



جامعة الموصل  
كلية علوم الحاسوب والرياضيات

# توظيف نموذج ماركوف المخفي في تمييز الجينات المشفرة للبروتينات

سرى محمد جمال الدين حسين العلوي

رسالة ماجستير  
الإحصاء

بإشراف  
الأستاذ المساعد  
د. مثنى صبحي سليمان

2022م

1443هـ

## المستخلص

تضمنت الرسالة توظيف نموذج ماركوف المخفي في دراسة وتحليل بيانات المعلوماتية الحيوية في تمييز 13 جيناً مسؤولاً عن تشفير بروتينات السلسلة التنفسية في الميتوكوندريا وتحديد نوعية القاعدة النروجينية المستبدلة للجين MT-ND5 للإنسان والفئران، والتي تم الحصول عليها من موقع NCBI ضمن قاعدة بيانات Data Base في مراكز عالمية متخصصة في الهندسة الوراثية ودراسة عمل الجينات، فضلاً عن التطرق الى المسائل الأساسية لهذا النموذج. ان اجراء التحليل يدوياً لمثل هذه التطبيقات يكون شاقاً جداً مقارنة مع خوارزميات نموذج ماركوف المخفي، مثل خوارزمية Viterbi، التي أعطت نتائج جيدة بنسبة تطابق عالية. وقد تم تصميم برنامج حاسوبي لهذا الغرض وتمت برمجة الخوارزمية باللغة البرمجية MATLAB R2017b. ومن نتائج التطبيق العملي على سلسلة الحامض النووي الرايبي منقوص الاوكسجين DNA وباستخدام الخوارزمية المقترحة للتمييز ولتحديد نوعية القاعدة النروجينية المستبدلة، تم التوصل الى أن نموذج ماركوف المخفي يعد نهج قوي بشكل خاص لتحديد نسبة التطابق تصل إلى دقة تصنيف عالية.

**University of Mosul  
College of Computer Sciences  
And Mathematics**



# **Employing Hidden Markov Model in Recognizing Protein - Coding Genes**

**Sura Mohammed Jamal Aldeen Hussain  
Alalwlia**

**M.Sc Thesis  
Statistics**

**Supervised by**

**Assistant Professor**

**Dr. Muthanna Subhi Sulaiman**

**2022 A.D.**

**1443 A.H.**

## **Abstract**

Included in this thesis are the Hidden Markov model in the study and analysis of Bioinformatics data in distinguishing 13 genes responsible for encoding respiratory chain proteins in mitochondria and determining the quality of the nitrogenous substitution of the MT-ND5 gene for humans and mice, which was obtained from the NCBI website within the database base in global centers specialized in genetic engineering and the study of the work of genes, as well as addressing the basic issues of this model. Manual analysis of such applications is very tedious compared to Hidden Markov model algorithms, such as Viterbi algorithm, which gave accurate results with a high matching rate. A computer program was designed for this purpose and the algorithm was programmed in MATLAB R2017b. From the results of the practical application on the DNA sequencing and using the proposed algorithm to distinguish and to determine the quality of the substituent nitrogenous base, it was concluded that the Hidden Markov model is a particularly powerful approach to determine the matching ratio up to high classification accuracy.