

تقدير الانموذج المشترك لبيانات البقاء الطولية عالية الابعاد بأستعمال الطريقة الجينية الهجينة HGA_EM

انوار داخل هندول

دائرة التنمية البشرية، وزارة التخطيط، بغداد، العراق.

Email: anwardakhelhandool@gmail.com , ORCID: <https://orcid.org/0009-0007-3415-5536>

سهاد علي شهيد

قسم الاحصاء، كلية الإدارة والاقتصاد، الجامعة المستنصرية، بغداد، العراق.

Email: dr.suhadali@uomustansiriyah.edu.iq , ORCID: <https://orcid.org/0000-0002-1437-6469>

المستخلص

معلومات البحث

تُعدّ النماذج المشتركة من الأساليب الإحصائية المتقدمة التي تتيح تحليل البيانات الطولية وبيانات البقاء بصورة متكاملة، لا سيما عند وجود ارتباط بنيوي بين المسار الزمني للمؤشر الحيوي وخطر حدوث الحدث السريري، وفي هذا السياق، تهدف هذه الدراسة إلى تطوير تقدير أكثر كفاءة لمعلمت النمذجة المشتركة عبر اعتماد الطريقة الجينية الهجينة (HGA-EM)، التي تمزج بين قدرات الخوارزمية الجينية في البحث الشامل وخوارزمية EM في تحسين الحلول وتثبيتها، وقد طُبّق هذا الأسلوب على بيانات سريرية لمرضى التلاسيميا تضمنت ثلاثة مؤشرات طولية (Urea, MCV, HB) على مدى ثلاث زيارات زمنية.

تواريخ البحث:

التقديم: 2025 / 12 / 15

المراجعة: 2026 / 01 / 27

قبول النشر: 2026 / 02 / 02

نشر الكتروني: 2026 / 03 / 01

تسلسل الصفحات: 113 - 103

الكلمات المفتاحية:

الانموذج المشترك، دالة المخاطرة، بيانات البقاء، البيانات الطولية.

المراسلة:

أسم الباحث: انوار داخل هندول

Email:

anwardakhelhandool@gmail.com

أظهرت النتائج أن مؤشر HB سجل زيادة تدريجية عبر الزمن (0.0551)، وأن ارتفاع قيمه يرتبط بانخفاض ملموس في خطر الانتكاس وفق معامل ارتباط طولي-بقيائي سالب (-0.0823)، كما بلغ متوسط MCV نحو (77.99) مع تأثير زمني محدود (0.1205)، فيما أظهر ارتباطاً بقائياً طفيفاً خافضاً للخطر (-0.0088)، أما مؤشر Urea فكان الأكثر تذبذباً، إذ ارتفع بمرور الزمن (0.2835) وبين ارتباطاً واضحاً بزيادة احتمال وقوع الحدث من خلال معامل بقائى سالب أكبر نسبياً (-0.02745)، وتبرز هذه النتائج كفاءة الطريقة الهجينة HGA-EM في تحقيق تقديرات مستقرة وأكثر دقة مقارنة بالطرائق التقليدية، وفي تمثيل العلاقة الديناميكية بين المؤشرات الحيوية وخطر الحدث السريري بدقة عالية.

1. المقدمة

تُعدّ البيانات الطولية وبيانات البقاء من أكثر أنواع البيانات استعمالاً في الدراسات الطبية والسريرية لما توفره من فهم شامل لتطور الحالة الصحية للمريض عبر الزمن، إلى جانب تحديد زمن حدوث الأحداث السريرية المهمة مثل الانتكاس أو الوفاة، غير أن تحليل هذين النوعين من البيانات بصورة منفصلة قد يؤدي إلى تقديرات متحيزة، لا سيما عندما يكون المسار الزمني للمؤشرات الحيوية ذا ارتباط مباشر بخطر وقوع الحدث، ومن هنا برزت النماذج المشتركة (Joint Models) بوصفها إطاراً إحصائياً متقدماً يجمع بين النموذج الطولي وانموذج البقاء داخل صياغة واحدة تتيح تمثيل العلاقة الديناميكية بين تطور المؤشر الحيوي واحتمالية الحدث بشكل أدق وأكثر واقعية، وتزداد الحاجة إلى هذا النوع من النمذجة عند التعامل مع بيانات سريرية عالية الأبعاد أو ذات تذبذب واضح، كما هو الحال في متابعة مرضى التلاسيميا، حيث تتغير المؤشرات الحيوية مثل خضاب الدم (HB)، ومتوسط حجم كريات الدم الحمراء (MCV)، ويوريا الدم (Urea) عبر الزمن بطريقة تؤثر على الخطر السريري وترتبط به، إلا أن التقدير التقليدي للنموذج المشترك يواجه تحديات تقنية معقدة بسبب وجود تكاملات عالية الأبعاد ومزيج من التوزيعات الإحصائية، مما يقلل من كفاءة بعض طرائق التقدير المتعارف عليها، تعتمد هذه الدراسة الطريقة الجينية الهجينة (HGA-EM) التي تمزج بين قدرة الخوارزمية الجينية على استكشاف فضاء الحلول بشكل شامل وبين دقة خوارزمية EM في تحسين المعلمت، بهدف الوصول إلى تقدير أكثر استقراراً وكفاءة للنموذج المشترك، وقد جرى تطبيق الأسلوب المقترح على بيانات حقيقية لمرضى التلاسيميا بهدف تقييم العلاقة بين المسارات الطولية للمؤشرات الحيوية واحتمال وقوع الحدث السريري، وتحديد المتغيرات الأكثر تأثيراً في الخطر.

2. مشكلة البحث

تواجه الدراسات الطبية والسريرية التي تتضمن بيانات طولية (قياسات متكررة خلال فترة زمنية) وبيانات البقاء تحدياً من دمج هذين النوعين من البيانات ضمن إطار إحصائي واحد يراعي فيه الارتباط بين المسار الزمني للمتغيرات الطولية واحتمالية وقوع حدث ما (وفاة، انتكاسة) من أجل الوصول إلى نتائج موثوقة، لذا فإن النماذج المشتركة (Joint Model) تمكن من الجمع بين معلومات البقاء (وقت حدوث الحدث) ومعلومات القياسات الطولية مما يعطي رؤية متكاملة عن حالة المريض خلال فترة المراقبة (الدراسة).

3. هدف البحث

تهدف هذه الدراسة إلى دمج العملية الطولية وعملية البقاء في أنموذج مشترك يعكس الارتباط الحقيقي بين تطور المؤشر الحيوي واحتمال حدوث الحدث، وذلك لتحسين دقة التقدير والتنبؤ مقارنة بالتحليل المنفصل، ويرتكز الهدف على ربط انموذج التأثيرات المختلطة الخطية لتمثيل البيانات الطولية مع انموذج كوكس للخطورة النسبية عبر تأثيرات عشوائية مشتركة توضح الاعتمادية بين المسارين، فضلاً عن بناء دالة إيمان مشتركة وتقدير معالم أنموذج باستعمال الطريقة الجينية الهجينة HGA_EM.

4. الدراسات السابقة

يستعرض هذا الجزء مراجعة تحليلية لأبرز الدراسات التي أسهمت في تطوير الأنموذج المشترك وطرائق تقديره، من حيث المفاهيم، والمنهجيات، والتطبيقات وكمايلي:

في عام (2021) درست الباحثة (Alkhatami) [1]، تحليل البيانات الطولية وبيانات البقاء باستعمال النماذج المشتركة بهدف فهم العلاقة بين القياسات الطولية ونتائج البقاء، المشكلة الرئيسية التي تم معالجتها هي الصعوبات الحسابية في استدلال المعالم بدقة، لحل هذه المشكلة تم استعمال طرائق الاحتمالية القصوى والاحتمالية التقريبية، ودرست تأثير التوزيع العشوائي الخاطي للجزء العشوائي، واقترحت استعمال توزيع t للتأثيرات العشوائية لزيادة الدقة، تم تطبيق النماذج على بيانات سريرية حقيقية وأظهرت النتائج أن التحليل المشترك يوفر تقديرات أكثر دقة وموثوقة.

في عام (2020) قَدَمَ، (Qiu, et al.) [2]، دراسة مقارنة بين انموذج كوكس للتناسب الخطي (Cox Proportional Hazards Model) وانموذج غابات البقاء العشوائية (RSF) للتنبؤ بتطور الورم الدبقي عالي الدرجة (High-Grade Glioma - HGG) بعد العلاج الإشعاعي بحزمة الجسيمات (PBRT)، شملت الدراسة (82) مريضاً عولجوا في مركز شنغهاي للبروتون والأيونات الثقيلة خلال المدة من يونيو 2015 إلى نوفمبر 2019، وتم تقسيم العينة إلى 80% تدريب و20% اختبار باستعمال (10) متغيرات سريرية وعلاجية.

تم تقييم أداء الانموذجين من خلال مؤشر التوافق (C-index) ودرجة بريير (Brier Score)، وأظهرت النتائج تفوق انموذج (CPH) بدقة أعلى (C = 62.9% مقابل 61.1% في RSF) ومعايرة أفضل (BS = 0.159 مقابل 0.174 في RSF)، أكدت الدراسة الاتجاه المتنامي نحو دمج خوارزميات التعلم الآلي في النمذجة الإحصائية لتحسين التنبؤ بتطور الأورام وتعزيز التحليل السريري التنبؤي.

5. الانموذج المشترك The Joint Model

تُعد البيانات الطولية (Longitudinal Data) من أكثر أنواع البيانات شيوعاً في الدراسات الطبية، إذ تمثل قياسات متكررة للمؤشرات الحيوية تسجل للمريض نفسه عبر فترات زمنية متعددة، بما يسمح بتتبع تطور حالته الصحية بمرور الزمن، أما بيانات البقاء (Survival Data) فتهدف إلى دراسة الزمن اللازم لحدث معين مثل الانتكاس أو الوفاة، مع الأخذ بنظر الاعتبار ظاهرة الرقابة (Censoring)، وإن تحليل كل نوع من هذه البيانات بصورة منفصلة قد يؤدي إلى تقديرات متحيزة، لا سيما عند وجود ارتباط بنيوي بين المسار الزمني للمؤشر الحيوي واحتمالية وقوع الحدث، ومن هنا برزت الحاجة إلى تطوير الانموذج المشترك (Joint Model) ليشكل إطاراً إحصائياً موحدًا يجمع بين تحليل البيانات الطولية وبيانات البقاء في آن واحد، حيث يربط هذا الانموذج بين انموذج طولي (عادةً خطي مختلط) وانموذج للبقاء (غالباً انموذج كوكس للخطورة النسبية) من خلال مؤثرات عشوائية أو متغيرات كامنة مشتركة، مما يسمح بفهم العلاقة الديناميكية بين سلوك المؤشر الحيوي وخطر وقوع الحدث عبر الزمن، وتكمن أهمية هذا الأنموذج في قدرته على تحسين دقة التقدير وتقليل التحيز الناتج عن المراقبة أو الانسحاب غير العشوائي، فضلاً عن تقديم تنبؤات ديناميكية دقيقة لاحتمال البقاء باستعمال المعلومات المتجددة من القياسات الطولية. [2][3]

$$L_i(\theta) = \log \prod_{i=1}^n f(Y_i|b_i) \cdot f_b(b_i) \cdot f_T(T_i\delta_i|b_i) \quad (1)$$

اذ ان: $f(Y_i|b_i)$: دالة الإمكان للانموذج الفرعي لبيانات الطولية، $f_b(b_i)$: دالة الاحتمال للتأثير العشوائي. $f_T(T_i\delta_i|b_i)$: دالة الإمكان لانموذج الفرعي لبيانات البقاء، وعند فك الحدود في معادلة (1) يصبح لدينا دالة الإمكان الأعظم للانموذج المشترك (The Joint Likelihood Function)، وكما في الصيغة الآتية:

$$L_i(\theta) = \int [\phi(Y_i(t_{ij}); x_i^T(t_{ij})\beta + b_i^T z_i(t_{ij}), \sigma_\epsilon^2)] \cdot \frac{1}{(2\pi)^{\frac{q}{2}} |\Sigma_b|^{\frac{1}{2}}} \exp\left(\frac{-1}{2} (b_i^T \Sigma_b^{-1} b_i)\right) \cdot [h_i(T_i|b_i)]^{\delta_i} \cdot \exp\left(-\int_0^{T_i} h_i(u|b_i) du\right) \quad (2)$$

اذ ان: $Y_i(t_{ij})$: متغير عشوائي يمثل القيم المشاهدة لمتغير الاستجابة الطولي عند الزمن (t_{ij}) ، أي الاستجابة للمشاهدة (i) في المقطع العرضي (j) عند الزمن (t)، θ : تمثل متجه المعالم الكلي في الانموذج (الطولي-البقاء-التأثير العشوائي).

$$\theta = \{\beta, \mu_b, \Sigma_b, \eta, \psi, b_i, \sigma_\epsilon^2\}$$

$W_i(t_{ij})$: دالة المسار للمفردة (i) عند الزمن (t_{ij}) وتمثل القيمة الحقيقية للمتغير الطولي في تلك النقطة الزمنية سواء كان لدينا مشاهدة لهذا الزمن ام لا، وتتكون دالة المسار من متغيرات ثابتة وعشوائية، $x_i^T(t_{ij})$: مصفوفة التأثيرات الثابتة وتسمى (المتغيرات المشتركة) وتشمل كل المتغيرات الثابتة التي يشترك بها الافراد مثل (العمر، الجنس، نوع العلاج)، وتقاس للمشاهدة (i) في المقطع العرضي (j) عند الزمن (t)، β : متجه المعلمات المرتبطة بالمتغيرات التأثيرات الثابتة بضمنها معلمة الحد الثابت، $Z_i(t_{ij})^T$: يمثل مصفوفة متغيرات التأثيرات العشوائية (القياسات المكررة خلال الزمن) للمشاهدة (i) في المقطع العرضي (j) عند الزمن (t)، اذ تتغير قيم تلك المتغيرات مع الوقت لكل مريض والهدف هو تمثيل الفروق الفردية بين المرضى خلال المسار الزمني للدراسة.

المتغيرات العشوائية ابعاده ($1 \times t$)، ويشير الى التغير مع الزمن يتم تمثيله بواسطة المتجه (z) وان b_i : تمثل متجه معلمات التأثيرات العشوائية اذ يرافق كل متغير عشوائي متجه عشوائي، بافتراض ان (b_i) يتوزع توزيعاً طبيعياً بمتوسط (0) وتباين Σ_b ، ويشتمل (b_i) على معلمتي الحد الثابت (b_{0i}) التي تشير الى انحراف المفردة (i) عن المتوسط العام عند قيم متغير الاستجابة الطولي قيد الدراسة، ومعلمة الميل الحدي (b_{qi}) وتشير الى معلمات التأثيرات العشوائية للمفردة (i) ويرتبط بزمن حدوث الحدث. u_i : زمن البقاء المشاهد للمفردة (i)، $\epsilon_i(t_{ij})$: حد الخطأ العشوائي ويتوزع توزيعاً طبيعياً بمتوسط صفر وتباين σ_{ϵ}^2 ، وهو يمثل الاخطاء للمفردة (i) عند نقطة الزمنية (j).

6. الانموذج الفرعي الطولي (Longitudinal Submodel)

يُستعمل هذا الانموذج لنمذجة البيانات الطولية (Longitudinal Data) التي تمثل القياسات المتكررة لوحدة الدراسة على فترات زمنية متعددة، مثل قياس المؤشرات الحيوية أو رصد استجابة المريض للعلاج خلال فترة معينة، يهدف هذا الانموذج إلى وصف التغيرات داخل الأفراد (Within-Subject Variations) وكذلك الاختلافات بين الأفراد (Between-Subject Variations)، مما يتيح فهماً دقيقاً لتطور الحالة مع مرور الوقت. وغالباً ما تُنمذج البيانات الطولية باستعمال انموذج التأثيرات المختلطة الخطية (Linear Mixed-Effects Model)، الذي يجمع بين نوعين من التأثيرات: [4]

- التأثيرات الثابتة (Fixed Effects): وتمثل الاتجاه العام للتغير عبر جميع الأفراد في الدراسة (Overall Trend).
- التأثيرات العشوائية (Random Effects): وتمثل التباين بين الأفراد أو الوحدات، أي الفروق الفردية في الاستجابة (Individual Differences).

تقدم يتم نمذجة البيانات الطولية وفق انموذج التأثير المختلط الخطي (Linear Mixed Effect Model (LME))، إذ يفسر انموذج التأثير المختلط التباين بين المشاهدات من خلال تقدير التأثيرات العشوائية الخاصة بالمشاهدات بوجود المعلمات الثابتة عبر المشاهدات (التأثيرات الثابتة)، وكما موضح سابقاً فان البيانات الطولية تتكون من قياسات متكررة كنتائج لكل مفردة (مشاهدة) خلال فترة زمنية معينة وغالباً ما تكون تلك القياسات لنفس المفردة مترابطة، بالتالي فإن المتوقع ان يكون لكل مفردة في مجتمع الدراسة نمط استجابة خاص به مع مرور الزمن. [2][5] يمكن صياغة انموذج (LME) على ضوء البيانات الطولية، كالاتي:

$$Y_i(t_{ij}) = x_i^T(t_{ij})\beta + b_i^T z_i(t_{ij}) + \epsilon_i \quad (3)$$

اذ ان: $Y_i(t_{ij})$: متغير الاستجابة وهو متغير عشوائي ويمثل القيمة المشاهدة للمتغير الطولي للمفردة (i) خلال اخذ القياس (الزيارة) (j) عند الزمن t_{ij} ، اذ ان ($i = 1, 2, \dots, n$) وان ($j = 1, 2, \dots, J_i$)، $x_i^T \beta$: يمثل مكون التأثير الثابت (Fixed Effect) اذ ان (x_i^T) متجه المتغيرات المشتركة (Common Variables) والتي تؤثر على متغير الاستجابة تأثيراً ثانوياً وليس رئيسياً. وان $x_i^T = (x_{i1}, x_{i2}, \dots, x_{ip})^T$.

اذ ان: p : عدد المتغيرات في المكون الثابت، اما (β) تمثل متجه معلمات الانحدار للمكون الثابت المرتبطة بمتجه المتغيرات المشتركة، وان $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$ ، $b_i^T z_i(t_{ij})$: يمثل مكون التأثير العشوائي (Random Effect) اذ ان ($z_i(t_{ij})$) متجه المتغير العشوائي الذي يقيس التغير المسجل بمرور الزمن للمفردة (i) خلال اخذ القياس (الزيارة) (j) عند الزمن t_{ij} ، اما (b_i) يمثل متجه المعلمات التي تلتقط التغير العشوائي على مستوى كل مفردة (i) ويرتبط مع متجه ($z_i(t_{ij})$)، وتتوزع (b_i) توزيعاً طبيعياً بمتوسط (μ_b) وتباين (Σ_b) لكل ($i = 1, 2, \dots, n$). اي ان $b_i \sim N(\mu_b, \Sigma_b)$.

مكون التأثير الثابت (Fixed Effect) يكون ثابت للعينة ككل في مجتمع الدراسة، اذ تفسر لنا المعلمة (β) تأثير متجه المتغيرات المشتركة للعينة المبحوثة ككل من خلال متوسط المتغيرات الطولية.

اما متوسط المكون العشوائي (μ_b) فيصف الخصائص الحقيقية لسلوك هذا المتغير وتغيره بمرور الزمن اي ان لكل مفردة استجابتها الخاصة، هذا الاختلاف في الاستجابة يتم التعبير عنه من خلال متجه التأثيرات العشوائية (b_i) ولكل مفردة على حدة ($i = 1, 2, \dots, n$). [4]

وتحت افتراض ان $\epsilon_i(t_{ij})$ يتبع التوزيع الطبيعي وأن حد الخطأ العشوائي مستقل بشكل تام (Mutually Independent) بالنسبة لـ (b_i) ، وعليه فإن التوزيع الشرطي لمتغير الاستجابة بوجود التأثير العشوائي $(Y_i(t_{ij})|b_i)$ هو أيضاً يتوزع توزيعاً طبيعياً وكالاتي:

$$\left. \begin{aligned} E(Y_i(t_{ij})|b_i) &= W_i(t_{ij}) \\ E(Y_i(t_{ij})|b_i) &= x_i^T(t_{ij})\beta + b_i^T z_i(t_{ij}) \\ \text{Var}(Y_i(t_{ij})|b_i) &= \sigma_\epsilon^2 \end{aligned} \right\} \quad (4)$$

7. الانموذج الفرعي للبقاء (Survival Submodel)

يختص هذا الانموذج بنمذجة زمن حدوث الحدث (Time-to-Event) استناداً إلى تحليل بيانات البقاء (Survival Data Analysis)، إذ يُراقب الوقت حتى وقوع الحدث محل الاهتمام، مثل الوفاة أو الفشل العلاجي، ومن أبرز النماذج المستعملة في هذا السياق:

- انموذج المخاطر النسبية (Proportional Hazards Model).
- انموذج وقت الفشل المتسارع (Accelerated Failure Time Model).
- ويُعد انموذج كوكس شبه المعلمي للمخاطر النسبية (Cox Proportional Hazards Semiparametric Model) الأكثر استعمالاً في التطبيقات العملية، إذ يتيح دراسة تأثير العوامل المختلفة على معدل الخطر (Hazard Rate) دون افتراض شكل محدد لتوزيع أزمنة الأحداث. [2]
- ومن اجل تهيئة الانموذج الفرعي لبيانات البقاء وملائمته مع الانموذج الفرعي للبيانات الطولية ذات التأثير المختلط فان دالة الخطورة (Hazard Function) للمفردة (i) بوجود (b_i) تكتب كالاتي:

$$h_i(t_{ij}|b_i) = h_0(t) \exp(z^T \psi + \alpha m_i(t)) \quad (5)$$

اذ ان: $h_i(t_{ij}|b_i)$: دالة الخطورة للمفردة (i) بوجود التأثير العشوائي (b_i) ، $h_0(t)$: دالة الخطورة الأساسية (Baseline Hazard)، α : معلمة ربط (Link Parameter) بين الانموذجين (الطولي والبقاء)، z^T : متجه المتغيرات المشتركة المعتمد على الزمن، ψ : معلمات المرتبطة بالمتغيرات المشتركة، $m_i(t)$: القيمة المتوقعة لمتغير الاستجابة $Y_i(t_{ij})$ في الانموذج الطولي عند الزمن (t) .
وتحسب كالاتي:

$$m_i(t) = x_i^T(t_{ij})\beta + z_i^T(t_{ij})b_i \quad (6)$$

دالة الخطورة التجميعية (Cumulative Hazard Function) يمكن صياغتها كالاتي:

$$H_i(t_{ij}|b_i) = \int_0^t h_i(u|b_i) du \quad (7)$$

دالة البقاء (Survival Likelihood Function) تكتب وفق الصيغة الاتية:

$$S_i(t_{ij}|b_i) = \exp\left(-\int_0^t h_i(u|b_i) du\right) \quad (8)$$

الدالة في معادلة (9) تستند على حدوث الحدث من عدمه لكل مفردة (i) ، بالتالي فان الدالة الاحتمالية لحدوث الحدث عند الزمن المشاهد الكلي (T_i) وللمشاهدة (i) تكتب كالاتي:

$$f_T(T_i \delta_i | b_i) = \begin{cases} h_i(T_i | b_i) \cdot \exp(-H_i(T_i | b_i)) & , \text{if } \delta_i = 1 \\ \exp(-H_i(T_i | b_i)) & , \text{if } \delta_i = 0 \end{cases} \quad (9)$$

الصيغة العامة لمعادلة (9) تكتب كالاتي:

$$f_T(T_i \delta_i | b_i) = [h_i(T_i | b_i)]^{\delta_i} \cdot \exp\left(-\int_0^{T_i} h_i(u|b_i) du\right) \quad (10)$$

ففي حال ان زمن البقاء (Survival Time) يتبع توزيع Weibull فان دالة الخطورة الأساسي هي:

$$h_0(t) = \lambda \rho t^{\rho-1}$$

بالتالي فان معادلة (5) تكتب وفق المعادلة الاتية:

$$h_i(t_{ij}|b_i) = \lambda \rho t^{\rho-1} \exp(z^T \psi + \alpha m_i(t)) \quad (11)$$

$$H_i(t_{ij}|b_i) = \int_0^{T_i} \lambda \rho u^{\rho-1} \exp(z^T \psi + \alpha m_i(u)) du \quad (12)$$

دالة الأماكن الأعظم لدالة البقاء وفق معادلة (10) تكتب كالآتي:

$$f_T(T_i \delta_i | b_i) = [\lambda \rho t^{\rho-1} \exp(z^T \psi + \alpha m_i(T_i))]^{\delta_i} \cdot \exp\left(\int_0^{T_i} \lambda \rho u^{\rho-1} \exp(z^T \psi + \alpha m_i(u)) du\right) \quad (13)$$

ويمثل (u) الحد الأعلى للتكامل والذي يساوي (T_i) أي وقت المشاهد الكلي. ان دمج الانموذجين الفرعيين في الانموذج المشترك يتم من خلال السماح لانموذج البقاء بالاعتماد على بعض خصائص الانموذج الطولي وفق طرائق ربط مختلفة، بحيث يتم نمذجة البيانات الطولية وزمن الحدث بصورة مترابطة ضمن دالة كثافة مشتركة مشروطة (Joint Conditional Density Function) وليس بشكل منفصل كالمودجين مستقلين، ويمتاز هذا الترابط بقدرته على تحليل البيانات الطولية بدقة عالية حتى زمن وقوع الحدث. [5] [6]

8. تقدير الانموذج المشترك The Joint Model Estimation

بعد ان تم دراسة مفهوم الانموذج المشترك والتعرف على جميع مكوناته، سيتم دراسة طرائق التقدير للانموذج المشترك للبيانات المشتركة بالاعتماد على معادلة الأماكن الأعظم للانموذج المشترك في معادلة (2).
قدم الباحث (Rizopoulos) عام (2012) منهجية مفصلة لتقدير الانموذج المشترك وفق طريقة تقدير الأماكن الأعظم (MLE)، فدالة الأماكن الأعظم في معادلة (1) هي: [7][8][1]

$$L_i(\theta) = \prod_{i=1}^n \int f(Y_i | b_i) \cdot f_b(b_i) \cdot f_T(T_i \delta_i | b_i) \quad (14)$$

اذ ان دالة الأماكن الأعظم للانموذج المشترك أعلاه لا تمتلك حل مغلق نتيجة للأسباب الآتية:

1. وجود التكامل متعدد الابعاد على المؤثرات العشوائية (b_i) داخل دالة الأماكن.
 2. شكل العلاقة غير خطية بين الانموذج الطولي وانموذج البقاء وبالتحديد عند الربط بين الانموذجين من خلال ($m_i(T_i)$).
 3. دالة الأماكن تتضمن دمجاً بين التوزيع الطبيعي (للانموذج الطولي) وتوزيع أسي أو كوكس أو ويبل (لانموذج البقاء)، وهذا المزيج لا يؤدي إلى شكل مغلق.
 4. لانموذج يحتوي على تكاملات زمنية (داخل دالة البقاء) مما يزيد من تعقيد الدالة.
- ونتيجة للأسباب أعلاه نلجأ إلى الطرائق العددية (Numerical Methods) لتقريب هذا التكامل، وفي دراستنا سيتم التطرق الطريقة الجينية المقترحة (HGA_EM).

1.8. الطريقة المقترحة خوارزمية التقدير الجينية الهجينة (HGA-EM)

Hybrid Genetic Algorithm – Expectation Maximization

نظراً للتعقيد البيئي للانموذج المشترك وصعوبة تقدير معالمه بسبب وجود تكاملات عالية الأبعاد للتأثيرات العشوائية، تصبح الطرائق التقليدية مثل خوارزمية EM عرضة لمشكلات الأمثلية المحلية وتعتمد بدرجة كبيرة على القيم الابتدائية للمعلمات، وللتغلب على هذه الإشكالات، تم اقتراح الخوارزمية الجينية الهجينة (HGA-EM) التي تجمع بين قوة البحث الشامل للخوارزمية الجينية GA، وبين دقة التحسين الاحتمالي لخوارزمية EM في الوصول إلى حلول أكثر استقراراً وقرابة من القيمة المثلى. ينتج هذا الدمج اعتماد قيم ابتدائية محسنة تُنتجها GA، ثم تقوم EM بصقلها عبر خطواتها التكرارية للوصول إلى أفضل تقدير ممكن للمعلمات، مما يرفع كفاءة التقدير ويقلل احتمالية الوقوع في القيم المثلى المحلية، وبالتالي يحقق تقديرات أكثر دقة للانموذج المشترك.

خطوات التقدير باستعمال الخوارزمية الجينية الهجينة (HGA-EM) كالآتي: [9][10].

المرحلة الأولى: تهيئة الحلول (المقدرات) الأولية (θ)، اذ يتم توليد P من المتجهات تمثل ($\theta^1, \theta^2, \dots, \theta^P$) بشكل عشوائي.

المرحلة الثانية: حساب دالة الهدف السابقة لكل متجه (θ^P).

المرحلة الثالثة: اختيار الافراد (المعلمات) الأفضل (Selection) والتي تحقق اقل قيمة لدالة الهدف (Objective Function).

المرحلة الرابعة: دمج المعلمات او ما يسمى التزاوج (Crossover) بين أفضل الحلول (Optimal Solution) لتكوين جيل جديد.

المرحلة الخامسة: تعديل بعض المعلمات عشوائياً وتسمى مرحلة الطفرة (Mutation) لتوسيع مساحة البحث عن الحلول المثلى في فضاء

فضاء الحلول الشامل (Global Minimum) لجميع الأجيال وتجنب القيم المثلى المحلية (Local Optima) في فضاء الحلول التابع لكل جيل على حدة.

المرحلة السادسة: تحديث الجيل الجديد من خلال استبدال الافراد (المعلمات) الأقل كفاءة بالافراد الجدد وتحسين دالة الهدف تدريجياً.

المرحلة السابعة: تنفيذ معيار التوقف (Stopping Criterion) على الخوارزمية عندما يتحقق الشرط الآتي:

$$|Objective_{k+1} - Objective_k| < \varepsilon \quad (15)$$

وبالتالي تبدأ خوارزمية EM من القيم الابتدائية الناتجة من GA

$$\theta^{(0)} = \theta_{GA}^*$$

وتكرر خطوتين رئيسيتين:

المرحلة الثامنة: الخطوة التوقعية: (E-Step)

يتم حساب القيمة المتوقعة لدالة اللوغاريتم الإمكانية الكاملة بناءً على المعلمات الحالية كما يأتي:

$$Q(\theta | \theta^{(k)}) = E_k[\log L_i(\theta) | Y_i(t_{ij}), T_i, \delta_i, b_i; \theta] \quad (16)$$

المرحلة التاسعة: الخطوة التعظيمية: (M-Step)

يتم إيجاد المعلمات التي تُعظم $Q(\theta | \theta^{(k)})$

$$\theta^{(k+1)} = \arg \max_{\theta} Q(\theta | \theta^{(k)})$$

المرحلة العاشرة: معيار التوقف: (Stopping Criterion)

تتوقف الخوارزمية عندما يتحقق الشرط الآتي:

$$\theta^{(k+1)} - \theta^{(k)} < \varepsilon \quad (17)$$

وبذلك يُعد الناتج النهائي:

$$(\hat{\theta}) = \theta^{(k)} \quad (18)$$

وهو يمثل المقدرات المثلى لمعلمات الامنودج المشترك.

بعد الحصول على أفضل متجه للمعلمات $(\hat{\theta})$ يمكن تقريب مصفوفة التباين من خلال مصفوفة Hessian التقريبية والتي يتم اشتقاقها من دالة الهدف، وكالاتي:

$$Var(\hat{\theta}) = [H(\hat{\theta})]^{-1} \quad (19)$$

9. الجانب التطبيقي

تناول الجانب التطبيقي بيانات عن مرض ثلاثيميا الدم لعينة حجمها (52) مريض لثلاث زيارات بيانات حقيقية اخذت من مستشفى مدينة الطب / بغداد وتم تقدير معلمات الامنودج المشترك باستعمال طريقة الجينية الهجينة (HGA_EM) باستعمال برنامج لغة R (الاصدار 4.5.1) لثلاث مؤشرات طولية خضاب الدم HB، ومتوسط حجم كريات الدم الحمراء MCV، ويوريا الدم Urea_Blood [11][1].

1.9. تقدير معلمات الامنودج المشترك للمؤشر الطولي (Hb)

كل امنودج طولي فرعي يرتبط بدالة الخطورة (1,0) بالتالي فان (Longitudinal Submodel)

$$y_i^{Hb}(t_{ij}) = B_{0,Hb} + B_{1,Hb}(Sex)_i + B_{2,Hb}(age)_i + B_{3,Hb}(treat.)_i + B_{4,Hb}(t_{ij}) + b_{0,Hb} + b_{1,Hb}(t_{ij}) + \epsilon_i^{Hb}(t_{ij}) \quad (20)$$

اما الامنودج البقاء الفرعي المرتبط بدالة الخطورة الطولية الفرعية لمستويات (Hb) (Survival Submodel)

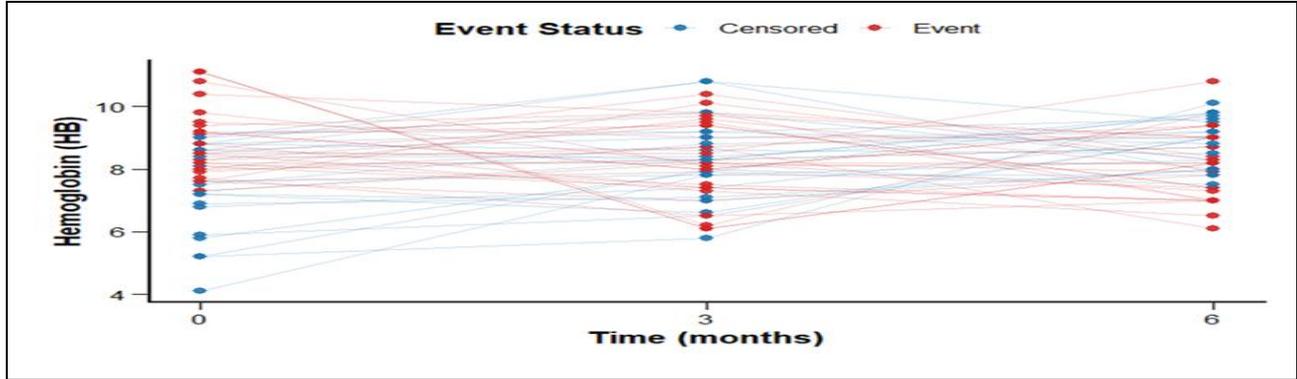
$$h_{Hb,i}(t) = h_{Hb,01} \exp\{\psi_{11,Hb}(Sex)_i + \psi_{12,Hb}(ag)_i + \psi_{13,Hb}(treat.)_i + \alpha_{Hb,1} m_{Hb,i}(t)\} \quad (21)$$

إذ ان : $\alpha_1, \psi_{11}, \psi_{12}, \psi_{13}$: تأثير المتغيرات (الجنس والعمر والعلاج و Hb) على التوالي على خطر حدوث الحدث الانتكاسة.

الجدول (1) نتائج تقدير معلمات الامنودج المشترك بالطريقة (HGA-EM) للمؤشر الطولي (Hb)

Methods	Type model	Variable	Estimation (Mean)	St-Dev	Confidence Interval 95%	
HGA-EM	Longitudinal Sub Model	(Intercept)	7.9397	0.58	(7.94, 8.68)	
		Time	0.0551	0.10	(-0.15, 0.26)	
		Age	0.0117	0.024	(-0.037, 0.061)	
		Gender	0.2085	0.21	(-0.20, 0.62)	
		Drug_Type	-0.0835	0.11	(-0.29, 0.12)	
		Age	-0.0789	0.024	(-0.127, -0.030)	
		Gender	-0.1847	0.22	(-0.62, 0.23)	
		Survival Sub Model	Drug_Type	0.1039	0.11	(-0.14, 0.34)
			value (HB) = α	-0.0823	0.01	(-0.10, -0.06)

تشير نتائج الجدول (1) إلى أن المؤشر الطولي Hb يمتلك مستوى ابتدائياً مستقرًا نسبيًا بلغ 7.94 مع فاصل ثقة ضيق، ويُظهر اتجاهًا زمنيًا موجبًا ضعيفًا (0.0551) يعكس تحسنًا تدريجيًا محدودًا في مستوى الهيموغلوبين عبر فترات المتابعة، دون دلالة إحصائية قوية، كما لم تُظهر المتغيرات الديموغرافية والعلاجية (العمر، الجنس، نوع العلاج) تأثيراً معنوياً واضحاً على مساره الطولي، في انموذج البقاء، تبين أن العمر له تأثير سلبي دال إحصائياً على خطر الحدث، بينما لم يكن للجنس ونوع العلاج تأثير جوهري، وتُعد النتيجة الأهم هي معامل الارتباط السلبي الدال إحصائياً ($\alpha = -0.0823$)، الذي يدل على أن ارتفاع القيم المتوقعة لمؤشر Hb يرتبط بانخفاض ملموس في خطر حدوث الحدث، مما يؤكد الدور الوقائي للهيموغلوبين وأهميته كمؤشر حيوي تنبؤي، ويعزز كفاءة استعمال الانموذج المشترك وطريقة HGA-EM في توصيف العلاقة البنوية بين التطور الطولي للمؤشر واحتمالية البقاء.



الشكل (1) تطور مستوى المؤشر الطولي (Hb) حسب حالة حدوث الحدث

يوضح الشكل (1) المسارات الزمنية لمستوى الهيموغلوبين (Hb) للمرضى عبر ثلاث نقاط زمنية (0، 3، 6 أشهر)، مع التمييز بين المرضى الذين تعرضوا للحدث (باللون الأحمر) والمرضى تحت المراقبة (باللون الأزرق)، يُلاحظ وجود تباين فردي واضح في مستويات Hb ومساراتها عبر الزمن، وهو ما يعكس اختلاف استجابة المرضى للعلاج، كما يُظهر الشكل أن المرضى الذين تعرضوا للحدث يميلون إلى امتلاك قيم Hb أقل أو مسارات غير مستقرة مقارنة بالمرضى الرقابيين الذين غالبًا ما يحافظون على مستويات أعلى أو أكثر استقرارًا بمرور الزمن، هذا النمط البصري يدعم فرضية الانموذج المشترك بوجود علاقة بنوية بين المسار الطولي لـ Hb وخطر حدوث الحدث، حيث يرتبط انخفاض Hb بزيادة احتمال الحدث، وهو ما يتسق مع نتائج معامل الارتباط السلبي المقدر في الانموذج المشترك.

2.9 تقدير معلمات الانموذج المشترك للمؤشر الطولي (MCV)

$$y_i^{MCV}(t_{ij}) = B_{0,MCV} + B_{1,MCV}(Sex)_i + B_{2,MCV}(age)_i + B_{3,MCV}(treat.)_i + B_{4,MCV}(t_{ij}) + b_{0,MCV} + b_{1,MCV}(t_{ij}) + \epsilon_i^{MCV}(t_{ij}) \quad (22)$$

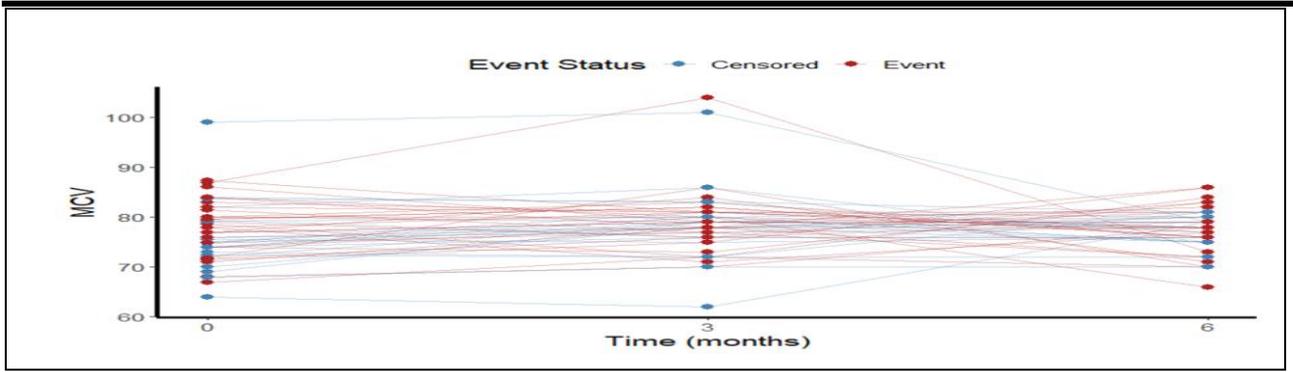
اما الانموذج البقاء الفرعي المرتبط بدالة الخطورة الطولية الفرعية لمستويات (MCV) (Survival Submodel)

$$h_{MCV,i}(t) = h_{MCV,01} \exp\{\psi_{11,MCV}(Sex)_i + \psi_{12,MCV}(ag)_i + \psi_{13,MCV}(treat.)_i + \alpha_{MCV,1} m_{MCV,i}(t)\} \quad (23)$$

الجدول (2) نتائج تقدير معلمات الانموذج المشترك بالطريقة (HGA-EM) للمؤشر الطولي (MCV)

Methods	Type model	Variable	Estimation (Mean)	St-Dev	Confidence Interval 95%
HGA-EM	Longitudinal Sub Model	(Intercept)	77.99	2.62	(75.36, 80.62)
		time	0.1205	0.46	(-0.78, 1.02)
		Age	0.0024	0.11	(-0.21, 0.22)
	Survival Sub Model	Gender	-0.5526	0.95	(-2.43, 1.33)
		Drug_Type	-0.5893	0.49	(-1.57, 0.39)
		Age	-0.0789	0.024	(-0.127, -0.030)
Survival Sub Model	Gender	-0.1847	0.22	(-0.62, 0.23)	
	Drug_Type	0.1039	0.11	(-0.14, 0.34)	
	value(MCV)= α	-0.0088	0.002	(-0.012, -0.006)	

تشير نتائج جدول (2) إلى أن المؤشر الطولي MCV يتمتع باستقرار نسبي عبر الزمن، إذ بلغ متوسطه الابتدائي نحو 77.99 مع تأثير زمني موجب ضعيف وغير دال إحصائياً، كما لم تُظهر المتغيرات الديموغرافية والعلاجية (العمر، الجنس، نوع العلاج) تأثيراً جوهرياً على مساره الطولي، في المقابل، أوضح انموذج البقاء وجود تأثير سلبي دال للعمر على خطر الحدث، بينما لم يكن للجنس ونوع العلاج تأثير معنوي، وتُعد النتيجة الأهم هي معامل الارتباط السلبي الدال إحصائياً ($\alpha = -0.0088$)، الذي يؤكد أن ارتفاع القيم المتوقعة لـ MCV يرتبط بانخفاض خطر حدوث الحدث، مما يبرهن على وجود علاقة بنوية حقيقية بين المسار الطولي للمؤشر واحتمالية البقاء، ويعزز ملاءمة استعمال الانموذج المشترك وطريقة HGA-EM في توصيف هذه العلاقة بدقة.



الشكل (2) تطور مستوى المؤشر الطولي (MCV) حسب حالة حدوث الحدث

يوضح الشكل (2) المسارات الزمنية للمؤشر الطولي MCV عبر ثلاث نقاط زمنية (0، 3، 6 أشهر) مع التمييز بين المرضى تحت المراقبة (باللون الأزرق) والمرضى الذين تعرضوا للحدث (باللون الأحمر)، يلاحظ أن معظم المرضى يتمركزون ضمن نطاق متقارب من قيم MCV مع تغيرات زمنية محدودة نسبياً، مما يعكس استقراراً عاماً في المسار الطولي للمؤشر مقارنةً بمؤشرات أكثر تقلباً مثل Hb أو Urea، كما أن التداخل الواضح بين مسارات المجموعتين يشير إلى أن MCV لا يميز بشكل حاد بصرياً بين حدوث الحدث وعدم حدوثه، رغم وجود بعض الحالات الفردية ذات القيم المتطرفة، هذا السلوك يدعم نتائج الأنموذج المشترك التي أظهرت أن تأثير MCV الطولي محدود، لكنه ما يزال يمتلك ارتباطاً بنيوياً معنوياً مع خطر الحدث عند دمجها ضمن الإطار المشترك، وهو ما لا يمكن استنتاجه بدقة من الرسم الوصفي وحده دون النمذجة الإحصائية.

3.9. تقدير الانموذج المشترك للمؤشر الطولي (Urea_Blood)

$$y_i^{Urea}(t_{ij}) = B_{0,Urea_{Blood}} + B_{1,Urea_{Blood}}(Sex)_i + B_{2,Urea_{Blood}}(age)_i + B_{3,Urea_{Blood}}(treat.)_i + B_{4,Urea_{Blood}}(t_{ij}) + b_{0,Urea_{Blood}} + b_{1,Urea_{Blood}}(t_{ij}) + \epsilon_i^{Urea_{Blood}}(t_{ij}) \quad (24)$$

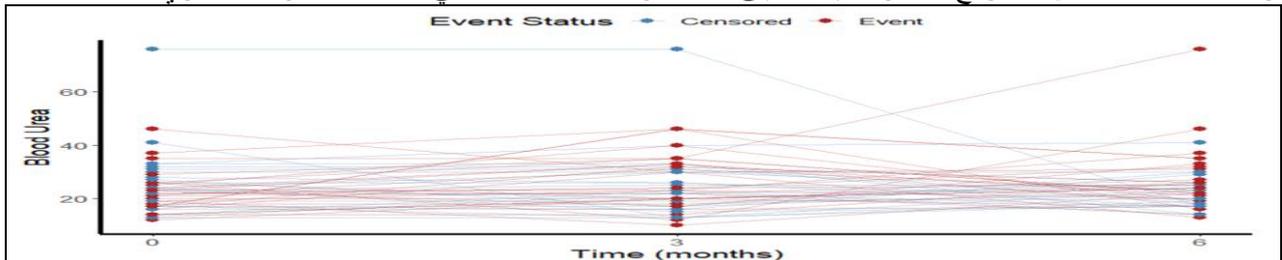
اما الانموذج البقاء الفرعي المرتبط بدالة الخطورة الطولية الفرعية لمستويات (Urea_Blood) (Survival Submodel)

$$h_{Urea,i}(t) = h_{Urea_{Blood},01} \exp\{\psi_{11,Urea_{Blood}}(Sex)_i + \psi_{12,Urea_{Blood}}(age)_i + \psi_{13,Urea_{Blood}}(treat.)_i + \alpha_{Urea_{Blood},1} m_{Urea_{Blood},i}(t)\} \quad (25)$$

الجدول (3) نتائج تقدير معلمات الانموذج المشترك بالطريقة (HGA-EM) للمؤشر الطولي (Urea_Blood)

Methods	Type model	Variable	Estimation (Mean)	St-Dev	Confidence Interval 95%
		(Intercept)	20.73	5.48	(9.94, 31.51)
	Longitudinal Sub Model	Time	0.2835	0.86	(-1.42, 1.99)
		Age	0.2211	0.23	(-0.23, 0.67)
		Gender	-1.6559	2.00	(-5.59, 2.28)
HGA-EM		Drug_Type	0.4428	1.04	(-1.62, 2.50)
		Age	-0.0789	0.024	(-0.127, -0.030)
	Survival Sub Model	Gender	-0.1847	0.22	(-0.62, 0.23)
		Drug_Type	0.1039	0.11	(-0.14, 0.34)
		value(Blood_Urea)	-0.02745	0.004	(-0.037, -0.021)

تشير نتائج جدول (3) إلى أن المؤشر الطولي Urea_Blood يمتلك مستوى ابتدائياً متوسطاً بلغ 20.73 مع فاصل ثقة واسع نسبياً، ويُظهر اتجاهاً زمنياً موجباً (0.2835) يدل على ازدياد واضح في قيم اليوريا مع مرور الزمن، وإن كان هذا التأثير غير دال إحصائياً بقوة. كما تبين أن العمر والجنس ونوع العلاج ليس لها تأثيراً معنوياً واضحاً على المسار الطولي للمؤشر. في انموذج البقاء، يظهر العمر بتأثير سلبي دال على خطر الحدث، في حين لم يكن للجنس ونوع العلاج تأثير إحصائي مهم. وتُعد النتيجة الأبرز هي معامل الارتباط السلبي الدال إحصائياً بين قيم Urea_Blood وخطر الحدث ($\alpha = -0.02745$) ما يشير إلى أن ارتفاع مستويات اليوريا يرتبط بزيادة خطر حدوث الحدث السريري، وهو ما يعكس حساسية هذا المؤشر للتدهور الصحي مقارنةً بالمؤشرات الأخرى، ويؤكد أهمية دمجها ضمن الأنموذج المشترك، إضافةً إلى كفاءة طريقة HGA-EM في التقاط هذا الارتباط البنيوي بدقة.



الشكل (3) تطور مستوى المؤشر الطولي (Urea_Blood) حسب حالة حدوث الحدث

يوضح الشكل (3) المسارات الزمنية لمؤشر Urea_Blood عبر ثلاث فترات متابعة (0, 3, 6 أشهر) مع التمييز بين المرضى تحت المراقبة (أزرق) والمرضى الذين تعرضوا للحدث (أحمر)، يُلاحظ وجود تقلبات زمنية واضحة وتباين فردي كبير في مستويات اليوريا مقارنة بالمؤشرين Hb و MCV، حيث تظهر لدى بعض المرضى الذين تعرضوا للحدث زيادات ملحوظة في قيم اليوريا مع الزمن، خاصة عند نهاية المتابعة، في حين تميل مسارات المرضى الرقابيين إلى أن تكون أكثر استقراراً نسبياً، هذا النمط البصري يشير إلى أن ارتفاع اليوريا وتزايدها عبر الزمن يرتبط بازدياد خطر الحدث، وهو ما ينسجم مع نتائج النموذج المشترك التي أظهرت وجود معامل ارتباط سلبي دال إحصائياً بين Urea_Blood ودالة البقاء، مما يؤكد أن هذا المؤشر يُعد من أكثر المؤشرات حساسية للتدهور الصحي ويبرر دوره المهم في تفسير مخاطر الحدث ضمن الإطار المشترك.

10. الاستنتاجات

1. أثبتت النموذج المشترك كفاءته في تمثيل العلاقة البنوية بين تطور المؤشرات الحيوية وخطر الحدث السريري، إذ أظهر مؤشر Hb تحسناً زمنياً محدوداً (0.0551) وارتباطاً بقائياً سالباً دالاً ($\alpha = -0.0823$) يدل على انخفاض خطر الانتكاس مع ارتفاعه، كما اتسم MCV باستقرار نسبي بمتوسط (77.99) وتأثير زمني ضعيف (0.1205) مع ارتباط بقائي سالب دال لكنه محدود ($\alpha = -0.0088$).
2. بينت النتائج أن مؤشر Urea_Blood هو الأكثر تذبذباً، إذ ازداد بمرور الزمن (0.2835) وارتبط بزيادة خطر الحدث من خلال معامل بقائي سالب أكبر نسبياً ($\alpha = -0.02745$)، كما أظهرت الطريقة الجينية الهجينة HGA-EM قدرة عالية على تحقيق تقديرات مستقرة ودقيقة لمعاملات النموذج المشترك وتمثيل العلاقة الديناميكية بين المؤشرات الحيوية واحتمالية البقاء بكفاءة عالية.

11. التوصيات

1. يُوصى باعتماد النموذج المشترك مع الطريقة الجينية الهجينة (HGA-EM) في الدراسات السريرية التي تتضمن بيانات طولية وبيانات بقاء، لما أظهرته النتائج من كفاءة عالية في تحقيق تقديرات مستقرة ودقيقة وتمثيل العلاقة البنوية بين تطور المؤشرات الحيوية وخطر الحدث بصورة أفضل من التحليل المنفصل.
2. يُنصح بالتركيز على المؤشرات الحيوية الأكثر حساسية للتدهور الصحي، ولا سيما Hb و Urea_Blood، عند المتابعة السريرية لمرضى الثلاثيميا، إذ بينت النتائج أن انخفاض Hb وارتفاع Urea_Blood يرتبطان بشكل معنوي بزيادة خطر الحدث، مما يعزز دورهما كمؤشرات تنبؤية مهمة في اتخاذ القرار العلاجي المبكر.
3. توصي الدراسة بتوسيع التطبيق المستقبلي للطريقة المقترحة ليشمل عينات أكبر وعدد زيارات زمنية أكثر، إضافة إلى دمج مؤشرات حيوية إضافية، بهدف تقييم أداء النموذج المشترك في سياقات طبية أوسع وبيانات عالية الأبعاد.

12. مواد تكميلية

(لا يوجد).

13. مساهمات المؤلفين

انوار داخل هندول: قامت بكتابة البحث وجمع العينات والعمل عليها، و سهاد علي شهيد: الاشراف العام على البحث.

14. التمويل

(لا يوجد).

15. بيان توافر البيانات

تم الحصول على البيانات الحقيقية من مستشفى مدينة الطب / بغداد.

16. شكر وتقدير

لا يسعني إلا أن أقدم بكلمات الشكر والامتنان إلى صرح مدينة الطب في بغداد، وتحديدًا العاملين في شعبة البيانات والإحصاء، الذين لم يخلوا علينا بوقتهم وجهدهم في توفير البيانات اللازمة للجانب التطبيقي من البحث، والشكر موصول لكل القلوب المخلصة التي ساندتني ووفرت لي البيئة المناسبة لإتمام هذا الجهد الأكاديمي.

17. تضارب المصالح

يُعلن المؤلفون عن عدم وجود أي تضارب في المصالح.

References

- [1] Alkhatami, A. (2021). Joint modeling of longitudinal and survival data [Doctoral dissertation, Carleton University]. Ottawa-Carleton Institute for Mathematics and Statistics.
- [2] Qiu, X., Gao, J., Yang, J., Hu, J., Hu, W., Kong, L., & Lu, J. J. (2020a). A comparison study of machine learning (Random Survival Forest) and classic statistic (Cox Proportional Hazards) for predicting progression in high-grade glioma after proton and carbon ion radiotherapy. *Frontiers in Oncology*, 10, 1–10. doi: <https://doi.org/10.3389/fonc.2020.551420>

- [3] Baghfalaki, T., Ganjali, M., & Martins, R. (2025). Approximate Bayesian inference for joint partially linear modeling of longitudinal measurements and spatial time-to-event data. *Journal of Statistical Computation and Simulation*. doi:[10.1080/00949655.2025.2538116](https://doi.org/10.1080/00949655.2025.2538116)
- [4] Caillebotte, A., Kuhn, E., & Lemler, S. (2023). Estimation and variable selection in a joint model of survival times and longitudinal outcomes with random effects. *arXiv preprint*. <http://arxiv.org/abs/2306.16765>
- [5] Ahmed Meri, M., Al-Hakeem, A. H., & Al-Abeadi, R. S. (2022). Overview on thalassemia: A review article. *Medical Science Journal for Advance Research*, 3(1), 26–32. doi: <https://doi.org/1046966/msjar.v3i1.36>
- [6] Shen, N., & González, B. (2021). Bayesian information criterion for linear mixed-effects models. *arXiv preprint*. <http://arxiv.org/abs/2104.14725>
- [7] Leiva-Yamaguchi, V., & Alvares, D. (2021). A two-stage approach for Bayesian joint models of longitudinal and survival data: Correcting bias with informative prior. *Entropy*, 23(1), 50. doi: <https://doi.org/103390/e23010050>
- [8] Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. Chapman & Hall/CRC.
- [9] Taher, A. T., Farmakis, D., Porter, J. B., Cappellini, M. D., & Musallam, K. M. (2025). *Thalassaemia International Federation guidelines for the management of transfusion-dependent thalassemia*. (Report).
- [10] Nguyen, H. T., Vasconcellos, H. D., Keck, K., Reis, J. P., Lewis, C. E., Sidney, S., Lloyd-Jones, D. M., Schreiner, P. J., Guallar, E., Wu, C. O., Lima, J. A. C., & Ambale-Venkatesh, B. (2023). Multivariate longitudinal data for survival analysis of cardiovascular event prediction in young adults: Insights from a comparative explainable study. *BMC Medical Research Methodology*, 23(1). doi: <https://doi.org/10.1186/s12874-023-01845-4>
- [11] Crowther, M. J., Abrams, K. R., & Lambert, P. C. (2013). Joint modelling of longitudinal and survival data. *Stata Journal*, 13(1), 165–184. doi: <https://doi.org/10.1177/1536867x1301300112>

<https://doi.org/10.31272/jae.i151.1537>

<https://admics.uomustansiriyah.edu.iq>

P-ISSN: 1813-6729 E-ISSN: 2707-1359

JAE



Estimation of the Joint Model for High-Dimensional Longitudinal Survival Data Using the Hybrid Genetic Algorithm–EM Method (HGA-EM)

Anwar Dakhel Handool

Dept. of Human Development, Ministry of Planning, Baghdad, Iraq.

Email: anwardakhelhandool@gmail.com , ORCID: <https://orcid.org/0009-0007-3415-5536>

Suhad Ali Shaheed

Dept. of Statistics, College of Administration and Economics, Mustansiriyah University, Baghdad, Iraq.

Email: dr.suhadali@uomustansiriyah.edu.iq , ORCID: <https://orcid.org/0000-0002-1437-6469>

Article Information

Article History:

Received: 15 / 12 / 2025

Revised: 27 / 01 / 2026

Accepted: 02 / 02 / 2026

Available Online: 01 / 03 / 2026

Pages no: 103 – 113

Keywords:

Joint Model, Survival Hazard, Survival Data, Longitudinal Data.

Abstract

Joint models are advanced statistical methods that enable the integrated analysis of longitudinal and survival data, particularly when there is a structural association between a biomarker's temporal trajectory and the risk of a clinical event. In this context, the present study aims to develop a more efficient estimation of joint model parameters by adopting the Hybrid Genetic Algorithm–Expectation Maximization (HGA–EM) approach, which combines the global search capability of genetic algorithms with the optimization and stabilization properties of the EM algorithm. This approach was applied to clinical data from patients with thalassemia, including three longitudinal biomarkers (HB, MCV, and Urea) measured over three follow-up visits. The results showed that the HB index increased gradually over time (0.0551). Higher HB values were associated with a noticeable reduction in the risk of relapse, as indicated by a negative longitudinal–survival association parameter (−0.0823). The mean MCV was approximately 77.99 with a limited time effect (0.1205), and it demonstrated a slight risk-reducing survival association (−0.0088). In contrast, Urea was the most variable biomarker, increasing over time (0.2835) and showing a clearer association with an increased probability of the event, reflected by a relatively larger negative survival coefficient (−0.02745). Overall, these findings highlight the efficiency of the HGA–EM hybrid approach in achieving more stable and accurate estimates than conventional methods and in precisely capturing the dynamic relationship between longitudinal biomarkers and the risk of clinical events.

Correspondence:

Researcher name:

Anwar Dakhel Handool

Email:

anwardakhelhandool@gmail.com

