

استخدام الأسلوب الهجين MLR-GA للتكهن ببيانات التلوث الجوي

أسامة بشير شكر الحنون**
drosamahannon@gmail.com

فنار عبد الرزاق محمد نجيب*
asbasb9980@gmail.com

المستخلص

ان نمذجة جودة الهواء اكتسبت أهمية كبيرة في تلوث الهواء الجوي بسبب الآثار السلبية على البيئة وصحة الإنسان. في هذه الدراسة تم التطرق الى ملوثات الهواء التي كان لها تأثير مباشر على بيانات PM₁₀ في الهواء. تم استخدام تسعة متغيرات التفسيرية من بيانات الارصاد الجوية على مدى ثلاث سنوات في تطبيق نماذج الانحدار الخطي المتعدد. حيث كان تحليل البيانات لهذه الفترة صعب التكهن مما ادى الى وجوب النظر في البيانات على انها بيانات ذات فصول موسمية و بذلك تم استخدام اسلوب التراصف الزمني (TS) Time-Stratified. في هذه الدراسة تم اقتراح استخدام نموذج الانحدار الخطي المتعدد (MLR) Multiple Linear Regression من الطرق الأكثر شيوعا لدراسة مثل هكذا عدد كبير من المتغيرات كما وتم اقتراح استخدام الخوارزمية الجينية (GA) Genetic Algorithm لتقليل عدد المتغيرات مما يؤدي الى نتائج اكثر دقة. وكذلك اقترح استخدام الطريقة الهجينة MLR-GA. ومن خلال تقليل عدد المتغيرات التفسيرية فقد حسنت الخوارزمية الجينية أداء MLR عبر الطريقة الهجينة MLR-GA.

الكلمات المفتاحية: Multiple Linear Regression, Genetic algorithm, Forecasting.

This is an open access article under the CC BY 4.0 license

<http://creativecommons.org/licenses/by/4.0/>

Using the Hybrid MLR-GA Approach for Air Pollution Forecasting ABSTRACT

Air Quality Modeling gained great importance in atmospheric pollution because of its negative effects on the environment and human health. In our study, the relationship between (Particulate Matter PM₁₀) and other nine variables over three years is studied to applied the multiple linear regression models. The seasonal influences for seasonally periods lead to difficult analyzing and forecasting. Therefore, Time-stratified (TS) approach is used into seasonally. The multiple linear regression

* باحثة / المديرية العامة للتربية في محافظة نينوى

** مدرس / قسم الاحصاء والمعلوماتية / كلية علوم الحاسوب والرياضيات / جامعة الموصل

(MLR) model is the most common for studying like this number of variables. Genetic algorithm (GA) as well as their hybrid method such as MLR-GA, is proposed to reduce the number of studied variables. Reducing the number of variables may also lead to more accurate results. The genetic algorithm has improved the performance of MLR method separately. GA also improved MLR performance by using hybrid method MLR-GA.

1- المقدمة Introduction

يعرف نموذج MLR بأنه احد الاساليب الاحصائية للتحقق من العلاقة بين المتغيرات، ويستخدم لتشكيل نموذج للعلاقة بين المتغير المعتمد والعديد من المتغيرات التفسيرية وفي الأرصاد الجوية، فعادة ما تكون بيانات الارصاد الجوية متعددة المتغيرات وكبيرة الحجم ان كانت يومية أو أسبوعية أو أنماط موسمية اخرى، او بيانات متعددة المحطات مما يؤدي الى صعوبة التكهن بالبيانات المستقبلية (Güler and Güneri İşçi, 2016). فضلا عن استخدام الأنماط الموسمية أو الأسبوعية ومواجهة الآثار السلبية على البيانات ونتائج التحليل.

واجه الكثير من الباحثين في الإحصاء صعوبة في استخدام نماذج الانحدار المتعدد، فتم اللجوء الى استخدام بعض الطرق للحد من عدد المتغيرات التفسيرية والوصول إلى أفضل الحلول، فاستخدموا عددا من الطرق منها التقليدية مثل تحليل المكونات الرئيسية وتحليل السلاسل الزمنية، واخرى حديثة مثل اسلوب الذكاء الاصطناعي كالشبكات العصبية الاصطناعية Artificial Neural Networks (ANN) و الطرق الهجينة (GA-ANN) فضلا عن الأنظمة المضبية (Fuzzy) وبعض النماذج الهجينة المستخدمة لتحسين أداء نموذج الانحدار الخطي المتعدد وفي الواقع كان الأداء أفضل عند استخدام الأساليب الحديثة بالمقارنة مع استخدام الطرق التقليدية، كان استخدام الخوارزمية الجينية مع الشبكات العصبية (GA-ANN) كان أفضل من أداء GA و ANN كلا على حدى كما في (Zhao *et al.*, 2010) الا ان استخدام تقنية الذكاء الاصطناعي ادى الى تحسين اكثر في النتائج لأن الأساليب التقليدية قد لا توفر حساب دقيق للبيانات.

في دراسة بيانات الارصاد الجوية غالبا ما يستخدم اسلوب الانحدار الخطي المتعدد الذي يتكون من عدد كبير من المتغيرات التفسيرية و يؤدي ذلك الى تنبؤ غير دقيق وصعوبة في تحديد البيانات المدروسة في الماضي وفي حساب النماذج و دراستها وتحليلها، فضلا عن التأثيرات السلبية النمط الاسبوعية والموسمية.

الهدف من هذه الدراسة هو تحسين أداء الطرق التقليدية من خلال اقتراح MLR و GA وطرقها الهجينة (MLR-GA). ولتجنب الآثار السلبية للنتائج الموسمية وتحليل البيانات سيتم استخدام اسلوب الترافف الزمني لمعالجة عدد كبير من بيانات الأرصاد الجوية.

2-الطرق المقترحة.

يستعرض هذا القسم الانحدار الخطي المتعدد MLR ونماذجه، واستخدام الخوارزمية الجينية GA والتي هي احدى انواع الذكاء الاصطناعي وطريقة MLR-GA الهجينة للحصول على نتائج وتنبؤات دقيقة.

2-1: الانحدار الخطي المتعدد (Multiple Linear Regression MLR)

يعرف تحليل الانحدار بشكل عام بأنه أسلوب رياضي لتوضيح العلاقة بين المتغير المعتمد ومتغير أو متغيرات أخرى تسمى المتغيرات التفسيرية. يهتم تحليل الانحدار بوصف العلاقة بين المتغيرات على هيئة نموذج فقد يكون النموذج يحتوي على متغير توضيحي واحد فيسمى في هذه الحالة بنموذج الانحدار البسيط، أما في حالة احتواء النموذج على متغيرات توضيحية عدة (اثنين أو أكثر) فإنه يسمى بنموذج الانحدار المتعدد (Nathans *et al.*, 2012). عادة ماتظهر ظاهرة الارتباط الذاتي في أغلب الدراسات التي تأخذ شكل السلاسل الزمنية وغالبا ما تكون المشاهدات المتتالية للسلسلة الزمنية غير التفسيرية، ولذلك يجب استخدام اساليب احصائية خاصة عند دراسة مثل هذه البيانات والتكهن بقيمها في المستقبل، وتظهر هذه المشكلة في البيانات المقطعية وخاصة التي تكون في شكل مجاميع واسباب هذه المشكلة عادة تكون ل

- حذف بعض المتغيرات التفسيرية من العلاقة المدروسة.
- التشخيص غير الدقيق للعلاقة بين المتغير المعتمد والمتغيرات التفسيرية.
- قد تكون هنالك عوامل عشوائية تؤثر على القيم المتتالية للبواقي كما يحدث في حالات عدم الاستقرار والجفاف وغير ذلك من الحالات التي يمتد أثرها على مدى مشاهدات عديدة متعاقبة مما يحدث مشكلة ارتباط ذاتي بين أخطاء تلك المشاهدات
- إجراء تعديلات في البيانات او اللجوء الى تقدير قيم بعض المشاهدات اعتمادا على قيم مشاهدات اخرى.

استخدم العديد من الباحثين الانحدار الخطي المتعدد ويمكن صياغة النموذج

العام لمعادلة MLR على النحو الاتي:

$$Y = B_0 + B_1X_1 + B_2X_2 \dots \dots + B_iX_i + E \quad (1)$$

حيث Y_i هو المتغير المعتمد و X_1, X_2, \dots, X_i هي المتغيرات التفسيرية و B_0, B_1, \dots, B_i هي معاملات الثوابت او معاملات الانحدار E هو حد الخطأ لاقبل تنبؤ رقم i حيث $i=1,2,\dots,n$ ان عدد المشاهدات هو n يكون عدد المعادلات هو n (الراوي 1987) ويمكن صياغة المعادلة (1) في صورة مصفوفة كما يلي:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1x_{11} & x_{12} \dots x_{1p} \\ \dots & \dots \\ 1x_{n1} & x_{n2} \dots x_{np} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \end{bmatrix} \quad (2)$$

اذ Y حجم $(n \times 1)$ و المصفوفة X من درجة $(n \times (p+1))$ وحجم β هي $(p \times 1)$ و درجة ε هي $(i \times 1)$. ويلاحظ أن العمود الاول في مصفوفة البيانات يحتوي على قيمة الواحد الصحيح عند كل المشاهدات من (1) إلى (n) وذلك لتقدير المعامل الثابت والعمود الثاني من المصفوفة يحتوي على قيم المتغير الاول (X_1) ، وبذلك كل عمود يحتوي على قيم متغير تفسيري محدد، وباستخدام رموز المصفوفات يمكن اختصار كتابة نموذج الانحدار الخطي كما يلي:

$$Y = \beta X + \varepsilon \quad (3)$$

2-2: الخوارزمية الجينية:

تعد الخوارزمية الجينية أحد أساليب الذكاء الاصطناعي الحديثة التي تعتمد اعتمادا دقيقا على أفكار الهندسة الوراثية التي تتميز بالإنتاج المقصود للمجموعات الموروثة بهدف تكوين أفراد ذوي صفات جيدة. تعرف الخوارزميات الجينية على انها طريقة البحث عن أفضل الحلول للوصول إلى الخوارزمية وتجد الحلول لتحسين المشاكل من خلال استخدامات مستوحاة نتيجة للتطور الطبيعي للبقريات، مثل الاختيار، التصلب والطفرة (Zhao et al., 2010).

3-2: طريقة (MLR-GA) الهجينة

الخوارزميات الجينية (GA) تكمن في السماح لاستكشاف الفضاء التجريبي وهو كل شيء يسبب حدوث الطفرات، والتي تقوم على مبادئ النظرية الداروينية للتطور كأداة لتحسين العديد من

المشاكل في تصميم العلاج في الآونة الأخيرة يمكن استخدامه لتحسين نماذج MLR والتكهن بالمتغيرات التفسيرية الأكثر فعالية (Sattari and Gharagheizi, 2008).

يتم استخدام الخوارزمية الجينية لتحسين أداء الانحدار الخطي المتعدد باقتراح طريقة MLR-GA الهجينة وهي طريقة تطويرية تستخدم على نطاق واسع في العديد من المجالات. في هذه الدراسة سوف تستخدم قدرة الخوارزميات الجينية لتحسين نماذج MLR. وسيتم اجراء بعض المقارنات بين أداء MLR و MLR-GA لمعرفة مدى التحسن الحاصل في نتائج الانحدار بعد ادخال الخوارزمية الجينية واثبات ذلك التحسن من خلال الطرق الهجينة المقترحة لإدخالها واعتمادها في مجالات اخرى.

3- النتائج والمناقشة.

ان اولى الخطوات الرئيسية في عمل البحوث هي الحصول على البيانات التي تتناسب مع الجانب العملي من الدراسة من جانب ودقتها من جانب اخر وهي الركيزة الأساسية للوصول إلى النتائج التي يمكن الاعتماد عليها في أي عمل بحثي، ولذلك كانت دراسة بيانات الأرصاد الجوية في ماليزيا التي غطت ثلاث سنوات من 2013-2015. وتم اختيار البيانات الأكثر تأثيراً على بيانات (Particulate Matter PM_{10}) الذي يشير الى جسيمات المادة في الهواء هذه الجسيمات هي اشياء مثل الغبار العضوي والبكتريا المحمولة جوا وغبار البناء، و تحديد أهم الفترات في هذه السنوات الثلاث لتمثل بداية انفراجه في هذه الدراسة لتصبح مجموعات أكثر تجانسا وملائمة للدراسة. وتم اعتماد النمط الموسمي واستخدامه في طريقة التراصف الزمني، وتمت مقارنة نتائج البيانات الكلية والبيانات المتراصفة لنمط الفصول الموسمية. في نمط الفصول الموسمية المتراصفة فقد حددت أربع مجاميع متراصفة ايضا بالاعتماد على الفصول الموسمية في كل سنة ومراصفتها زمنيا مع جميع الفصول المشابهة في السلسلة الزمنية. وفي هذه الدراسة سيتم عرض نتائج التكهنتات للبيانات الكلية وكذلك لفصل موسمي واحد متراصف زمنيا تم اختياره من بين بقية الفصول الموسمية المتراصفة زمنياً وليكن الفصل الموسمي الأول. وسيتم استخدام مقاييس خطأ التكهن MAE و RMSE للطرق المستخدمة في الدراسة MLR و MLR-GA ومناقشتها، وتمثيل الاتساق بمخططات توضيحية بين سلسلة الدالة الاصلية PM_{10} وسلسلة التكهن y باستخدام الطرق المذكورة. تم تنفيذ الجانب التطبيقي من الدراسة باستخدام البرامج Minitab, MATLAB, Excel لتنفيذ الاوامر التي تخدم هذه الدراسة. حيث صممت هذه الدراسة للتنبؤ بقيم PM_{10} في الغلاف الجوي وبعض ملوثات الهواء التي تؤثر في بيانات PM_{10} وهي عدد من العناصر

العضوية وتكون صلبة او سائلة وتختلف في حجمها مثل (CO, O3, SO2, NO, NOX) وبعض المؤثرات منها درجة حرارة الهواء والرطوبة وسرعة الرياح بمقياسين مختلفين 10 m و xxm (حيث تغطي البيانات اليومية لمدة ثلاث سنوات من 1 كانون الثاني 2013 الى 31 تشرين الاول 2015 في احدى الولايات الماليزية. وسيتم حساب نماذج MLR باستخدام برنامج Minitab لإيجاد نماذج الانحدار وبتطبيق المعادلة (1) تم الحصول على النموذج التالي للبيانات الكلية.

$$PM_{10} = -B0 + B1CO - B2O3 + B3SO2 + B4NOx - B5NO + B6AT - B7H + B8WS10m - B9WSxxm \quad (4)$$

حيث ان WS10m, WSxxm يمثلان سرع الرياح بمقياسين مختلفين وأن AT تمثل درجة الحرارة وكذلك تمثل H الرطوبة النسبية. ويمثل PM_{10} المتغير y. أما مجموعة المتغيرات CO، O3، SO2، NOX، NO درجة حرارة الهواء و الرطوبة وسرعة الرياح بقياسين مختلفين فتمثل المتغيرات التفسيرية x_1, x_2, \dots, x_p على التوالي، و كذلك $\beta_0 + \beta_1 + \dots + \beta_p$ معاملات الانحدار للمتغيرات التفسيرية، و سيتم استخدام برنامج Minitab لإيجاد نماذج الانحدار الخطي المتعدد، والذي يتضمن عددا كبيرا من البيانات والمتغيرات التي يصعب حلها دون استخدام الحاسوب.

وتم استخدام جزء من البيانات الكلية لجعلها بيانات تدريب وإيجاد نماذج الانحدار لهذه البيانات ليتم إيجاد افضل النماذج والجزء الباقي للاختبار للتأكد من قوة التكهن بالنماذج. في هذه الدراسة تم اخذ بيانات التدريب بدء من 1 كانون الثاني 2013 إلى 30 نيسان 2015 والجزء المتبقي من البيانات المستخدمة للاختبار بدء من 1 ايار 2015 إلى 31 تشرين الاول ثم وجد نموذج الانحدار الخطي المتعدد للبيانات الكلية الخاصة بالتدريب كما في النموذج التالي (Dolatabadi et al., 2010).

$$PM_{10} = -152 + 110CO + 55.8O3 + 2336SO2 + 397NOX - 1866NO + 4.08AT - 0.108H + 4.25WS10m + 0.0048WSxxm \quad (5)$$

وبعد التكهن بالبيانات الخاصة بالمتغير المعتمد PM_{10} للبيانات الكلية الخاصة بالتدريب كان من الافضل ترتيب البيانات بطرق احصائية اكثر تجانسا فتم توجيه الاهتمام استخدام البيانات المترافقة بنمط الفصول الموسمية. وبإيجاد نموذج MLR للبيانات الكلية سوف يكون الهدف التالي هو البيانات المترافقة لنمط الفصول الموسمية وإثبات أن البيانات المترافقة بهذا النمط أفضل من البيانات الكلية. وتم اختيار الصل الموسمي الأول ممثلاً عن باقي الفصول الموسمية الاخرى.

ويعد اسلوب الترافف الزمني باستخدام انماط الفصول الموسمية من كل عام من أهم الأساليب الإحصائية المستخدمة من قبل العديد من الباحثين وباستخدام هذا الاسلوب الذي اتبعه العديد من

الباحثين للحصول على نتائج اكثر دقة كما اشار لها العديد من الباحثين مثل (Perez, 2012) و(Huebnerova and Michalek, 2014).

حيث تم الكشف عن اربعة انماط موسمية من خلال بيانات PM_{10} من حيث الارتفاع والانخفاض المتكرر خلال السنوات الثلاث بنفس الفترة الزمنية وحسب التأثيرات الجوية من امطار او جفاف او اعتدال، ويضم الفصل الموسمي المتراصف الاول الاشهر كانون الاول وكانون الثاني و شباط، و الفصل الموسمي المتراصف الثاني اذار و نيسان و ايار، و الفصل الموسمي المتراصف الثالث حزيران و تموز و اب، و الفصل الموسمي المتراصف الرابع ايلول و تشرين الاول و تشرين الثاني وعلى مدى ثلاث سنوات.

في هذه الدراسة سيتم استخدام التراصف الزمني لتشكيل انماط فصول موسمية متراصفة موسميا واختيار الموسم الأول ممثلا عن باقي الفصول الموسمية الأخرى. و قبل ايجاد MLR يتم ايجاد مجموعتين جزئيتين فرعيتين جديدتين من البيانات الجزء الأكبر يستخدم للتدريب والجزء الآخر للاختبار تستخدم مجموعة التدريب للحصول على أفضل النماذج و تستخدم مجموعة الاختبار للتحقق من دقة النماذج النموذج التالي يمثل افضل نموذج انحدار للبيانات المتراصفة زمنيا للموسم الأول:

$$PM_{10} = -31.4 + 8.19CO + 597O_3 + 4241SO_2 + 2253NOX - 1916NO \quad (6)$$

$$+ 1.79AT - 0.396H + 1.05WS_{10m} - 0.0498WS_{xxm}$$

و سيتم عرض نتائج MAE و RMSE لطريقة MLR في الحالتين التدريب والاختبار للبيانات الكلية، و نمط البيانات المتراصفة وفق نمط الفصل الموسمي الأول. فضلا عن نتائج MAE و RMSE لطريقة MLR-GA الهجينة والمقارنة بين النتائج التي تم الحصول عليها للبيانات الكلية والبيانات المتراصفة.

سيعرض الجدول (1) النتائج التي تم التوصل اليها لطريقة MLR للمقياسين MAE و RMSE لبيانات التدريب والاختبار.

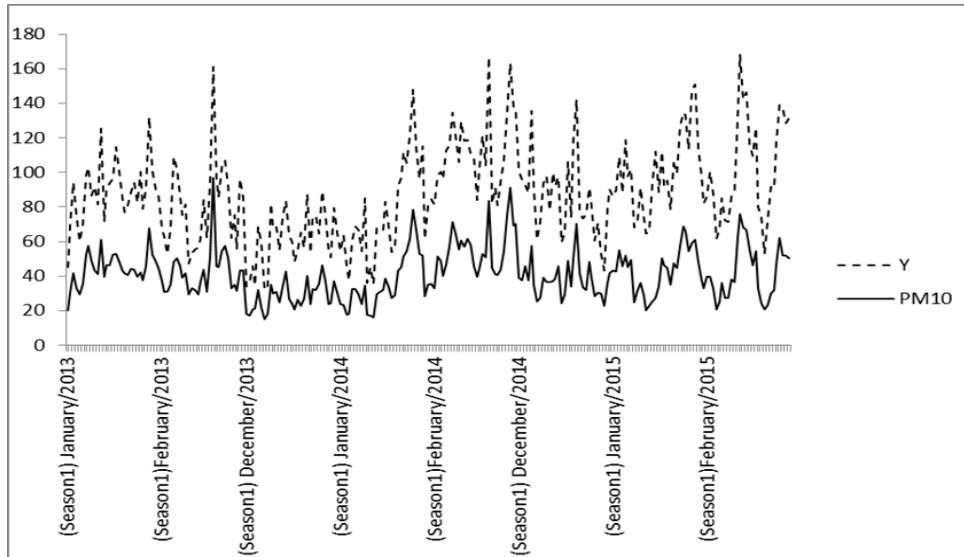
الجدول (1): قياس مفسرات الخطأ MAE, RMSE لطريقة MLR للتدريب والاختبار

Method	Training		Testing	
	MAE	RMSE	MAE	RMSE
Full data	11.3	15.3	16.3	23.3
S1	6.7	8.4	14.3	17.5

الجدول (1) يعرض MAE و RMSE لطريقة MLR للبيانات الكلية حيث كانت MAE هي 11.3 و 16.3 و RMSE هي 15.3 و 23.3 على التوالي، في حين كانت نتائج نمط الفصل الموسمي الأول المتراصف (S1) تبين ان MAE هي 6.7 و 14.3 و RMSE هي 8.4 و 17.5 لبيانات لتدريب و الاختبار على التوالي. ونلاحظ بأن هناك تحسنا في نتائج نمط الفصول الموسمية في بيانات التدريب وكذلك في الاختبار. وبذلك يمكن القول ان البيانات التي خضعت لأسلوب التراصف الزمني كانت افضل من البيانات الكلية ولذلك تم استخدام التراصف الزمني للبيانات للحصول على نتائج افضل. ولذلك تم استنتاج ان البيانات المتراصفة افضل من البيانات الكلية لطريقة MLR.

في الاشكال التالية سيتم توضيح الاتساق بين السلسلتين الاولى الاصلية PM_{10} و الثانية سلسلة التكهن Y لبيانات التدريب والاختبار للبيانات الكلية و المتراصفة زمنيا للفصل الموسمي الأول ممثلا عن بقية الفصول لطريقة MLR وطريقة MLR-GA الهجينة ورسمها باستخدام برنامج Excel.

الشكل (1) يعرض الاتساق للفصل الموسمي الأول كعينة من بيانات نمط الفصول الموسمية المتراصفة، وذلك بسبب العدد الكبير من الرسومات في المجاميع، الذي يوضح جزئي التدريب والاختبار حيث جزء التدريب الذي يبدأ من 1 كانون الثاني 2013 إلى 28 شباط 2014 وجزء الاختبار بدءا من 1 كانون الاول 2014 إلى 28 شباط 2015.



الشكل(1): الاتساق بين السلسلة الاصلية PM_{10} للنمط S1 وسلسلة التكهن Y للتدريب والاختبار باستخدام طريقة MLR.

الشكل (1) يعرض اتساقا جيدا وتجانس كبير بين السلسلة الاصلية PM_{10} و سلسلة التكهن Y باستخدام طريقة MLR للفصل الموسمي الأول من نمط الفصول الموسمية المتراففة لبيانات التدريب، والاختبار حيث PM_{10} تم تمثيلها بالسلسلة المتصلة و Y مثلت بسلسلة منقطعة لتميزهم عن بعض.

في هذه الدراسة تم استخدام GA لتحسين أداء MLR وتم التوصل للنموذج MLR-GA الهجين وبأقل الاخطاء وذلك بتوظيف الخوارزمية الجينية باستخدام البرنامج الذي تم بناؤه على اساس برنامج MATLAB لإيجاد مفسرات الخطأ التي سيتم على اساسها المقارنة بين النتائج لنماذج MLR.

ويتم كتابة برنامج بناء على برنامج MATLAB و حساب النماذج بالبداية من قبل احد البرامج الخاصة ومنهم برنامج Minitab لمجموعة البيانات بأكملها بعد ازالة المتغيرات ذات العلاقة الخطية باستخدام عامل التضخم للتباين (VIF) ويكون تحسين GA للانحدار الخطي المتعدد من خلال الاختلاف والاختيار عن طريق تقييم دالة اللياقة وتم حساب الخطأ من خلال حساب مقاييس الخطأ (MAE) و (RMSE) (Ahmadi, 2012).

وفيما يلي شرح مبسط لطريقة صياغة الخوارزمية لطريقة MLR-GA و طريقة عملها:

1. يبدأ البرنامج بإدخال قيم المتغيرات (المعتمد، المستقلة) وتحديد حجم مصفوفة المتغيرات من قبل المستخدم وتدعى هذه العملية بإنشاء المجتمع.
2. ثم يتم حساب دالة اللياقة البدنية (Fitness Function).
3. وبعدها يتم ايجاد \hat{y} التي تقابل fval في البرنامج.
4. ثم يتم حساب error لتقييم دالة اللياقة البدنية.
5. ثم تحصل بقية العمليات كالتقاطع (التزواج) والطفرة للحصول على افضل النتائج وبأقل خطأ.
6. تعاد هذه الخطوات لتحسين اداء MLR وحساب دالة اللياقة لنمطي البيانات المتراففة.

وبعد تطبيق طريقة GA للوصول الى MLR-GA. وستعرض نتائج MAE و RMSE التي تم التوصل لها للبيانات الكلية ونمط الفصول الموسمية المتراففة والفصل الاول سيكون ممثلا عنها لطريقة MLR-GA الهجينة كما ستعرض في الجدول (2).

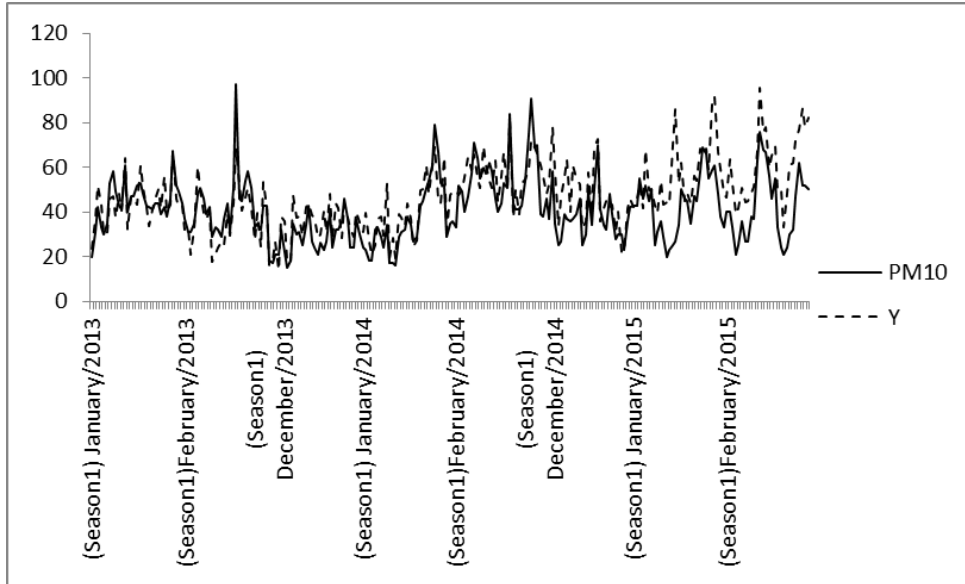
الجدول (2): قياس مفسرات الخطأ MAE, RMSE لطريقة MLR-GA للتدريب والاختبار

Method	Training		Testing	
	MAE	RMSE	MAE	RMSE
Full data	11.4	15.3	16.3	23.3

S1	6.8	8.5	14.3	17.7
----	-----	-----	------	------

الجدول (2) يبين MAE و RMSE للطريقة MLR-GA للتدريب و الاختبار للبيانات الكلية حيث كانت نتائج MAE هي 11.4 و 16.3 و كانت RMSE هي 15.3 و 23.3 على التوالي، اما في النمط و في نمط الفصول الموسمية المتراصة حيث كانت نتائج MAE في (S1) هي 6.8 و 14.3 و كانت RMSE هي 8.5 و 17.7 على التوالي. وتم ملاحظ أن هناك تحسنا في النتائج عند تطبيق اسلوب التراسف الزمني للبيانات المتراصة زمنياً. ولذلك يمكن استنتاج ان البيانات المتراصة اثبتت نتائج افضل من البيانات الكلية.

وباستخدام البيانات الناتجة عن MLR-GA سيتم رسم الاتساق للبيانات المتراصة زمنياً التدريب والاختبار، وسيتم عرض الشكل (2) للفصل الموسمي المتراصف الاول فقط كعينة لنمط الفصول الموسمية. ويتضمن الشكل (2) الجزء التدريبي الذي يبدأ من 1 كانون الثاني 2013 إلى 28 شباط 2014 وجزء الاختبار الذي بدءا من 1 كانون الاول 2014 إلى 28 شباط 2015.



الشكل (2): الاتساق بين السلسلة الاصلية PM_{10} للنمط S1 للتدريب والاختبار وسلسلة

التكهن Y باستخدام طريقة MLR-GA.

حيث يعرض الشكل (2) اتساقا واضحا وجيدا في جميع المواضع الموضحة في الشكل اعلاه بين السلسلة الاصلية PM_{10} للفصل الموسمي المتراصف الاول من نمط الفصول الموسمية المتراصة وسلسلة التكهن Y باستخدام طريقة MLR-GA لكلا البيانات للتدريب والاختبار.

بالنسبة للبيانات المتراصة فستتم المقارنة بين الطرق المستخدمة في الدراسة وهي (MLR، MLR-GA) مع الفصل الموسمي المتراصف الاول من بيانات نمط الفصول الموسمية المتراصة وتمثلها S1 وسيتم تفسيرها بالجدول (3) التالي:

الجدول(3): قياس مفسرات الخطأ MAE, RMSE لبيانات S1 في التدريب والاختبار

Methods	Training		Testing	
	MAE	RMSE	MAE	RMS E
MLR	6.7	8.4	14.3	17.5
MLR-GA	6.8	8.5	14.3	17.7

الجدول (3) يعرض نتائج MAE، RMSE لبيانات الفصل الموسمي المتراصف الاول من نمط الفصول الموسمية المتراصفة لبيانات التدريب والاختبار وكانت نتائج طريقة MLR للمقياس MAE هي 6.7 و 14.3 و قيم RMSE هي 8.4 و 17.5 على التوالي، وفي طريقة PC كانت قيم MAE هي 10.9 و 29.5 و قيم RMSE هي 14.0 و 34.1 على التوالي، اما في طريقة MLR-GA كانت نتائج MAE هي 6.8 و 14.3 و نتائج RMSE هي 8.5 و 17.7 على التوالي، وبمقارنة نتائج MLR لوحده مع طريقة MLR-GA الهجينة للتدريب والاختبار (لتقييم اداء GA) حيث كانت نتائج MLR لقيم MAE هي 6.7 و 14.3 و قيم RMSE هي 8.4 و 17.5 على التوالي، و نتائج طريقة MLR-GA للمقياس MAE هي 6.8 و 14.3 و نتائج RMSE هي 8.5 و 17.7 على التوالي، فكانت النتائج متقاربة جدا في التدريب والاختبار.

هكذا اثبتت النتائج انه لما كان MLR-GA مبني على اساس MLR فان جودته تعد واحدة من أفضل الطرق لتحسين الأداء و كذلك اثبتت النتائج للبيانات المتراصفة تحسن في نتائج MLR-GA عن MLR لوحده.

4-الاستنتاجات Conclusion

1. ان مراصفة البيانات عكس نتائج أفضل و كان له دور كبير في إظهار نتائج ادق عند تطبيق الطرق المقترحة مقارنة مع البيانات الكلية.
2. اثبت استخدام GA تحسنا ونتائج أدق عند ادخاله في نماذج التكهن الهجينة مقارنة مع استخدام الطرق التقليدية لوحدها.

المصادر

1-Güler, N. and Güneri İşçi, Ö. (2016). The Regional Prediction Model of Pm10 Concentrations for Turkey. *Atmospheric Research*, 180, 64-77.

- 2-Huebnerova, Z. and Michalek, J. (2014). Analysis of Daily Average Pm10 Predictions by Generalized Linear Models in Brno, Czech Republic. *Atmospheric Pollution Research*, 5(3), 471-476.
- 3-Nathans, L. L., Oswald, F. L. and Nimon, K. (2012). Interpreting Multiple Linear Regression: A Guidebook of Variable Importance. *Practical assessment, research & evaluation*, 17(9).
- 4-Perez, P. (2012). Combined Model for Pm10 Forecasting in a Large City. *Atmospheric environment*, 60, 271-276.
- 5-Sattari, M. and Gharagheizi, F. (2008). Prediction of Molecular Diffusivity of Pure Components into Air: A Qspr Approach. *Chemosphere*, 72(9), 1298-1302.
- 6-Zhao, H., Zhang, J., Wang, K., Bai, Z. and Liu, A. (2010). A Ga-Ann Model for Air Quality Predicting. Proceedings of the 2010 2010 *International Computer Symposium (ICS2010)*, 693-699.