



التنبؤ بقابلية وصول المحل للبروتين

جامعة دمشق
كلية الهندسة الميكانيكية والكهربائية
قسم الهندسة الطبية

رسالة أعدت لنيل درجة الماجستير في الهندسة الطبية-

إعداد: م. ريتا نايف شاهين

بإشراف: د.م. هاني عماشة و د. مجد الجمالي

الاستنتاجات:

• تشير نتائج الدراسة إلى فعالية استخدام الأنظمة العصبونية العائمة القابلة للتكيف ANFIS لحل إحدى أهم المشاكل في المعلوماتية الحيوية ألا وهي التنبؤ بقابلية الوصول للمحل.

• وصلت دقة التنبؤ المسجلة للنظام المقترح والمعتمد على هذه التقنية إلى %71 وهي نتيجة تبين الأداء الجيد بالمقارنة مع باقي نظم التنبؤ بقابلية الوصول للمحل الموجودة حالياً.

• تأثير الطريقة المستخدمة للحصول على الخرج على دقة النظام. فقد أظهرت نتيجة المقارنة التي أجريناها تحسناً كبيراً في الدقة بمقدار %3 لدى تغيير طريقة الحصول على الخرج، وبالتالي توصلنا لدقة تنبؤ تساوي %74.

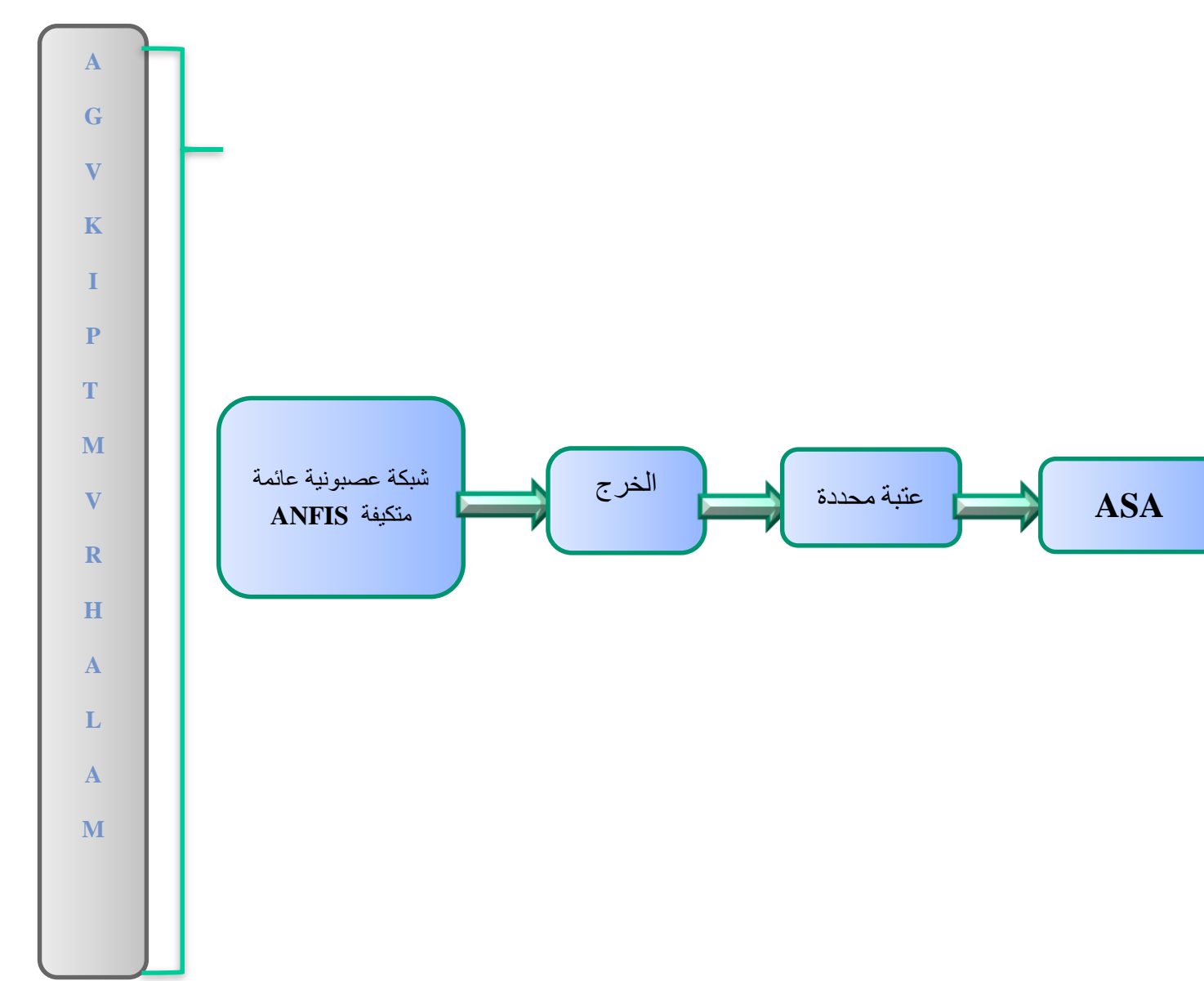
ما قمنا به في هذا البحث وأنجزناه يتيح لنا العمل في المستقبل على تطوير عملية التنبؤ بقابلية الوصول للمحل وزيادة دقتها عن طريق تطوير النظام من ناحيتين:

• الأولى هي العمل على استخدام معلومات بيولوجية إضافية لترميز الحموض الأمينية في سلسلة بروتين، منها المعلومات عن كل حمض أميني سواء كانت القطبية أو غيرها من الخصائص الفيزيوكيميائية...، أو باستخدام تشكيلات البروتين Profiles، أو أي طريقة ترميز أخرى.

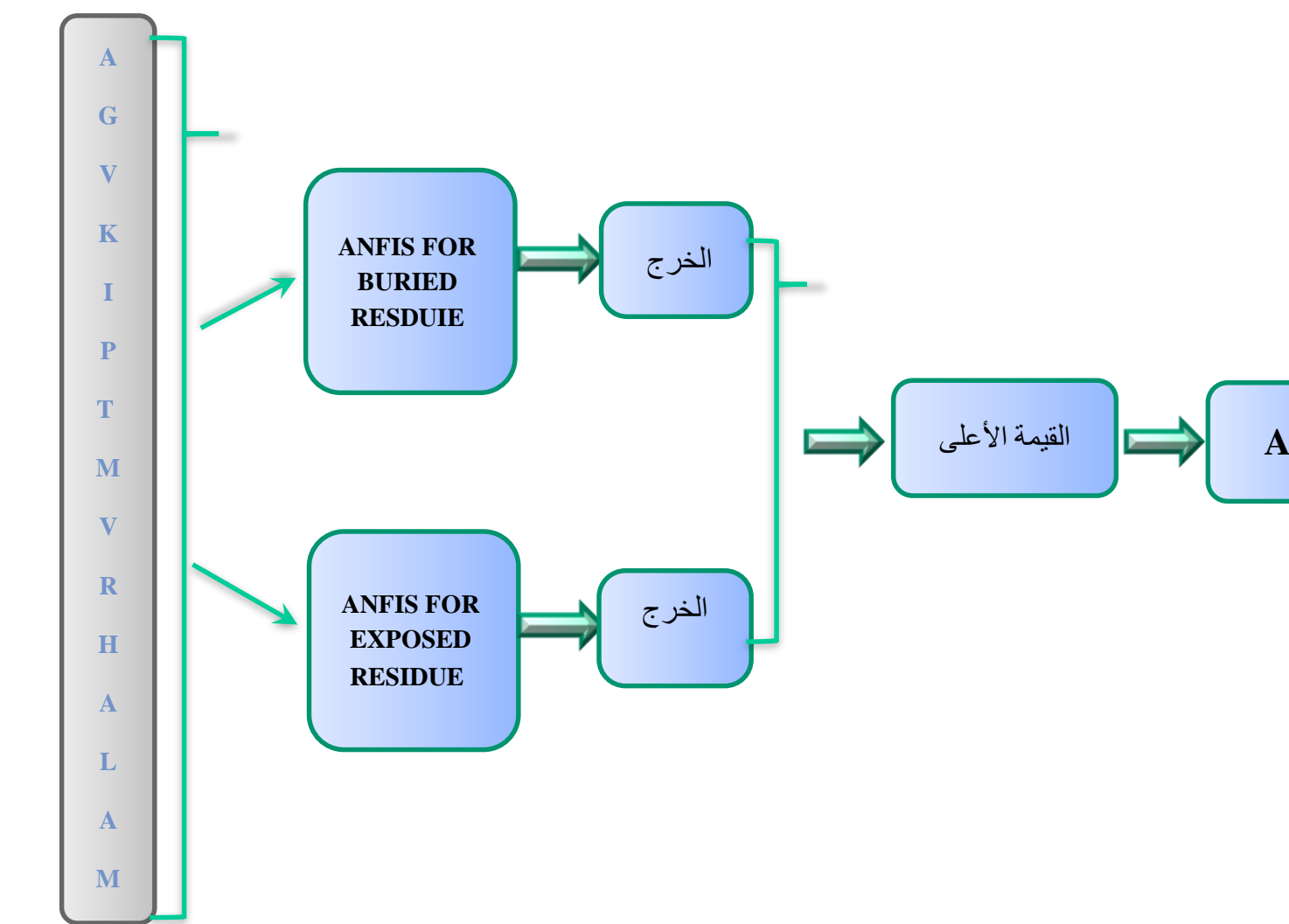
• أما الناحية الثانية فستكون بالعمل على تطوير التقنية المستخدمة للتنبؤ، أو اعتماد تقنية أخرى من تقنيات الذكاء الصناعي لتحقيق هذه الوظيفة، والوصول إلى دقة تنبؤ أعلى.

المراجع

1. Minh N. Nguyen, Jagath C. Rajapakse. (2007) "Two-Stage SVR Approach for Predicting Accessible Surface Areas of Amino Acids". *Proteins*, Volume 63, Issue 3
2. James R. A. Hutchins, (2014 Apr 15) "What's that gene (or protein)? Online resources for exploring functions of genes, transcripts, and proteins". *Mol Biol Cell*. 25(8):pp1187-120.
3. Srayania Mukherjee, Yang Zhang. (13 July 2011) "Protein-Protein Complex Structure Predictions by Multimeric Threading and Template Recombination". *Structure*, Volume 19, Issue 7, pp955-966.
4. Raif MF, Ahmad S, Zheng R, Mohamed. (2005) "Solvent accessibility in native and isolated domain environments: general features and implications to interface predictability". *Biophys Chem*.
5. Ralph A Wheeler. (2011) "Annual reports in computational chemistry". Volume 7. Amsterdam; Boston: Elsevier. ISBN: 9780444543028 0444543023
6. Kersse K, Verspurten J, Vanden Berghe T, Vandenabeele P. (2011) "The death-fold superfamily of homotypic interaction motifs". *Trends in biochemical sciences*. 36(10):541-52
7. Nickolay A. Khazanov, Heather A. Carlson. (2013) "Exploring the Composition of Protein-Ligand Binding Sites on a Large Scale". *PLoS Computational Biology*.
8. Chao Fan, Diwei Liu, Rui Huang, Zhigang Chen and Lei Deng. (2015) "PredRSA: a gradient boosted regression trees approach for predicting protein solvent accessibility". *BMC Bioinformatics*.
9. Thompson MJ, Goldstein RA. (1996) "Predicting solvent accessibility: higher accuracy using bayesian statistics and optimized residue substitution classes". *Proteins* 25:38-47.
10. Gianluca Pollastri, Alberto JM Martin, Catherine Mooney and Alessandro Vullo. (2007) "Accurate prediction of protein secondary structure and solvent accessibility by consensus combiners of sequence and structure information". *BMC Bioinformatics*.
11. Ahmad S, Gromiha MM. (2002) "NETASA: neural network based prediction of solvent accessibility". *Bioinformatics*. 18(6):819-24.
12. Won, Kyoung Jae, Hanleyck, Thomas, Prugel-Bennett, Adam and Krogh, Anders. (2005) "Evolving Hidden Markov Models for Protein Secondary Structure Prediction". *Edinburgh, IEEE Congress on Evolutionary Computation*, pp. 33-40.
13. Adamczak R, Porolito A, Meller J. (2004) Accurate prediction of solvent accessibility using neural networks based regression. *Proteins*.
14. Ghafari., Alireza Meshkin and Hossein. (2010) "PREDICTION OF RELATIVE SOLVENT ACCESSIBILITY BY SUPPORT VECTOR REGRESSION AND BEST-FIRST METHOD". *EXCLI Journal* . 9:29-38, ISSN 1611-2156



الشكل (1)



الشكل (2)

الجدول (1)

العتبة %	Qb	Qe	Q
5	57.0008	73.2887	57.9885
9	57.6236	73.0185	58.8594
10	57.8851	73.2817	59.2293
16	59.0787	72.6674	60.7544
20	60.0874	71.9871	61.9140
25	61.4761	71.1256	63.3443
36	65.6541	69.0928	66.6922
50	71.6487	65.2591	70.7948
60	75.7739	62.5398	68.6550
70	79.2787	59.5596	67.0351
80	81.8041	55.7499	63.3218
90	83.7613	51.7657	58.0066

الجدول (2)

مجموعة الاختبار	النظام الأول (بوجود عتبة)	النظام الثاني (بدون وجود عتبة)
R126	70.7 %	73.9 %

الجدول (3)

مجموعة الاختبار	النظام الأول (بوجود عتبة)	النظام الثاني (بدون وجود عتبة)
TRAIN	80.1%	90.9%

منهجية البحث:

تم بناء نظامين مختلفين للتنبؤ بقابلية الوصول للمحل متشابهين في البنية ومختلفين في آلية الحصول على الخرج وذلك لدراسة تأثير آلية الحصول على الخرج على دقة التنبؤ.

المرحلة الأولى: تصميم وبناء نظام برمجي للتنبؤ بقابلية الوصول للمحل باستخدام الأنظمة العصبونية العائمة القابلة للتكيف ANFIS الشكل (1).

المرحلة الثانية: بناء نظام آخر للتنبؤ بقابلية الوصول للمحل باستخدام الشبكات العصبونية العائمة المتكيفة أيضاً مع تغير بسيط في مكونات النظام حيث يتألف النظام الثاني من وحدتي ANFIS بدلاً من وحدة واحدة كما في النظام الأول الشكل (2).

دُرِب كل من النظامين باستخدام مجموعة التدريب المختارة وتمّ اختبارهما باستخدام مجموعة الاختبار R126 ثم رُمزت بيانات الدخل في كلا النظامين باستخدام قيم الهيدروفوبية للحموض الأمينية.

يتمّ اختيار عتبة محددة في النظام الأول واعتبار الحموض الأمينية ذات القيمة الأعلى من هذه العتبة على أنها حموض أمينية مغمورة، والأقل من هذه العتبة على أنها حموض أمينية سطحية.

يتمّ في النظام الثاني إيجاد أعلى قيمة بين خرجي الـ ANFIS الاثنان واعتبار التصنيف المقابل لهذه القيمة هو خرج النظام، وبالتالي تم الاستغناء عن العتبة.

النتائج:

✓ يبين الجدول (1) دقة النظام الأول، حيث نلاحظ أن تزايد دقة الكشف بالحموض الأمينية السطحية مترافق بتناقص دقة الكشف عن الحموض الأمينية المغمورة، حيث تقوم العتبة بتقسيم الحموض الأمينية إلى معرضة أو مغمورة تبعاً لقيمة محددة، والتي تكون غالباً غير واضحة في بنية البروتين الحقيقية، الأمر الذي يقيد أداء مصنف قابلية الوصول للمحل على عتبة مختارة بشكل عشوائي، وبالتالي سعينا في النظام الثاني المقترح على إلغاء تأثير العتبة.

✓ وصلت دقة النظام الثاني للتنبؤ بقابلية وصول المحل للبروتينات إلى %73.9، وبالمقارنة بين الدقة المسجلة لدى اختبار النظام الأول المقترح باستخدام مجموعة الاختبار R126، ولدى اختبار النظام المقترح الثاني باستخدام مجموعة الاختبار نفسها نلاحظ ازدياد دقة التنبؤ فالعتبة تؤثر بشكل واضح على دقة التنبؤ الجدول (2).

✓ لدراسة تأثير التشابه بين مجموعة الاختبار ومجموعة التدريب على دقة النتائج قمنا باختبار كل من النظامين باستخدام مجموعة جزئية من مجموعة التدريب TRAIN، فازدادت دقة التنبؤ عند استعمال المجموعة الجزئية. ويعود هذا الازدياد إلى التماثل بين سلاسل مجموعة البيانات TRAIN الذي لا يزيد عن %50، بينما لا يزيد التماثل بين سلاسل المجموعة R126 عن %25. ومنه نرى جلياً تأثير التماثل بين سلاسل مجموعة التدريب ومجموعة الاختبار على دقة النظام في التنبؤ بقابلية وصول المحل للبروتينات.