وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم الرياضيات

التنبؤ بالتركيب الثانوي للبروتين من خلال الدمج بين نموذج ماركوف المخفي والشبكات العصبية الاصطناعية

رسالة مقدمة
إلى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة ماجستير علوم في
الرياضيات/ الرياضيات الحاسوبية

من قبل
الاء عبدالرزاق مهدي صالح

بإشراف
أ. د .عمر صابر قاسم يحيى م.د. فاطمة محمود حسن

١٤٤٥ هـ ٢٠٢٣ م

المستخلص

يعد تحليل بنية البروتين أمرًا بالغ الأهمية في فهم الترتيبات المعقدة للهياكل ثلاثية الأبعاد للبروتينات، يعتمد التحديد الدقيق للهياكل الثلاثية بشكل كبير على معرفة الهياكل الثانوية. في هذه الرسالة، تم تقديم خوارزمية مقترحة Combining Hidden Neural (CHN) تعتمد على دمج نموذج ماركوف المخفي Hidden Markov Models (HMM) والشبكات العصبية الاصطناعية Artificial Neural Networks (ANN) للتنبؤ بالبنية الثانوية للبروتين Protein Secondary Structure Predict (PSSP). تتضمن آلية المعالجة للطريقة المقترحة CHN الجمع بين مصفوفة الانبعاث Emission Matrix التابعة لنموذج HMM، مع مخططات الترميز الخاصة بالشبكة العصبية الاصطناعية ANN لتوليد خصائص مشتركة بين HMM وANN، وبالتالي تعزيز عملية التدريب، حيث تم معالجة القيم الصفرية في مصفوفة الانبعاث مسبقا من خلال مفاهيم المنطق الضبابي Fuzzy Logic. تكشف التجارب التي تم اجرائها أن خوارزمية CHN تتفوق على الأساليب التقليدية، وتظهر مرونة فائقة في التعامل مع مجموعات البيانات المتنوعة، مع تحقيق دقة ملحوظة، عند تقييمها باستخدام مجموعة بيانات RS126 مقارنة ببعض الدراسات السابقة. تسلط هذه النتائج الضوء على التطورات المهمة التي تقدمها خوارزميةCHN Combining Hidden Neural في مجال التنبؤ ببنية البروتين، مما يساهم في فهم أعمق لهياكل البروتين والوظائف المرتبطة بها.

Ministry of Higher Education
and
Scientific Research
University of Mosul
College of Computer Science
and Mathematics
Department of Mathematics

Predication of secondary strustures of protein by combining Hidden Markov Model and Artificial neural networks

A Thesis Submitted to the Council of the College of Computer Science and Mathematics University of Mosul as a Partial Fulfillment of Requirements for the Degree of Master of Science in Mathematics/Computational Mathematics

By

Alaa. A. Razzaq Mahdi Saleh

Supervised by

Assistant Professor

Dr Omar .S. Qasim **Dr .Fatima .M. Hasan**

2023 A.D. **1445 A.H.**

Abstract

Protein structure analysis is of utmost importance in understanding the intricate arrangements of the three-dimensional structures of proteins. Accurate determination of the tertiary structures heavily relies on knowledge of the secondary structures. In this master's thesis, a proposed algorithm called Combining Hidden Neural (CHN) is presented, which combines Hidden Markov Models (HMM) and Artificial Neural Networks (ANN) to predict the protein's secondary structure (PSSP). The processing mechanism of the CHN method involves merging the emission matrix of the HMM model with the encoding schemes of the ANN to generate common features between HMM and ANN, thereby enhancing the training process. Zero values in the emission matrix are preprocessed using Fuzzy Logic concepts. Experimental results demonstrate that the CHN algorithm outperforms traditional methods and exhibits remarkable flexibility in handling diverse datasets, when applied to the Intersect-Pieces and Rots-Sander datasets. Furthermore, the algorithm achieves a prediction when evaluated using the RS126 dataset, surpassing some previous studies. These results shed light on the significant advancements offered by the CHN algorithm in the field of protein structure prediction, contributing to a deeper understanding of protein structures and their associated functions.