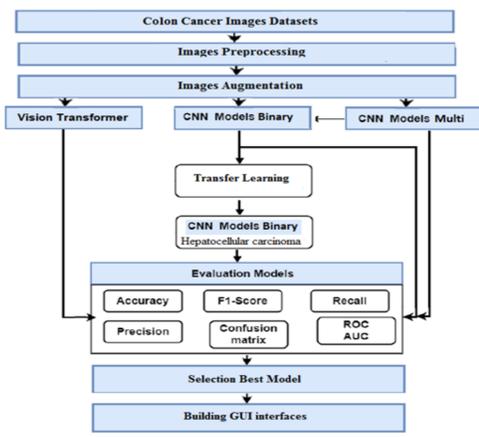


تطوير نظام خبير للكشف المبكر عن سرطان القولون Development of an Expert System for Colon Cancer Early Detection

إعداد عدنان أسبر
بإشراف أ.د. رشا مسعود

القسم العملي

تُقدم هذه الفقرة استراتيجية مقترحة للكشف عن سرطان القولون والمستقيم باستخدام صور ناتجة عن RT-PCR كما هو موضح في الشكل (1)، تتضمن الطريقة المقترحة عدة خطوات: جمع الصور؛ المعالجة المسبقة للصور واستخلاص السمات، تعزيز الصور؛ ثم استخدام ثلاث نماذج للتعلم العميق (CNN) أحدهما نموذج للتصنيف ثنائي والأخر للمتعدد، والثاني استخدام المحول Vision Transformer (ViT) لتعزيز أداء التنبؤ الثنائي بسرطان القولون والمستقيم، وأخيراً يتم مقارنة النتائج وتقييم النماذج المقترحة من أجل اعتماد النموذج الأفضل للتنبؤ بسرطان القولون والمستقيم، الشكل يتضمن المنهجية المقترحة للكشف عن سرطان القولون والمستقيم



الشكل (1) يتضمن المنهجية المقترحة

القسم العملي

استُخدمت في هذا البحث الطرق والأجهزة والبرمجيات التالية:

تم استخلاص الحمض النووي الرايبوزي (RNA من ٣٠ عينة عن طريق عزل الرنا باستخدام العتيقة - عتيقة InviTrap® Spin Cell RNA Mini Kit شركة STRATEC الألمانية والحصول على RNA بجودة عالية، ومن ثم تحويل RNA إلى cDNA باستخدام أنزيم FIREPOL reverse transcriptase من شركة Solis BioDyne، وتم تصميم المشرع primer من أجل تضخيم جينة الـ C-FLIP من أجل انجاز تفاعل البوليميراز السلسلي بالزمن الحقيقي، وتم العمل على جهاز الـ IMx3005P QPCR System المصنوع من قبل Agilent Stratagene، حيث يمكن متابعة نتائج كل دورة من RT-PCR من خلال استخدام صبغات متفلورة ثم ترجم النتائج بشكل مخططات بيانية على شكل صورة، تم معالجة البيانات وفق تابع يربط الفلورة وما يقابلها من عدد الدورات الحرارية على مقياس لوغاريتمي حيث تظهر قيمة Ct على شكل قيمة رقمية، وبالتالي حساب قيمة التعبير الجيني لبروتين C-FLIP وفق قاعدة Livak، تم حساب التعبير الجيني لكل العينات وتصنيفها، وتم بناء شبكة تعلم عميق (CNN) تستخدم صور ناتجة عن (RT-PCR) وتدريب هذه الشبكة واختبارها على صور (qPCR) المأخوذة لمنحنى تضاعف مورثة بروتين C-Flip للحصول على نتائج تفيد بوجود أو عدم وجود سرطان القولون حيث كانت دقة التدريب ٩٠,٢%، ودقة الاختبار ٩١,٣%.

البرمجيات المستخدمة: تم كتابة الكود بلغة بايثون (Python)، وذلك لاحتوائه على كافة المكتبات اللازمة لعمل المشروع الـ (مكتبة Keras ومكتبة TensorFlow) في الخلفية.

الملخص

تضمنت الدراسة الحالية محورين: الأول هو التحري عن بروتين الـ C-Flip وعلاقته بحدوث أورام القولون كواسم حيوي، والثاني استخدام المعلوماتية الحيوية في تقييم الواسم الحيوي بروتين C-Flip في الكشف المبكر والتنبؤ بسرطان القولون عن طريق قياس التعبير الجيني لهذا البروتين باستخدام تقنية RT-PCR وقسمت العينات في هذه الدراسة إلى ثلاثة مجموعات أولاً: جمع ٣٠ عينة مضافاً لها بيانات أخذت من جهاز qPCR عددها ٢٣٦ صورة (٢٠٠ صورة لأشخاص أصحاء، ٩٦ صورة لأشخاص مرضى)، ثانياً: تعزيز الصور Images Augmentation التصحيح ٤٣٢ صورة (٢٩١ صورة لأشخاص أصحاء، ١٤١ صورة لأشخاص يعانون من سرطان القولون)، ثالثاً: بيانات أخذت من جهاز qPCR لأشخاص مصابين بسرطان الكبد ١٤٢ صورة، فتمت هذه الدراسة نظام خبير يساعد الطبيب في التشخيص والكشف المبكر عن سرطان القولون (الورم وتحديد درجته)، ويدعم قراره في تصنيف نوع الورم فيما إذا كان ورم قولوني حميد أو ورم قولوني خبيث، باستخدام خوارزميات التعلم العميق، وهذا النظام عبارة عن تطبيق قائم على الشبكة العصبونية الالتفافية العميقة مخصص لاستقبال صور من RT-PCR وتحليلها مكتوباً بلغة الـ Python واستخدامه في التشخيص، حيث تم في هذا البحث اقتراح منهجية تتألف من ثلاث نماذج للتعلم العميق نموذجين للشبكة العصبونية الالتفافية (CNN) أحدهما نموذج للتصنيف ثنائي والأخر للمتعدد، والثاني استخدام المحول Vision Transformer (ViT) لتعزيز أداء التنبؤ الثنائي بسرطان القولون، أظهرت نتائج تدريب واختبار النموذج وفقاً للتصنيف الثنائي باستخدام الشبكة العصبونية الالتفافية (CNN) على مجموعتي البيانات المذكورتين (الأولى والثانية)، أن مقياس الأداء تشير إلى أن دقة النظام على مجموعة البيانات الثانية كانت ٩٠,٥٤%، في بيانات التدريب و٩٤,٥٢% في بيانات الاختبار، وعند إجراء تدريب واختبار النموذج على مجموعتي البيانات (الأولى والثانية) باستخدام التصنيف الثنائي عبر محول الرؤية (ViT)، أظهرت مقياس الأداء أن دقة النظام على مجموعة البيانات الثانية بلغت ٩٦,١٠% في بيانات التدريب و٩٤,٤٣% في بيانات الاختبار،

النتائج والمناقشة

في هذه الدراسة تم التركيز على الكشف المبكر عن سرطان القولون والمستقيم وتصنيفه باستخدام صور PCR عن طريق قياس التعبير الجيني لـ بروتين C-Flip. أثبتت المنهجية المقترحة في هذه الدراسة فعاليتها في الكشف عن سرطان القولون والمستقيم بالاعتماد على خوارزميات التعلم العميق.

نتيجة للمقارنة بين النموذجين CNN و ViT اللذين تم استخدامهما في تطوير نظام خبير يهدف إلى الكشف المبكر عن سرطان القولون، حيث يُعتبر نموذج CNN أداة فعالة وسريعة في عملية التنبؤ، ويتميز بدقة ٩٤,٥٢% في تصنيف وتشخيص الصور المدخلة، في المقابل يوفر النموذج اللغوي الكبير دقة أعلى في تصنيف وتشخيص الصور، إلا أنه يتطلب وقتاً أطول في عملية التنبؤ مقارنة بنموذج CNN.

المراجع

1. Khan J, Wei JS, Ringnér M, et al (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nature Medicine, 7(6), 673-679.
2. Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D., & Levine, A. J. (1999). Broad patterns of gene expression were revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proceedings of the National Academy of Sciences, 96(12), 6745-6750.
3. Ma, E., Sasazuki, S., Iwasaki, M., Inoue, M., & Shoichiro Tsugane for the Japan Public Health Center-based Prospective Study Group. (2010). 10-Year risk of colorectal cancer: development and validation of a prediction model in middle-aged Japanese men. Cancer epidemiology, 34(5), 534-541.
4. Ryu, B. K., Lee, M. G., Chi, S. G., Kim, Y. W., & Park, J. H. (2001). Increased expression of cFLIP in colonic adenocarcinoma. The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland, 194(1), 15-19.

القسم النظري

يعتبر قياس التعبير الجيني عنصراً أساسياً في فهم كيفية عمل الخلايا والأنسجة والكائنات الحية، ويُشير التعبير الجيني إلى عملية تحويل المعلومات المشفرة في الحمض النووي (DNA) إلى بروتينات وظيفية [1]، حيث تُستخدم تقنيات مختلفة لقياس التعبير الجيني، مثل: هناك العديد من الدراسات المرجعية التي تناولت الكشف المبكر عن سرطان القولون والمستقيم، ويمكن تصنيف هذه الدراسات إلى ثلاثة أقسام حسب البيانات المستخدمة لبناء نظام التنبؤ: فمنها ما استخدم بيانات وراثية مثل المصفوفة الميكروية أو تسلسل الـ DNA وهذا النوع من الدراسات يعتمد على طرق الذكاء الصناعي وتقنيات إيجاد الحل الأمثل في إيجاد الجينات المرتبطة بالمرض وانتفاها وطرق التصنيف المختلفة للتنبؤ والمشكلة الأساسية التي تواجه هذا النوع من الدراسات هي أن عدد الجينات يكون أكبر من عدد العينات المدروسة [2]، ومنها ما استخدم عوامل الخطورة والبيانات السريرية لبناء نموذج قادر على التنبؤ باحتمال الإصابة بالمرض وحساب مؤشر ذو درجات لهذا الاحتمال وهناك العشرات من هذه النماذج، وتعتمد هذه النماذج على التحليل الخطي [3]، أما القسم الثالث فيشمل أنظمة للتنبؤ تعتمد على نوعي البيانات مع الوراثة مع السيريبي وهي دراسات كثيرة، حيث قام الباحث Kyu Ryu و زملائه في جامعة كيونغ هي في كوريا الجنوبية) في عام ٢٠٠١ حيث تم أخذ ٥٢ عينة لسرطان القولون متباينة المراحل و١٥٠ من الأورام الحميدية الغدية، ودراسة مستويات التعبير الجيني لهذه العينات ومقارنتها مع الجينة الاستتبابية الثابتة للتعبير، تم إجراء التحليلات الإحصائية وفق اختبار Wilcoxon test وتم استنتاج ان التعبير الجيني لبروتين C-FLIP مرتفع جداً في العينات السرطانية وبالتالي ان بروتين الـ C-FLIP هو واهم حيوي لسرطان القولون [4].