tumor area. In the third stage, the most important 6 optimal features were extracted in order to be relied upon in classification and diagnosis. The final stage classifying the extracted tumor as benign or malignant nodules by means of the SVM algorithm.

The segmentation algorithm was tested on a database consisting of 137 tumoral and non-tumoral images and achieved a high accuracy of up to 96.2%. The experimental results of applying the classification system to 24 random cases (containing 29 tumors), which were not trained, also found that the system also achieved high accuracy by 93.1%.

key words:

Automated Lung Cancer Detection, Wavelet Transformation, Digital Image Processing and Segmentation, FCM Cluster Algorithm, SVM Algorithm, Pulmonary Nodule Classification.

Syrian Arab Republic
Damascus University
Faculty of Mechanical and Electrical Engineering
Department of Computers and Automation Engineering
Master in Computers Engineering and its Networks

Developing an Algorithm for Segmentation of Pulmonary Nodule Images for Early Detection of Lung Cancer

A dissertation submitted To Damascus University in partial fulfillment of the requirements for the degree of Master in Computers Engineering and its Networks at Faculty of Mechanical and Electrical Engineering

By:
Eng. Mohammad Jameel Ayoush

Main Supervisor:
Prof. Eng. Dr. Maamoun Younes
Assistant Professor in the
Department of Computers &
Automation Engineering

Associate Supervisor:
Prof. Eng. Dr. Ayman Saboni
Assistant Professor in the
Department of Biomedical
Engineering

2022-2023