جامعة الموصل
كلية علوم الحاسوب والرياضيات

# نموذج ذكائي هجين مقترح لتحليل الإنحرافات الجينية

**ليلـى هـادي جويـر**

رسالة ماجستير

الرياضيات / حاسوبية

**إشراف**

**د. شـذى عـبدالله محمـد رمضـان**

**مدرس**

**1443هـ** **2021م**

# الخـلاصـــة

اكتسبت تقنيات التعلم الآلي واستخلاص البيانات مؤخرًا المزيد من الأهمية في مجال التشخيص الطبي، وخاصة تحليل الجينوم البشري. أحد أهم مصادر اختلاف الجينوم البشري هو تعدد أشكال النيوكليوتيدات المنفردة Single Nucleotide Polymorphisms (SNPs)، والمرتبطة بأمراض بشرية متعددة. تم تطوير العديد من التقنيات للتمييز بين العينات المتأثرة والعينات الصحية لبيانات SNP. إذ يعد تحقيق دقة تصنيف عالية في مثل هذا الفضاء عالي الأبعاد أمرًا بالغ الأهمية للتشخيص والعلاج الناجح.

تم في هذا البحث إقتراح طريقة لإستخلاص الميزات المطلوبة لـSNPs من خلال اعتماد نواة متعددة الحدود في "خوارزمية تحليل المكون الأساسي Kernel Principle Component Analysis KPCA لغرض التمييز، وكذلك دمجها مع "خوارزمية البحث الذكائية أمثلة سرب الجسيمات Particle Swarm Optimization PSO"، وتم تسميتها PSO_KPCA. طبقت الطريقة المقترحة على عشرة بيانات تجريبية حقيقة من موقع HapMap العالمي، ووجدت أن الطريقة المقترحة حققت دقة بنسبة 98.22% في حالة تمييز الـSNPs المصابة، وتم تقدم الطريقة المقترحة في نموذج برمجي متكامل يتوافر على GUI ومكتوب بلغة Matlab®2019a، يقوم بقراءة ملف البيانات الذي يكون بامتداد hmp.*، فضلاً عن إجراء عدد من التحليلات الإستكشافية الشاملة للبيانات على مستوى الجينوم.

University of Mosul
Colloge for Computer Science
and Mathematics

# A Proposed Hybrid Intelligent Model for Analyzing Genetic Deviations

**Layla Hadi Jwear**

**MSc. \ Thesis**

**Computational\ Mathematics**

**Supervised By**

**Dr. Shatha A. M Ramadhan**

**Lecturer**

---

**2021 A.M.** **1443 A.H.**

## Abstract

Machine learning and data mining techniques have recently gained more and more importance in the field of medical diagnosis, especially human genome analysis. One of the most important sources of human genome variation is single nucleotide polymorphisms (SNPs), which are associated with multiple human diseases. Several techniques have been developed to distinguish between affected and healthy samples for SNP data. Achieving high classification accuracy in such a high-dimensional space is critical for successful diagnosis and treatment.

In this thesis, a method is proposed to extract the required features for SNPs by adopting a polynomial kernel in the "Kernel Principle Component Analysis KPCA" algorithm for the purpose of discrimination, as well as combining it with the "Particle Swarm Optimization PSO" intelligent search algorithm called PSO_KPCA . The proposed method was applied to ten real experimental data from the global HapMap website, and it was found that the proposed method achieved an accuracy of 98.22% in the case of identifying the affected SNPs. It has a *.hmp extension, as well as a number of exploratory analyzes of the genome-wide data.