

نظام التحديد الآلي للطفرة الطورية في خلايا الكرموسومات للتعرف على حالات التشوهات الخلقية الأكثر شيوعاً

إعداد

ريم بنت محمد يوسف عبدالله باشميل

بحث مقدم لنيل درجة الماجستير في علوم الحاسبات

البحث بدعم من مدينة الملك عبدالعزيز للعلوم والتقنية
رقم المنحة (أط-٣٧-١١٦٥)

إشراف

د. لمياء عبد الله الرفاعي

د. وديع الحلبي

كلية الحاسبات وتقنية المعلومات

جامعة الملك عبدالعزيز

جدة - المملكة العربية السعودية

شوال ١٤٣٩ هـ - يونيو ٢٠١٨ م

نظام التحديد الآلي للطفرة الطورية في خلايا الكروموسومات للتعرف على حالات التشوهات الخلقية الأكثر شيوعاً

المستخلص

يعتبر تنميط الكروموزومات البشرية إجراءً سريريًا مهمًا في فحص وتشخيص السرطانات والاضطرابات الوراثية. تنميط الكروموزومات هو تقنية قياسية تستخدم لتصنيف الكروموسومات الطورية إلى ٢٤ نوعًا. وهو الجزء الأساسي من التعرف على التثلث الصبغي الأكثر شيوعاً. و الذي يعتبر أيضا مشكلة في مجال التعرف على الأنماط لعقود من الزمن. التثلث الصبغي هو نوع من التشوهات الخلقية حيث توجد ثلاث نسخ من أحد الكروموسومات، بدلا من اثنين في الحالات العادية. أكثر أنواع الكروموسومات شيوعاً في البشر التي تبقى حتى الولادة هي (التثلث الصبغي في كرموزوم ١٣ (متلازمة باتو)، التثلث الصبغي في كرموزوم ١٨ (متلازمة إدواردز)، التثلث الصبغي في كرموزوم ٢١ (متلازمة داون)، التثلث الصبغي في كروموسومات X and Y (XXX) (متلازمة اكس ثلاثية)، (متلازمة كلاينفيلتر)، (XYY).

وللمساهمة في تحسين سرعة ودقة عملية التصنيف، اقترحنا أن نقوم بتصميم نظام آلي للتعرف على التثلث الأكثر شيوعاً (نسخة إضافية من كروموسوم) من صورة الصبغيات الطورية. لبناء هذا النظام، هناك أربع خطوات. تتمثل الخطوة الأولى في تحسين الصورة من أجل إعداد هذه الصورة للحصول على أفضل أداء من اكتشاف وتصنيف الكروموسومات. الخطوة التالية هي تجزئة الكروموسومات التي تميز الكروموسوم عن الخلفية أو "غير الكروموسومات" أثناء عملية التحليل للصورة الرقمية. الخطوة الثالثة

هي استخراج الخصائص. يتم استخلاص أفضل خصائص الكروموسوم (طول الكروموسوم، الطول النسبي، مساحة الكروموسوم، المنطقة النسبية، مؤشر السنترومير ، ومظهر الكثافة) لتصنيف الكروموسومات باستخدام قاعدة بيانات الصور المحلية المجمعة التي تتكون من ١٥٧ صورة تحتوي على ٧٢٦٦ كروموسوم. رابعاً، تصنيف الكروموسومات، يقوم النظام المقترح بتصنيف الكروموسومات وفقاً للخصائص المذكورة عن طريق تنفيذ وتدريب مصنف الطبقات العصبية متعدد الطبقات مع autoencoder لتصنيف الكروموسومات إلى طبقات الكروموسومات المستهدفة. كانت دقة التصنيف للتدريب ١٠٠٪ وأفضل نتيجة سيناريو الاختبار ما بين (٨٨,٩٪ - ١٠٠٪).

Automated identification of abnormal metaphase chromosome cells for the recognition of most common trisomies

By

Reem Mohammed Yousuf Bashmail

**A Thesis Submitted in Partial Fulfillment of the
Requirements for the Master Degree in Computer Science**

Supervised By

Dr. Lamiaa Elrefaei

Dr. Wadee Alhalabi

Faculty of Computing and Information Technology

KING ABDULAZIZ UNIVERSITY

JEDDAH-SAUDI ARABIA

Shawwal 1439H – June 2018 G

Automated identification of abnormal metaphase chromosome cells for the recognition of most common trisomies

ABSTRACT

Karyotyping of human chromosomes has become an important clinical procedure in the screening and diagnosis of cancers and genetic disorders. Karyotyping is a standard technique utilized to classify metaphase chromosomes into 24 types. It is the essential part of the chromosome trisomy abnormality (an extra copy of a chromosome) recognition. It is also considered as important pattern recognition problem for decades. A trisomy is an abnormality type where there are three copies of a particular chromosome, instead of two in normal cases. The most common chromosomes trisomy types in humans that survive to birth are (trisomy 13 (Patau syndrome), trisomy 18 (Edwards syndrome), trisomy 21 (Down syndrome), and trisomy of sex chromosomes (XXX (Triple X syndrome), XXY (Klinefelter syndrome), XYY). To improve the speed and accuracy of the classification process, we propose an automated system to recognize the most common trisomy. To build such system, there are four steps. The first step is the chromosome's image enhancement in order to prepare the image before performing detection of metaphase spreads and classification of banded chromosome patterns. Difference of Gaussian (DoG) is used followed by inverse gradient of each pixel of the input image. The second step is chromosome segmentation in order to differentiate the chromosome from "non- chromosome" or background by applying OTSU's thresholding then applying series of morphological operations.

Third is feature extraction. The best chromosome's features (chromosome length, relative length, chromosome area, relative area, centromere index, and density profile) is extracted to classify the chromosomes using assembled local images database which Consisting of 157 image that involving 7266 chromosomes. Forth is, Chromosome classification, the proposed system is classifying the chromosomes according to the mentioned features by implementing and training multi-layer neural network classifier with autoencoder to classify chromosomes into targeted chromosomes classes. The classification accuracy for the training was 100% and the best testing scenario result is between (88.9% - 100%).