مدخل إحصائي لتحليل النماذج الوبائية

(دراسة تطبيقية)

Statistical Approach to Analyze Epidemiological Models
(Applied Study)

شيماء مصطفى عبد الرحيم مصطفى سرحان

المعيدة بقسم الإحصاء التطبيقي والتأمين

كلية التجارة – جامعة المنصورة

الأستاذ الدكتور/ فاطمة علي عبد العاطي

أستاذ الإحصاء التطبيقي

كلية التجارة – جامعة المنصورة

الدكتور/ محمد مصطفى شهاب الدين

مدرس الإحصاء التطبيقي

كلية التجارة – جامعة المنصورة

المستخلص:

تؤثر الأمراض المعدية في البداية على منطقة واحدة فقط وسرعان ما تنتشر في جميع أنحاد العالم، وقد تصبح في النهاية جائحة أو وباء كما حدث في وباء COVID – 19، وتعد المساهمة الأساسية لهذا البحث هي نمذجة الوباء في مصر، وكذلك القدرة على عمل تنبؤات بعدد حالات الإصابة والوفاة والشفاء في مصر على المدي طويل الأجل، من خلال استخدام أحد النماذج المتخصصة في النمذجة الوبائية وأحد نماذج التعلم العميق.

فلقد أجري البحث لدراسة ديناميكيات الوباء في مصر باستخدام خوارزميات التعلم العميق والنماذج الوبائية، ومن بين العديد من خوارزميات التعلم العميق، استخدمنا خوارزمية الشبكة العصبية متعددة الطبقات Multi – Layer Perceptron (MLP)، والنموذج الوبائي.Susceptible – Infected – Removed (SIR) وذلك لإظهار دور تلك النماذج في التنبؤ بتطور جائحة COVID-19 في مصر وتقدير الوقت اللازم لاختفائه، وأيضا لبيان التوزيع الوبائي له في مصر.

استهدفت تلك الدراسة الوصول إلى نموذج مناسب للتنبؤ بوباء COVID – 19في مصر، ولذلك تناولت الدراسة تطبيق نموذج SIR، ونموذج الشبكة العصبية متعددة الطبقات MLP، وتمت المقارنة بين تلك الأساليب باستخدام معايير قياس دقة التنبؤ، واتضح من خلالها أن نموذج MLP للتنبؤ بوباء COVID – 19أكثر كفاءة من نموذجSIR .

الكلمات المفتاحية: النمذجة الوبائية، نموذج SIR، الذكاء الاصطناعي، الشبكات العصبية الاصطناعية، نموذج MLP، وباء19 – COVID، رقم التكاثر الأساسي.

Abstract:

Infectious diseases initially affect only one specific region, but they quickly spread to all parts of the world and may eventually become a pandemic or epidemic, as was the case with the COVID-19 pandemic. The primary contribution of this study is to model the epidemic in Egypt and to be able to make predictions about the number of infected, deaths, and recovered cases in Egypt over the long term. This was achieved by using a specialized epidemiological modeling model and a deep learning model.

This study aimed to find a suitable model for predicting the evolution of COVID-19 epidemic in Egypt. Therefore, the study applied the following models in forecasting: The SIR epidemic model, and The Multi-Layer Perceptron (MLP) neural network model.

Comparison between these models has been done to reach the best model that can be used in prediction. These methods were compared using prediction accuracy measures, and it was found that the MLP model for predicting the COVID-19 epidemic was more efficient than the SIR model.

Keywords:

Epidemiological Modeling, SIR model, Artificial Intelligence, Neural Networks, MLP Model, COVID-19 Epidemic, Basic Reproduction Number.

المقدمة ومشكلة الدراسة

منذ ظهور جائحة COVID – 19 في ديسمبر 2019، أثرت على حياة الإنسان والاقتصاد العالمي، وتسببت في اضطرابات كبيرة. فلقد انتشر الوباء في كل قارات العالم، وتسبب في حوالي ما يقرب من 705 مليون إصابة، و701 مليون حالة وفاة حتى 30 أبريل 2023 في جميع أنحاء العالم. وتتسابق الدول لإبطاء انتشار المرض فلا يزال الوباء مستمرًا، وتبذل الجهود العالمية لفهم طبيعة الفيروس وإدراك علم الأوبئة وكذلك محاولة إبطاء انتشاره من خلال تطبيق التدابير الاحترازية اللازمة.

ومن هنا جاءت أهمية النمذجة الوبائيةEpidemiological Modeling، حيث تلعب النماذج الرياضية (النماذج الوبائية)، وكذلك الذكاء الاصطناعي Artificial Intelligence، دورًا هاما في الصحة العامة في العديد من الجوانب، مثل: التصور السريع للمعلومات الوبائية، وتقدير والتنبؤ بالوباء، ورصد سرعة انتشاره، وبالتالي المساعدة في اتخاذ القرارات بشأن الوقاية من الأوبئة ومكافحتها.

أجري البحث لدراسة ديناميكيات الوباء في مصر باستخدام خوارزميات التعلم العميق والنماذج الوبائية. ومن بين العديد من خوارزميات التعلم العميق استخدمنا خوارزمية الشبكة العصبية متعددة الطبقات Multi Layer Perceptron (MLP) –، والنموذج الوبائي Susceptible – Infected – Removed (SIR)، وذلك لإظهار دور نموذجي SIR، وMLP في التنبؤ بتطور جائحة COVID-19 في مصر وتقدير الوقت اللازم لاختفائه، وأيضا لبيان التوزيع الوبائي له في مصر.

أهداف الدراسة

يمكن بلورة أهداف الدراسة في النقاط التالية:

١) الوصول إلى نموذج مناسب لفهم ديناميكيات انتشار الوباء وتحليل بعض تفاصيل وخصائص الأمراض الوبائية.

٢) التنبؤ بعدد السكان المصابين والمتعافين والوفيات في المجتمع مع استبعاد الزيادات المحتملة في نفس الوقت في عدد الأفراد المعرضين للإصابة، ومن ثم التنبؤ بالاتجاه العام للوباء وتطوره.

٣) تقدير المعلمات الوبائية المختلفة مثل: معدل الانتشار(α)Transmission Coefficient or Rate، ومعدل الشفاء(γ)Recovery Coefficient، ومعدل الوفاة (μ) Death Rate، والتي منها يتم تقدير المقاييس الوبائية المختلفة كتقدير عدد التكاثر الأساسي (R_o) Basic Reproduction Number، ومستوى المناعة الجماعية Herd Immunity Limit (HIL)، ومستوي التطعيم اللازم للقضاء على المرض V_c، ومدة الوباء المتبقية لانتهائهEpidemic Time، ومن ثم يمكننا حساب والتنبؤ بحجم السكان

المعرضين للإصابةS، والمصابينI، والمتعافين من المرضR، وبالتالي القدرة على تتبع مسار الوباء والتنبؤ بالاتجاه العام له.

٤) عرض الخوارزميات الأساسية لنموذجي SIR، وMLP، واستخدام هذه الخوارزميات في مجال النمذجة الوبائية.

٥) عرض التحليل الرياضي بشكل أكثر وضوحا وبالتالي تحليل التفسيرات النظرية بشكل سليم.

٦) تحديد ما إذا كان الوباء يتبع التوزيع الطبيعي أم لا.

٧) عمل تنبؤات وتوقعات على المدي طويل الأجل.

٨) التعبير عن فائدة اللقاح وإجراءات الوقاية في تنمية المناعة الجماعية وزيادتها.

٩) حساب المؤشرات الإحصائية التي تسمح لمسؤولي الصحة معرفة ما إذا كانت السياسات المتبعة فعالة أم لا، وإذا كان معدل انتشار الوباء في تزايد أو تناقص أو في حالة مستقرة.

أهمية الدراسة

تكمن أهمية الدراسة فيما يلي :

أولا: الأهمية العلمية

تتمثل الأهمية العلمية والإحصائية لهذا البحث في عرض الأساس الرياضي والإحصائي لأحد أساليب السلاسل الزمنية وهو نموذج SIR، وأحد نماذج الذكاء الاصطناعي (التعلم العميق)، وهو أسلوب الشبكات العصبية الاصطناعيةANNوالمقارنة بين النموذجين لتحديد الأفضل في التنبؤ بالاتجاه العام لوباء COVID-19 في جمهورية مصر العربية.

ثانيا: الأهمية العملية

ترجع أهمية البحث العملية في بناء نموذج إحصائي للتنبؤ بالاتجاه العام لوباء COVID-19، وتحديد حدود التطور المستقبلي للوباء في جمهورية مصر العربية، مما له تأثير على الاقتصاد ونظم الصحة والموارد المادية والبشرية، ومن ثم يمكن تقييم درجة تأهب الدولة لمواجهة هذا الوباء.

النماذج المستخدمة

هناك العديد من نماذج التنبؤ بالأمراض المعدية، وفي هذا البحث سنقوم باستخدام أحد النماذج المتخصصة في النمذجة الوبائية، وأحد نماذج التعلم العميق، وهما نموذجي SIR، MLP.

أولا: نموذج SIR

نموذج SIR (Susceptible, Infected, Removed) هو أحد أبسط النماذج الوبائية، يستخدم لتقدير عدد الأفراد المصابين والمتعافين من مرض وبائي معين، وذلك في مجتمع كبير ومغلق، كما يعد

أيضا أحد النماذج المستخدمة في السلاسل الزمنية، وبالتالي فهو نموذج ديناميكي dynamic model ولتكوين نموذج ديناميكي فإننا نحتاج لما يسمي بمتغيرات الحالة State Variables، ونحتاج أيضا لمعرفة كيف تتغير هذه المتغيرات بمرور الوقت (معدل التغير أي الميل Slope). ومنه فإننا نستطيع التنبؤ بالاتجاه المستقبلي للوباء، والعدد المحتمل للمصابين في حالة وجود او عدم وجود تطعيم، وتقدير السمات الرئيسية للمرض، وكذلك التأثير المحتمل لتدخلات الصحة العامة المختلفة.

ولتحديد متغيرات الحالة الأساسية في هذا النموذج فإننا نقوم بتقسيم المجتمع بشكل أساسي إلى ثلاث فئات(Jakhar, et al., 2020) :

١- مجموعة السكان المعرضين للإصابة susceptible :تمثل مجموعة السكان الذين ليس لديهم مناعة ضد العامل المعدي، والمعرضة للإصابة بالمرض، ولكنها سليمة وتمثل هذه المجموعة سكان مجتمع الدراسة عند بداية الوباء.

٢- مجموعة السكان المصابة بالفعل Infected: مجموعة السكان المصابة بالمرض وتنقل الوباء لمن يتواصل معهم من أفراد المجموعة S.

٣- Removed تشمل مجموعة السكان المعافاة من الوباء والمتوفين أيضاً بسبب الوباء (Recovered, Deaths).

تتوقف جودة النماذج على دقة الافتراضات التي تستند إليها، إذا وجدنا أن توقعات النموذج لا تتماشي مع القيم الفعلية، يعني ذلك ضرورة تغيير الافتراضات الأولية ليصبح النموذج مفيدًا (Wang, Jiale, et al., 2022)،(Paul, Amit, et al., 2022, p4)

ويمكن عرض افتراضات النموذج في النقاط التالية:

١) مجتمع الدراسة كبير وثابت أي ثبات حجم السكان (Closed Population).
٢) تجاهل التغيرات الحيوية والسكانية للمجتمع (الوفيات والمواليد والهجرة)، عدا الوفيات بسبب المرض.
٣) لا يؤثر العمر والجنس والعرق والوضع الاجتماعي على احتمالية إصابة الشخص حيث إن الظاهرة تنتشر بشكل عشوائي، وبالتالي تتساوى فرص الإصابة لجميع السكان.
٤) لا توجد مناعة موروثة (مشتقة من الأم) في السكان.
٥) يعتبر اختلاط الأفراد في جميع السكان يحدث بشكل متجانس.
٦) الشفاء من المرض ينتج مناعة دائمة أي لا توجد فرصة للعدوى مرة أخرى.
٧) غالبا ما يقترح أن يكون رقم التكاثر (RO) يتبع التوزيع الأسي (من إحصائيات بولتزمان الكلاسيكية classical Boltzmann Gibbs (BG) statistics).

٨) لا توجد تحركات مباشرة لأفراد المجتمع من الفئة S الى الفئة R.

يعد نموذج SIR (أو SIRD) أحد النماذج التي تستخدم في نمذجة الأمراض المعدية وأبسطها، فيحسب عدد المصابين بمرض معدي في مجتمع ثابت ومغلق بمرور الوقت. ويحتوي النموذج على أربع معادلات تفاضلية متعلقة بعدد الأشخاص المعرضين للإصابة (S)، والمصابين (I)، وكذلك (R) ويمثل المتعافين و(D) الذي يمثل المتوفين (وهنا تم تقسيم المتغير Removed إلى المتغيرين Recovered, Deaths).

تم صياغة النموذج على النحو التالي:

$$\frac{dS}{dt} = -\alpha \frac{S}{N} I \rightarrow (1)$$

$$\frac{dI}{dt} = +\alpha \frac{SI}{N} - \beta I_t \rightarrow (2)$$

$$\frac{dR}{dt} = +\gamma I_t \rightarrow (3)$$

$$\frac{dD}{dt} = +\mu I_t \rightarrow (4)$$

ويوضح الجدول التالي متغيرات الدراسة وكذلك المعلمات كما يلي:

جدول (1) يوضح متغيرات الدراسة وكذلك المعلمات

S	**المعرضين للإصابة بالوباء**
I	**حالات الإصابة**
R	**يشمل حالات الشفاء**
D	**حالات الوفاة بسبب الوباء**
N	**حجم السكان (المجتمع محل الدراسة)**
α	**معدل انتشار الوباء (معدل الانتقال أو معدل الإصابة)**
β	**معدل الإزالة**
γ	**معدل الشفاء**
μ	**معدل الوفاة بسبب الوباء**
λ	**معدل النمو الأسي**
R_0	**رقم التكاثر الأساسي**
HIT	**عتبة المناعة الجماعية**
V_c	**مستوي التطعيم**
OV	**تجاوز الحد**

ويمكن توضيح بعض المقاييس الوبائية المستخدمة ذات الصلة كما يلي:

١) **رقم التكاثر الأساسي Basic Reproduction Number:** يعتبر مقياس أساسي لانتقال العدوي ويمثل متوسط عدد حالات الإصابة الثانوية الناجمة عن حالة إصابة واحدة ويشار إليه بالرمز R_o (Katris, Christos, 2021, p.4).

بصيغة أخري هو متوسط عدد الحالات الثانوية المكتسبة من الحالة الأولية في مجموعة سكانية معرضة للإصابة في جميع مراحل العدوى.

أو متوسط حالات العدوى الجديدة المسببة بواسطة كل حالة في مجموعة معرضة للعدوى بالكامل والتي هي متجانسة، أو مجموعة مختلطة جيدا، أي يمكن لكل شخص فيها الاتصال مع كل فرد آخر معرَّض للعدوى.

٢) **المناعة الجماعية Herd Immunity:** المناعة الجماعية أو مناعة القطيع (HI) هي شكلٌ من أشكال الحماية غير المُباشرة من مرضٍ معدٍ، وتحدث عندما تكتسبُ نسبةٌ كبيرةٌ كافية من المجتمع مناعةً لعدوًى معينة، إما بسبب الإصابة بها سابقًا أو التطعيم، مما يُوفر حمايةً للأفراد الذين ليس لديهم مناعةٌ ضد المرض (Randolph, H. E., & Barreiro, L. B., 2020).

حين تكتسب نسبة حاسمة من السكان المناعة ضد الوباء، يطلق على هذه الحالة عتبة المناعة الجماعية (HIT) أو مستوى المناعة الجماعية (HIL) وهي النقطة التي تنخفض فيها نسبة الأفراد المعرضين للإصابة في مجتمع ما عن الحد الأدنى المطلوب لانتقال المرض. وحينها لا يستمر المرض في الانتشار بين السكان ويتوقف المرض المتوطن. يمكن قياس قيمة هذه العتبة بضرب R_o في S وهو هنا نسبة السكان المعرضين للإصابة ومنها نجد:

$$\therefore \mathrm{HIT} = 1 - \frac{1}{R_O} = \frac{R_O - 1}{R_O}$$

٣) **التطعيم Vaccination:** لتحقيق المناعة الجماعية HI من خلال التطعيم، يهدف مصـــنعو اللقاحات إلى إنتاج لقاحات ذات معدلات فشل منخفضة، ويهدف صانعو السياسات إلى تشجيع استخدامها. بعد الإدخال الناجح للقاح واستخدامه على نطاق واسع، يمكن ملاحظة انخفاض حاد في الإصابة بالأمراض التي يقي من الإصابة بها، مما يقلل من عدد حالات الإصابة والوفيات التي تسببها مثل هذه الأمراض.

وإذا افترضنا أن اللقاح فعال بنسبة 100%، فإن المعادلة المستخدمة لحساب عتبة مناعة القطيع يمكن استخدامها لحساب مستوى التطعيم اللازم للقضاء على المرض، مكتوبًا بصيغة V_c، ولكن عادة ما تكون فعالية اللقاحات غير كاملة، لذلك يجب مراعاة فعالية اللقاح E عند حساب V_c:

$$V_c = \frac{HIT}{E} = \frac{1 - \frac{1}{R_o}}{E}$$

٤) **تجاوز الحد Overshoot:** تجاوز الحد هو نتيجة غير مرغوب فيها، قد تحدث بشكل كبير حتى بعد تحقيق المناعة الجماعية، مما يؤدي إلى عدد كبير من الإصابات والوفيات قبل أن ينخفض الانتشار. وهو الفرق بين النسبة التراكمية للأفراد المصابين والعتبة النظرية للمناعة الجماعية (Nguyen, M. M., 2024):

Overshoot=Proportion of Total Infections – HIT

٥) **الحالات النشطة Active Cases:** يقصد بالحالات النشطة الحالات التي لا تزال لا نتيجة لها (سواء بالشفاء أو الوفاة)، ويتم حساب الحالات النشطة اليومية من المعادلة التالية:

Daily Active Cases = Total Infected – [Total Recovered + Total Deaths]

طريقة تقدير النموذج

يعتبر تقدير معلمات النموذج خطوة أساسية لفهم ديناميكيات انتشار الأوبئة، وهناك العديد من الطرق التي تُستخدم لتقدير هذه المعلمات بناءً على البيانات المتوفرة. يمكن تصنيف هذه الطرق إلى (Sebbagh, A., & Kechida, S., 2022):

١) الطرق التقليدية: تُعتبر أدوات أساسية في التقدير، لكنها تتطلب افتراضات قد لا تتوافق غالبًا مع طبيعة البيانات الوبائية. ومنها: طريقة المربعات الصغرى العادية Ordinary Least Squares وطريقة الإمكان الأعظم Maximum Likelihood Method.

٢) الطرق العددية والتحسين الرياضي: توفر حلولاً مرنة وسريعة، لكنها قد تتطلب محاكاة مكثفة. ومنها: طريقة أويلر Euler Method، والتحسين الخطي Optimization Method.

٣) الطرق المتقدمة: تُقدم تقديرات دقيقة ومرنة عند التعامل مع البيانات الضوضائية أو غير المؤكدة. ومنها: الطرق البازية Bayesian Inference، وخوارزمية كالمان فلتر الموسعة Extended Kalman Filter

٤) تقنيات الذكاء الاصطناعي والشبكات الديناميكية: توفر إمكانيات كبيرة للتعامل مع البيانات الضخمة وتحليل الأنماط المعقدة، ما يجعلها أدوات فعالة لتقدير المعلمات خاصةً في الأنظمة الديناميكية.

لكل طريقة مزاياها وعيوبها، ويتم اختيار الطريقة المُثلى بناءً على: طبيعة البيانات، والأهداف البحثية، ومتطلبات التحليل، والدقة المطلوبة، وكذلك القدرة الحسابية حيث إن بعض الطرق تحتاج إلى قدرات حسابية أعلى.

ثانيا: الشبكات العصبية الاصطناعية ونموذج MLP

الشبكات العصبية الاصطناعية (ANNs) هي نماذج حسابية مستوحاة من عمل الدماغ البشري. تُستخدم ANNs في العديد من التطبيقات بما في ذلك التنبؤ وتحليل البيانات فلديها قدرة عالية علي تحديد وحل العلاقات غير الخطية Non – Linear Relations. وفي هذا البحث، سنناقش استخدام نموذج الشبكات العصبية متعددة الطبقات(Multilayer Perceptron (MLP في التنبؤ بوباء كورونا في مصر، مع التركيز على حالات الإصابة، التعافي، والوفاة.

الشبكات العصبية متعددة الطبقات Multilayer Perceptron

MLP هو نوع من الشبكات العصبية الاصطناعية تم اقتراحه بواسطة Minsky و Papert في الخمسينيات من القرن الماضيs1950 .ويمثل Perceptron اول نموذج بسيط للتعلم الآلي، حيث إن Perceptron هو العنصر الأصغر في الشبكة ويقوم بإجراء حسابات معينة لتحليل خصائص المدخلات، وتخضع عملية تدريب شبكة MLP الي التدريب الإشرافي والتي تعمل على تعديل الأوزان تدريجيا بناء على الفرق بين المخرجات المتوقعة والمخرجات الفعلية. تتكرر هذه العملية حتى يتم تقليل الخطأ إلى الحد الأدنى(Azeem, Muhammad, et al., 2023, p5) ولشبكة MLP ذاكرة ديناميكية dynamic memory وهي مناسبة جدا في مجال التنبؤ بالسلسلة الزمنية، وفي الوقت الحالي، تستخدم على نطاق واسع في مختلف المجالات وقد حققت نتائج ملحوظة وناجحة في مجال التنبؤ.

تتكون الشبكة من ثلاث طبقات رئيسية:

١) طبقة المدخلات (Input Layer): تستقبل البيانات الخام وتمثل المتغيرات المستقلة التي يتم تقديمها إلى النموذج وكل متغير يتم تمثيله بوزن محدد.

والأوزان هي قيم عددية مرتبطة بكل متغير يتم تعديلها خلال عملية التدريب لتحقيق أفضل أداء.

٢) الطبقة المخفية (Hidden Layers): تقوم بعمليات حسابية ومعالجة البيانات. يمكن أن تكون هناك واحدة أو أكثر من هذه الطبقات.

٣) طبقة المخرجات (Output Layer): تمثل القيم المتوقعة للمتغيرات التابعة بناء على قيم المدخلات.

وللتنبؤ بتطور حالات كورونا في مصر، نحتاج إلى بيانات تاريخية تشمل:

- حالات الإصابة اليومية.
- حالات الشفاء اليومية.
- حالات الوفاة اليومية.

وحيث إن الشبكة العصبية لا تبرمج، بل إنها تقوم بالتعلم والتدريب، ويقصد بتدريب الشبكة هو إجراء يتم فيه استخدام قواعد التدريب والتعلم لضبط الأوزان. وطرق التدريب الأساسية هي: التدريب الإشرافي supervised training (التدريب في وجود معلم)، التدريب غير الإشرافي unsupervised training (التدريب في عدم وجود معلم)، التدريب بإعادة التدعيم reinforcement training or hybrid training، التدريب المتأخر delay learning.

وفي هذا البحث سنقوم بتطبيق شبكة MLP، التي تخضع فقط للتدريب الاشرافي ويقصد به التدريب في وجود معلم، حيث تعد هذه الطريقة هي الأكثر انتشارا في تدريب الشبكة العصبية، وفيها تجمع عينات من البيانات تحتوي قيماً للمتغيرات المدخلة وقيماً للمتغيرات المخرجة، وتقوم الشبكة بمقارنة النتائج التي تقدرها للمتغيرات الخارجة لكل عينة مدخلة بالقيم الفعلية لهذه المتغيرات، وبناء على ذلك تقوم الشبكة بإجراء التعديلات على أوزان الاتصال بهدف تقليل الأخطاء في النتائج، ثم تعاد عملية التدريب عدة مرات إلى أن يتم الوصول إلى نتائج مقبولة ومقاربة للقيم الفعلية.

النهج المستخدم في بناء وتطبيق الشبكة العصبية متعددة الطبقات (MLP) لتحليل البيانات، وكذلك الأهداف التي يسعي كل نهج لتحقيقها.

لإجراء التحليل على برنامج SPSS، فإننا بعد جمع البيانات نحتاج إلى إعدادها كي نتأكد من خلوها من القيم المفقودة، ثم يتم تقسيم المتغيرات إلى مستقلة وأخري تابعة (مستهدفة)، وفي تحليلنا هذا نقوم بتحويل المتغيرات (I, R, D) الي Lagged Valuesوالتي يطلق عليها القيم السابقة أو التاريخية أو المتأخرة حيث تمثل [I (n – 1)، R (n – 1)،D (n – 1)] المتغيرات المستقلة وتستخدم في التنبؤ بالمتغير التابع.

وهناك ثلاث طرق لإجراء تحليل الشبكة العصبية متعددة الطبقات (MLP) وهم:

١) شبكة واحدة لكل متغير Separate Network for Each Variable : يتم وضع متغير واحد فقط كمتغير تابع، واستخدام المتغير المتأخر لهذا المتغير كمتغير مستقل.
٢) شبكة منفصلة لكل متغير تابع على حدة (تحليل منفصل لكل متغير) Separate Network for Each Dependent Variable: يتم بناء شبكة عصبية منفصلة لكل متغير تابع (I, R, D)، يكون لكل شبكة عصبية متغير واحد تابع الذي تحاول الشبكة التنبؤ به أو تصنيفه، بالإضافة الي مجموعة من المتغيرات المستقلة (ويتم استخدام المتغيرات المتأخرة كمتغيرات مستقلة للشبكة).
٣) شبكة واحدة لجميع المتغيرات (المستقلة والتابعة معا)Single Network for All Variables : هنا نقوم بالتعامل مع جميع المتغيرات ككل وبناء شبكة عصبية واحدة تشمل جميع المتغيرات ,Infected)

(Recovered, Deathsكمتغيرات تابعة، وجميع المتغيرات المتأخرة (Lagged Infected, Lagged Recovered, Lagged Deaths) كمتغيرات مستقلة. وبالتالي تدخل جميع المتغيرات الي الشبكة، ويتم معالجة العلاقة بينها داخل نفس الهيكل العصبي.

الخلاصة: نختار النهج الأول عندما يكون لدينا بيانات سلسلة زمنية، ويكون المتغير نفسه هو العامل الرئيسي الذي يؤثر على التوقعات المستقبلية، وفي الحالات التي تفتقر إلى بيانات خارجية ذات صلة، أما النهج الثاني فهو مناسب عندما يكون لكل متغير خصائص مختلفة جدا لتحسين الدقة والموثوقية لكل متغير على حدة وتحليل مخصص لكل نتيجة. ونختاره عندما تكون التفاعلات بين المتغيرات أقل أهمية أو يمكن معالجتها بطرق أخري، أما النهج الثالث فهو مناسب للحصول على رؤية شاملة ومراعاة العلاقات المتبادلة بين المتغيرات.

وباتباع أي نهج منهم، يمكننا تحقيق فهم أفضل وتنبؤات أكثر دقة باستخدام تحليل الشبكات العصبية متعددة الطبقات (MLP). ويعتمد الاختيار على الأهداف البحثية ومتطلبات التحليل. هنا سوف نختار النهج الثالث فهو مناسب أكثر عند مقارنته بنموذج SIR فهو نموذج ديناميكي يعتمد التغير في كل متغير على التغيرات في المتغيرات الأخرى؛ للحصول على رؤية شاملة ومراعاة العلاقات المتبادلة بين المتغيرات.

مصادر البيانات

يتمثل مصدر البيانات في البيانات اليومية المتعلقة بعدد حالات الإصابة، والشفاء، والوفيات جراء وباء COVID – 19 في الفترة من 14 فبراير 2020 الي 30 ابريل 2023، والمنشورة في موقع منظمة الصحة العالمية (WHO)، وموقع Worldometer، وموقع Our World in Data ، وكذلك وزارة الصحة المصريةMoHP، والجهاز المركزي للتعبئة العامة والإحصاء بجمهورية مصر العربية.

الدراسة التطبيقية

أولا: التحليل الوصفي للبيانات الفعلية لوباء COVID - 19 في مصر:

يقدم هذا القسم التحليل الوصفي للبيانات اليومية الفعلية عن حالات الإصابة والوفاة والتعافي لوباء COVID – 19 منذ بداية ظهوره في منتصف فبرير ٢٠٢٠ حتى ٣٠ ابريل ٢٠٢٣. وكما هو موضح في الجدول التالي الإحصاءات الوصفية خلال الموجات الأربع وأيضا خلال فترة الوباء بالكامل:

جدول (2) يوضح الإحصاءات الوصفية للبيانات الفعلية

	Wave 1			Wave 2			Wave 3		
Measures	**I**	**R**	**D**	**I**	**R**	**D**	**I**	**R**	**D**
Mean	**399.56**	**360.91**	**23.17**	**688.115**	**421**	**42**	**556.06**	**533**	**29.16**
Median	**170.5**	**147**	**14**	**627**	**400.85**	**46**	**569**	**495.99**	**31**
Max	**1774**	**1716**	**97**	**1418**	**813**	**64**	**1203**	**1005**	**68**
Min	**0**	**0**	**0**	**351**	**67**	**11**	**0**	**23**	**0**
Sd	**472.789**	**395.078**	**24.16**	**256.52**	**170.94**	**13.7666**	**389.52**	**274.096**	**19.521**
Sum	**112676**	**101777**	**6353**	**89455**	**52111**	**5460**	**101759**	**90767**	**5336**

	Wave 4			Total		
Measures	**I**	**R**	**D**	**I**	**R**	**D**
Mean	**368.47**	**342.84**	**13.02**	**440.76**	**399**	**21.2**
Median	**0**	**342**	**0**	**168**	**377.56**	**13**
Max	**5516**	**1812**	**132**	**5516**	**1812**	**132**
Min	**0**	**0**	**0**	**0**	**0**	**0**
Sd	**673.73**	**294.69**	**21.7999**	**565.23**	**313.03**	**23.36**
Sum	**212240**	**197473**	**7499**	**516023**	**442128**	**24830**

يوضح الجدول (2) أن المتوسط اليومي للحالات المصابة والوفيات والتعافي أكبر من قيم الوسيط للمتغيرات الثلاث خلال الأربع موجات وأيضا خلال فترة الوباء بالكامل. ونستنتج من ذلك أن البيانات لا تتبع التوزيع الطبيعي كما أنها ملتوية تجاه اليمين. وهذا يشير الي وجود عدد قليل من الأيام ذات عدد حالات مرتفع بشكل غير طبيعي نتيجة لتفشي مفاجئ أو موجات جديدة، هذه القيم العالية تسحب المتوسط الي أعلي، مما يجعله أكبر من الوسيط، مما يعني أن أعداد الحالات عند نهاية كل موجة مصحوبة بالتواء تجاه اليمين (أي في تزايد، وبالتالي أعداد الحالات عند نهاية كل موجة تميل الى القيم المرتفعة).

كما يُلاحظ ارتفاع قيم الانحراف المعياري، وهذا يشير الي وجود فترات من الزيادات والانخفاضات الكبيرة (ذروة وانتكاس) وهو ما يعكس تقلبا ملحوظا في أعداد حالات الإصابة، والوفيات، والشفاء، وهذا يشير إلى عدم استقرار الوضع الوبائي مما يتطلب اتخاذ إجراءات احترازية سريعة.

ويمكن أيضا من خلال الإحصاءات الوصفية توضيح العلاقة بين المتغيرات؛ على سبيل المثال، قد نلاحظ أنه مع ارتفاع حالات الاصابة، يمكن أن ترتفع معدلات الوفيات والتعافي، ولكن بشكل غير متساو. يمكن أن يساعدنا هذا في تقديم تفسير أعمق حول فعالية الاستجابة الصحية وتأثير الفيروس على المجتمع.

إلى جانب ذلك، يمكن للإحصاءات الوصفية أن تساعد في التنبؤ باتجاه الوباء على المدي القصير. فمثلا، إذا أظهر تحليل البيانات أن هناك ارتفاعا متزايدا في عدد الحالات المصابة، يمكن استخدام هذه المعلومات للتحذير من موجات جديدة، وبالتالي معرفة نمط التغير في البيانات بمرور الوقت، مثل ارتفاع أو انخفاض الوفيات، قد يساعد في تخصيص الموارد الصحية بشكل أفضل.

ثانيا: بالنسبة لنموذج SIR

تتمثل المشكلة الأساسية في النماذج الرياضية في تحديد قيم معلمات النموذج التي تُحقق أقرب تطابق بين التنبؤات الناتجة عن النموذج والبيانات الفعلية. التحدي هنا يكمن في إيجاد طريقة تتناسب مع الطبيعة غير الخطية للبيانات المستخدمة. ومن أبرز الطرق المُستخدمة لمعالجة هذه المشكلة هي خوارزمية التحسين غير الخطية (Nonlinear Optimization Algorithm).

في هذا الإطار، تم استخدام لغة البرمجة R إصدار (4.3.1) لتحليل البيانات، حيث تُستخدم بدايةً حزمة deSolve في حل المعادلات التفاضلية العادية (ODEs) الخاصة بالنموذج، مع تطبيق خوارزمية التحسين optim، التي تهدف إلى تقليل الفرق بين القيم المقدرة والقيم الفعلية المستمدة من البيانات اليومية، وبالتالي تُتيح تقدير قيم المعلمات المُثلي، وعليه تُحسن دقة النموذج في محاكاة الظاهرة وفهم الآليات الديناميكية المرتبطة بها. كما تم الاعتماد على عدد من الحزم الإحصائية Packages لإجراء التحليل واستخراج النتائج، وهي:.(EpiModel, reshape2, ggplot2, Covid– 19. analytics)

وفي نموذج مثل: SIR، تُقدَّر المعلمات اليومية $(\alpha, \beta, \mu, \gamma)$ يوميًا بناءً على البيانات التاريخية اليومية للمتغيرات، ومن خلال هذه التقديرات الخاصة بمعلمات النموذج، يُمكن تقدير المقاييس الوبائية المختلفة والتنبؤ بالمتغيرات [S(t+1), I(t+1), R(t+1), D(t+1)] اعتمادًا على المعادلات التفاضلية الخاصة بالنموذج كما يلي:

λ (Exponential Growth Rate) $= \hat{\alpha} - \hat{\beta}$

Epidemic Time= ln (2) / λ

$$R_o = \frac{\hat{\alpha}}{\hat{\beta}}$$

$$R_e = R_o . X$$

$$\text{HIT or } P_c = \frac{R_o - 1}{R_o} = 1 - \frac{1}{R_o}$$

$$D(\text{duration of infection}) = 1/\gamma$$

OV=Proportion of Total Infections – P_c

$$V_c = \frac{1 - \frac{1}{R_o}}{E} = \frac{HIT}{E} = \frac{P_c}{E}$$

$$S_{n+1} = S_n + \hat{\alpha} . \frac{S_{n-1}}{S_O} . \ I_{n-1}$$

$$I_{n+1} = I_n + \hat{\alpha}.\frac{I_{n-1}}{S_O}.\ I_{n-1} - \hat{\beta}\ I_{n-1}$$

$$(\text{Removed})R_{n+1} = R_n + \hat{\beta}\ I_{n-1}$$

$$(\text{Recovered})R_{n+1} = R_n + \hat{\gamma}\ I_{n-1}$$

$$D_{n+1} = D_n + \hat{\mu}\ I_{n-1}$$

استناداً إلى المعطيات المذكورة وبالتطبيق على برنامج R، تم تقدير المعلمات بشكل يومي خلال موجات الوباء الأربع، وكذلك على مدار فترة الوباء كاملة، بالاعتماد على البيانات اليومية الفعلية لحالات الإصابة، والوفاة، والشفاء. وقد أسفر التحليل عن القيم التالية:

جدول (3) يوضح قيم المعلمات المقدرة لنموذج SIR

Waves	α	γ	μ	β	λ	Epidemic Time	R_o	P_c and V_c	OV
A	**0.105**	**0.078**	**0.005**	**0.083**	**0.023**	**30**	**1.27**	**21.3%**	**0**
B	**0.185**	**0.0117**	**0.057**	**0.069**	**0.116**	**6**	**2.68**	**62.7%**	**0**
C	**0.258**	**0.107**	**0.058**	**0.165**	**.093**	**8**	**1.56**	**35. 9%**	**0**
D	**0.077**	**0.055**	**0.05**	**0.105**	**-.03**	**0**	**0.73**	**0**	**0**
Total	**0.625**	**0.252**	**0.17**	**0.422**	**0.203**	**3**	**1.48**	**32.5 %**	**0**

بالنسبة للموجة الأولي A: نجد أن معدل انتشار الوباء α=0.105 وكذلك معدل الإزالةβ=0.083 ومنها نجد أن R_o = 1.27 (أكبر من الواحد الصحيح)؛ مما يعنى أن الوباء انتشر بسرعه ويجب اتخاذ الإجراءات اللازمة؛ كي نتمكن من السيطرة عليه ووقف وإبطاء انتشاره؛ لذا نجد أننا بصدد تطعيم حوالي 22%من إجمالي حجم السكان.

بالنسبة للموجة الثانية B: نجد أن معدل انتشار الوباءα=0.185 وكذلك الإزالة β=0.069 نلاحظ زيادة معدل الانتشار خلال الموجه الثانية عن الموجة الأولى ومصحوبا أيضا بانخفاض معدل الإزالة؛ لذا فإننا نجد R_o =2.68 وذلك يعنى أن كل شخص مصاب يعدي في المتوسط ثلاث أشخاص تقريبا أي أننا نحتاج لتطعيم 62.7% من إجمالي حجم السكان. مما يعنى أنه وفق تقديرات النموذج أن الموجة الثانية أكثر حدة من الموجة الأولي مما يستدعي سرعة تصنيع على اللقاحات.

أما الموجة الثالثة C: فإننا نجد معدل انتشار الوباءα=0.258 وكذلك الإزالة β=0.165 وهنا نلاحظ ارتفاع معدل الانتشار، ولكنه مصحوبا بارتفاع معدل الإزالة الذي أدى إلى انخفاض قيمة R_o =1.56 عن الموجة الثانية وذلك بسبب ظهور اللقاحات حيث تمت أول جرعة لقاح في ٢٠٢١/١/٢٤.

وارتفاع معدل الانتشار هنا يعني زيادة مستوي التواصل بين الأفراد وذلك صحيحا لأن قرارات الدولة المصرية في تلك الفترة أشارت الي رجوع الحياة الي طبيعتها وتخفيف الإجراءات الصارمة المتخذة من قبل، وبالتالي عادت الحياة كما كان الحال قبل ظهور الجائحة وازداد التواصل بين الأفراد وبالتالي زيادة

في معدل الانتشار، ولكن وجود اللقاحات في تلك الفترة أدى إلى ارتفاع في معدل الشفاء وبالتالي انخفاض في متوسط أعداد الإصابة الناتجة عن التواصل بين الأفراد.

أما الموجة الرابعة D: فإننا نجد انخفاض في معدل الانتشار، ومصحوبا بارتفاع في معدل الإزالة مما يعنى فعالية اللقاحات المستخدمة. وبالتالي انخفاض في متوسط أعداد الإصابة الناتجة عن التواصل بين الأفراد حتى وصل إلى $R_o = 0.73$ (أي أن كل شخص مصاب يعدي في المتوسط أقل من شخص واحد فقط وهذه هي المرحلة المراد الوصول اليها) وذلك بسبب وجود مناعة جماعية لدي الناس ووفقا لهذه التقديرات فإننا نجد أننا لا نحتاج إلى مزيد من التطعيم HIT =0%.

ومن المعلوم أنه إذا كان $(\lambda=\alpha-\beta) > 0$ فان نسبة الأعداد المصابة تنمو بشكل أسي. ومن خلال قيمة هذا المعدل يمكننا معرفة الوقت المتبقي لانتهاء الوباء Epidemic Time من العلاقة $\ln(2) / \lambda$.

ومن الجدول السابق نلاحظ أن قيمة معدل النمو الأسي خلال أول ثلاث موجات موجبة، مما يوضح أن الحالات المصابة تنمو بشكل أسي ومنه فانه يوجد وقت متوقع لانتهاء الوباء، بينما أصبحت قيمته سالبة خلال الموجة الرابعة وهذا يعنى أن الحالات المصابة خلال هذه الموجة لا تتزايد ولا تنمو بشكل أسي، بل تقل تدريجيا، وعليه فإن وقت الوباء المتوقع هو صفر أي ينتهي الوباء بنهاية هذه الموجة. وبالنظر إلى قيمته خلال فترة الوباء كاملة نجدها 0.203، وكذلك الوقت المقدر لانتهاء الوباء هو 3 أيام فقط أي ينتهي الوباء في 3/05/2023.

ومن الملاحظ أيضا أن قيمة تجاوز الحد Overshoot خلال الموجات الأربعة وكذلك خلال فترة الوباء كاملة مساوية للصفر، مما يشير الي أن النسبة التراكمية للأفراد المصابة لم تتجاوز الحد الأدنى (النسبة اللازمة) المطلوب لتحقيق المناعة الجماعية، وذلك يوضح لنا أنه وفق تقديرات النموذج تم السيطرة على الوباء بفعالية قبل أن ينتشر بشكل واسع بين السكان وأن التدابير الوقائية كانت فعالة للغاية في كبح انتشار الوباء.

ملحوظة: تم افتراض أن فعالية اللقاح $E = 100\%$، لذا فإنه لا يوجد فرق بين تقديرات مستوي المناعة الجماعية HIT، ومستوي التطعيم المطلوب V_c فكلاهما واحد.

كما تم تقدير المتغيرات (S, I, R, D) بشكل يومي خلال الموجات الأربعة، ومنه استنتجنا الإحصاءات الوصفية المقدرة للمتغيرات بواسطة نموذج SIRD:

جدول (٤) يوضح الإحصاءات الوصفية المقدرة للمتغيرات بواسطة نموذج SIRD

SIR	Wave 1			Wave 2			Wave3		
Measures	I	R	D	I	R	D	I	R	D
Mean	402.88	370.41	24.24	731.66	457.04	44.19	556.05	568.23	30.86
Median	175.997	157.42	14.95	641.37	440.17	48.63	609.74	565.16	32.77
Max	1769.25	1806.03	100.50	1841.76	1236.4	67.51	1300.95	2434.48	71.98
Min	0	0	0	230.68	98	11.63	0	34.36	0
Sd	478.92	404.62	25.18	288.03	158.30	14.75	385.77	302.97	20.67
Sum	113612.82	104455.76	6836.60	95115.25	59415	5744.49	101757.42	103424.04	5648.24

SIR	Wave 4			Total		
Measures	I	R	D	I	R	D
Mean	368.42	364.57	13.80	445.94	440.063	22.36
Median	0	358	0	169.01	412.13	13.696
Max	5516	3975.89	139.51	5516	3975.89	139.51
Min	0	0	0	0	0	0
Sd	664.87	343.17	23.02	565.24	345.88	24.58
Sum	212209.54	209994.696	7949.196	522192.16	477289.5	26178.54

ووفقا لتقديرات النموذج، يشير الجدول السابق الى أن البيانات لا تتبع التوزيع الطبيعي كما أنها ملتوية لليمين، حيث نجد أن قيمة المتوسط أكبر من قيمة الوسيط بالنسبة للمتغير I، وهذا يشير الي أن هناك أياما قليلة تشهد ارتفاعا كبيرا في عدد الإصابات، مما رفع قيمة المتوسط عن الوسيط.

كما نلاحظ ارتفاع قيمة المتوسط عن قيمة الوسيط بالنسبة للمتغير R، وهذا يعني أن توزيع حالات الشفاء ملتوٍ تجاه اليمين، كما يشير إلى وجود أيام استثنائية قليلة شهدت ارتفاعا كبير في حالات الشفاء مقارنة بباقي الأيام، بينما تبقي أغلب الأيام عند مستويات تعاف أقل نسبيا، مما جعل المتوسط أكبر من الوسيط.

كما نجد ارتفاع قيمة الوسط عن الوسيط بالنسبة للمتغير D خلال الموجتين الأولي والرابعة وكذلك خلال فترة الوباء كاملة، أي يوجد التواء لليمين والعكس بالنسبة للموجتين الثانية والثالثة، يوجد التواء سالب.

ثانيا: بالنسبة لنموذج MLP

من خلال نتائج مجموعة التدريب تم تقدير المتغيرات (I, R, D) بشكل يومي من الفترة 14/02/2020 حتى 30/04/2023. ومنه استنتجنا الإحصاءات الوصفية المقدرة للمتغيرات بواسطة نموذجMLP كما يلي:

جدول(5) يوضح الإحصاءات الوصفية المقدرة للمتغيرات بواسطة نموذج MLP

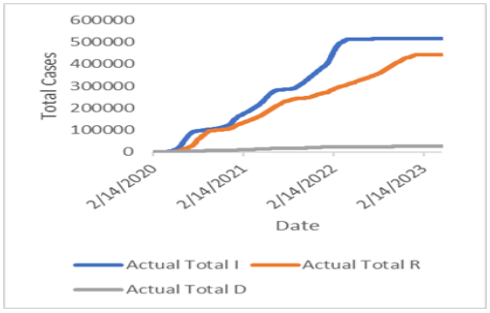
MLP	Wave 1			Wave 2			Wave 3		
Measures	I	R	D	I	R	D	I	R	D
Mean	436.71	387.05	22.30	808.5	459	42.5	623	524	31
Median	244.5	237	15	805.27	442.86	39.42	593.92٢	504.162	28.562
Max	1669	1269	81	1309	769	59	1238	897	61
Min	5	90	1	360	171	13	0	131	2
Sd	453.26	291.31	20.698	231.21	139.71	11.92	393.84	209.01	17.62
Sum	123153	109149	6289	104685	57571	5124	108688	92262	5227

MLP	Wave 4			Total		
Measures	I	R	D	I	R	D
Mean	330.03	385	14.34	449.62	538.76	27.535
Median	5	374.24	2	239.5	417	14
Max	2250	1388	95	2250	1388	95
Min	0	89	1	0	89	1
Sd	517.81	224.05	20.16	485.35	237.14	20.87
Sum	190429	215934	8272	526955	474916	24912

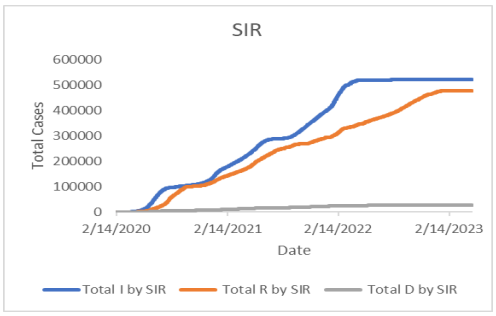
وفقا لتقديرات النموذج يتضح أن البيانات لا تتبع التوزيع الطبيعي كما أنها ملتوية لليمين، وبمقارنة نتائج الإحصاءات الوصفية في هذا الجدول بنظيرها في جدول (4) الخاص بنموذجSIR ، وكذلك جدول (2) الخاص بالقيم الفعلية، نجد أن نتائج نموذج MLP لا تختلف كثيرا عن جدول (2) وهذا يتوافق كثيرا مع الإحصاءات الوصفية للبيانات الفعلية، وعلى الرغم من عدم الاختلاف يبقي المعيار الأساسي في المقارنة بين النموذجين MLP،SIR هو مقاييس دقة التنبؤ.

المقارنة بين النماذج

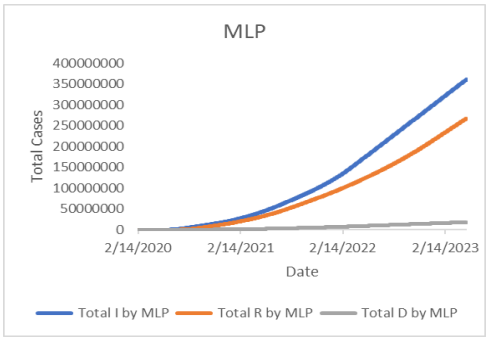
يعرض الشكل (١)، والشكل (٢)، والشكل (٣) الرسم البياني للقيم الفعلية والمقدرة والمتنبئ بها للمتغيرات الثلاثة (I, R, D) بواسطة نموذج SIR، ونموذج MLP. ومنه تم التنبؤ بعدد الحالات النشطة كما هو موضح في الجدول (٦)، وكذلك يوضح الشكل (٤) الفرق بين الحالات النشطة الفعلية والمقدرة بواسطة النماذج المستخدمة.



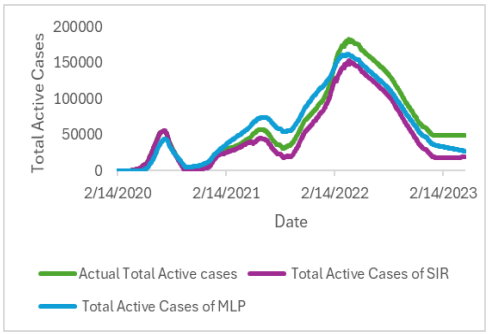
شكل (١) إجمالي حالات الإصابة، والشفاء، والوفاة الفعلية لوباء COVID – 19 في مصر



شكل (٢) إجمالي حالات الإصابة، والشفاء، والوفاة المقدرة بواسطة نموذج SIR لوباء COVID – 19 في مصر



شكل (3) إجمالي حالات الإصابة، والشفاء، والوفاة المقدرة بواسطة نموذج MLP لوباء COVID – 19 في مصر



شكل (٤) إجمالي الحالات النشطة الفعلية والمقدرة بواسطة نموذج SIR، ونموذج MLP لوباء COVID – 19 في مصر

جدول (٦) الحالات النشطة الفعلية والمتنبئ بها بواسطة نموذج MLP ونموذج SIR

	Actual		SIR		MLP	
Waves	**Peak Active Cases**	**Peak Active Cases Date**	**Peak Active Cases Estimation**	**Peak Active Cases Date Estimation**	**Peak Active Cases Estimation**	**Peak Active Cases Date Estimation**
A	**1295**	**6/19/2020**	**1282.44**	**6/20/2020**	**1071**	**6/16/2020**
B	**973**	**12/25/2020**	**1044.5**	**12/31/2020**	**818**	**12/26/2020**
C	**854**	**5/25/2021**	**1101.93**	**9/30/2021**	**672**	**5/9/2021**
D	**4803**	**3/19/2022**	**4803**	**3/20/2022**	**1483**	**3/27/2022**
Total	**49065**		**18724.12**		**27127**	

يمكن حساب معايير قياس دقة التنبؤ من خلال القيم الفعلية والقيم المتنبئ بها للمتغيرات الثلاث وفقا لنموذج SIR ونموذج MLP. وبالنسبة لنموذج MLP يتم قياس دقة النموذج باستخدام معايير قياس دقة الأداء على مجموعة الاختبار، هذه البيانات التي لم يستخدمها النموذج أثناء التدريب، وبالتالي تعطى مؤشرا حقيقيا على أداء النموذج في التنبؤ ببيانات جديدة. أما جودة توفيق البيانات يتم قياسها من خلال مجموعة التدريب. كما هو موضح في الجدول التالي:

جدول (٧) مقارنة معايير قياس دقة التنبؤ لنموذج SIR، ونموذج MLP

	SIR			MLP		
Measures	**I**	**R**	**D**	**I**	**R**	**D**
R^2	**0.75**	**0.57**	**0.84**	**0.98**	**0. 97**	**0.94**
R	**0.86**	**0.75**	**0.91**	**0.988**	**0.986**	**0.9699**
MSE	**109.05**	**13843.96**	**2.51**	**37.796**	**0.0007**	**0.22**
RMSE	**10.44**	**117.66**	**1.58**	**6.15**	**0.026**	**0.47**
MAE	**14.17**	**24.49**	**184.69**	**3.72**	**13.33**	**5.56**
MAPE	**7.38**	**101.93**	**1.41**	**1.37**	**0.003**	**0.11**

ونلاحظ من الجدول (٧) انخفاض معايير المقارنة (MEA، MSE، RMSE،MAPE) الخاصة بقياس خطأ ودقة التنبؤ لنموذج MLP مقارنة بنموذج SIR للثلاث متغيرات على الترتيب (I, R, D) مما يؤيد أفضلية نموذج MLP.

وبالنسبة لقيمة معامل التحديدR^2 وهي خاصة بجودة توفيق البيانات، فنجد أن نموذج MLP هو الأفضل من حيث جودة التوفيق لبيانات المتغيرات الثلاث وقيمته على الترتيب (0.98, 0.97, 0.94)

وهي الأقرب من 100% مقارنة بنموذج SIR حيث يتضح من الجدول أن نموذج MLP هو الأفضل والأكثر ملائمة للبيانات.

ملحوظة:

الشبكة العصبية نفسها لا يمكنها تقدير معدلات الإصابة والشفاء والوفاة، ولكنها تقوم بالتنبؤ بالمتغيرات (عدد الحالات)، وبالتالي لا يمكنها تقدير المقاييس الوبائية الأخرى كرقم التكاثر الأساسي، وعتبة المناعة الجماعية، وتجاوز الحد، وغيرها. بل تقوم بعملية تعلم غير خطي بين المدخلات والمخرجات من خلال عملية التدريب للبيانات التاريخية؛ هذا يعني أنها تقوم بنمذجة هذه البيانات بناءً على الأنماط الموجودة فيها دون استخدام معادلات صريحة، كما هو الحال في النماذج الرياضية التقليدية، مما يجعل من الصعب ربط المخرجات بمعادلات معينة، مثل تلك المستخدمة في حساب معدلات الإصابة والشفاء والوفاة، ولكن يمكن حساب هذه المعدلات يدويًا بالاعتماد على نتائج الشبكة المستخرجة.

بناءً على ما سبق، نستنتج أن الشبكات العصبية الاصطناعية قادرة على التنبؤ بالقيم التي يمكن استخدامها لاحقاً لتقدير المعدلات يدويًا، لكنها لا توفر معادلات رياضية تقليدية مثل النماذج الرياضية (الإحصائية)، فبدلاً من ذلك تقوم الشبكات العصبية بالتعلم من البيانات وتقديم تنبؤات بناءً على الأنماط المعقدة التي تكتشفها، وللتغلب على هذا التحدي، يمكن دمج الشبكات العصبية مع الأساليب الإحصائية التقليدية مما يتيح حساب المعدلات والمعلمات بشكل دقيق، وتفسير النتائج بطريقة أكثر دقة ووضوحًا.

أهم نتائج الدراسة:

يتم تقسيم نتائج الدراسة بناء على الطريقة المستخدمة حيث تناولت الدراسة أسلوبين للتنبؤ، هما أسلوب SIR، وأسلوب MLP وذلك بغرض التعرف على تأثير كل طريقة على نتائج الدراسة، وهدف التحليل هو مقارنة التقديرات وكذلك دقة وقدرة النموذجين على التنبؤ بعدد حالات الإصابة، والوفيات، والشفاء في فترات زمنية مختلفة بناءً على البيانات المتاحة، وتم التوصل إلى أن كلا النموذجين يقدمان رؤى قيمة حول ديناميكيات انتشار الوباء في مصر، حيث أظهر نموذج SIR فعالية في فهم الأنماط العامة لانتشار الفيروس، بينما تفوق نموذج MLP في تقديم توقعات دقيقة اعتمادًا على البيانات التاريخية. ويمكن توضيح ذلك من خلال مجموعة من النقاط:

أولاً: بالنسبة لنموذج SIR

١) هو نموذج رياضي يُستخدم على نطاق واسع في تحليل انتشار الأمراض المعدية. يعتمد على المعادلات التفاضلية لتقسيم المجتمع إلى ثلاث فئات رئيسية هي: المعرضين للإصابة، والمصابين، والمتعافين.

٢) على الرغم من قدرته على تقديم صورة عامة عن انتشار الوباء، إلا أن نتائجه كانت أقل دقة في التنبؤ بالتقلبات اليومية في عدد الحالات. ومن الجدير بالذكر، أنه يعتمد على افتراضات مبسطة حول معدلات الانتقال والشفاء، مما يجعله أقل مرونة في مواجهة التغيرات السريعة في طبيعة انتشار المرض.

٣) عند مقارنة توقعاته بالبيانات الفعلية، أظهر فروقات كبيرة بين القيم المتوقعة والقيم الفعلية، خاصة في الفترات التي شهدت زيادة حادة أو تراجعًا مفاجئ في عدد الحالات.

ثانياً: بالنسبة لنموذج MLP

١) يُعد من النماذج الذكية القائمة على الشبكات العصبية الاصطناعية، ويتعلم من البيانات التاريخية لتوقع انتشار الوباء.

٢) أظهر أداءً أفضل مقارنة بنموذج SIR في التنبؤ بأعداد حالات الإصابة، والوفيات، والشفاء. وتمكن من محاكاة التقلبات اليومية بدقة أعلى، بفضل قدرته على التعامل مع البيانات المتغيرة والأنماط غير الخطية في الانتشار.

٣) علاوة على ذلك، أظهر كفاءة واضحة في التكيف مع الزيادات المفاجئة في عدد الحالات، مما يشير إلى أن النموذج يمكنه التعلم والتكيف مع أنماط انتشار الوباء المتغيرة، مما يجعله أكثر مرونة ودقة في التنبؤ.

٤) يُعد استخدام نماذج الذكاء الاصطناعي كنموذج MLP خطوة مهمة نحو تحسين التنبؤات الصحية واتخاذ القرارات المستنيرة خلال الأوبئة.

ثالثاً: مقارنة الأداء باستخدام مؤشرات الدقة

عند مقارنة أداء النموذجين باستخدام مؤشرات الدقة (MAPE, RMSE, MSE, MEA) الخاصة بقياس خطأ ودقة التنبؤ، تبين ما يلي:

١) تفوق نموذج MLP بشكل واضح على نموذج SIR في توقع الحالات اليومية.

٢) أظهر نموذج MLP قدرة أفضل على التقاط التعقيدات والتقلبات غير الخطية في البيانات، بينما اعتمد نموذج SIR على معادلات ثابتة لم تستطع التعامل مع جميع التغيرات المفاجئة في مسار انتشار الوباء.

٣) باستخدام نموذج MLP اتضح أن قيم أخطاء التنبؤ الخاصة به أقل من قيمها لنموذج SIR، مما يعكس تحسنًا كبيرًا في دقة التوقعات وقدرته على التعامل مع البيانات المتغيرة والتقلبات اليومية في الإصابات والوفيات والشفاء.

٤) انخفاض قيم مؤشرات الدقة لنموذج SIR خلال فترة التحليل، مما يشير إلى محدودية قدرة النموذج على التنبؤ الدقيق في ظل التغيرات الديناميكية في أعداد الحالات.

توصيات ومقترحات الدراسة بناء على النتائج:

بناءً على ما تقدم من نتائج التحليل لوباء COVID – 19، تتقدم الباحثة بالتوصيات والمقترحات التالية لتعزيز دقة التنبؤات المستقبلية وتحسين استجابة القطاعات المختلفة:

١) استخدام تقنيات الذكاء الاصطناعي في التقدير والتنبؤ؛ ليعطي نتائج أفضل وتصغير حجم الأخطاء إلى أقل قيمة لها.

٢) استخدام أسلوب الشبكات العصبية الاصطناعية وجميع تقنيات الذكاء الاصطناعي لعمل تنبؤات طويلة الأجل، كبديل للنماذج الرياضية التي لا تصلح إلا للتنبؤات قصيرة الأجل، وكذلك لتجنب الافتراضات المعقدة لتلك النماذج كثبات حجم المجتمع، وتجاهل التغيرات الحيوية، وتساوي فرص الإصابة لجميع الأفراد وغيرها من الافتراضات غير المتوافقة مع الواقع، مع العلم إننا لا ننكر أهمية التنبؤات قصيرة الأجل في تقديمها نظرة ثاقبة حول ديناميكيات الانتشار من حيث طبيعة الوباء، وحجمه، ومدي تأثيره.

٣) بناءً على نتائج التحليل، تُوصي باستخدام نموذج MLP كأداة أكثر فعالية للتنبؤ بتطور الوباء في المستقبل في مصر، خاصةً في الحالات التي تشهد تغييرات ديناميكية في أعداد الحالات.

٤) يمكن لنموذج MLP أن يكون جزءًا من أنظمة التحذير المبكر؛ لتوجيه القرارات المتعلقة بالإجراءات الاحترازية وتوزيع الموارد الصحية.

٥) على الرغم من أن نموذج SIR يوفر فهمًا عامًا لديناميكيات الوباء، إلا أنه يحتاج إلى تحسينات أو دمج مع نماذج أخرى مثل نموذج MLP لزيادة دقته.

٦) إجراء العديد من الدراسات باستخدام تقنيات الذكاء الاصطناعي والتعلم الآلي، والتي ستساعد الكثير من الأشخاص في العديد من القطاعات وخاصة القطاع الصحي، على الاستعداد الجيد واتخاذ الاحتياطات اللازمة لتقليل انتشار العديد من الأمراض والأوبئة.

٧) ضرورة تدريس النماذج الرياضية للأمراض المعدية وتقنيات الذكاء الاصطناعي في قسم الإحصاء، حيث إن الذكاء الاصطناعي حالياً أثبت كفاءته مقارنةً بالعديد من النماذج الرياضية والاحصائية؛ لذا يمكن لقسم الإحصاء الاستفادة من ذلك في تدريس الطلاب كلا المنهجين، الأمر الذي يودي إلى ابتكار العديد من النماذج المدمجة. ومن ثم إنشاء قنوات اتصال فعالة بين مراكز تحليل البيانات الصحية والهيئات الحكومية والمؤسسات الصحية، لتنفيذ توقعات الأبحاث المطبقة مما يضمن اتخاذ القرارات الصحية بناءً على بيانات دقيقة وموثوقة، مما يُقلل من التأخر في الاستجابة للأزمات الصحية.

٨) نظراً للطبيعة الديناميكية للأوبئة، فإننا دائماً بحاجة إلى نماذج مبتكرة؛ لعمل تنبؤات قصيرة، ومتوسطة، وطويلة الأجل. ونماذج الذكاء الاصطناعي أثبتت جودتها، ولكننا دائما نطمع فيما هو أفضل، حيث إن استخدامها للتنبؤ لا يمكنه التقاط نمط تغير الوقت لانتشار الأمراض المعدية، ولحل هذه المشكلة يُفضل استخدام النماذج المدمجة أي دمج نماذج الذكاء الاصطناعي بالنماذج الرياضية، كاستخدام نموذج SEIRVD – DL، الذي يجمع بين النموذج الرياضي SEIRVD، وأحد نماذج التعلم العميق DL.

يمكن تقدير المعلمات الوبائية في النموذج الرياضي باستخدام أحد نماذج DL مثل: نموذج شبكة LSTM (long – short term memory) بدلاً من طرق التقدير التقليدية للتقدير مثل: OLS، MLE. فلقد وُجد أنه مناسب للتنبؤات قصيرة ومتوسطة وطويلة الأجل، وهذا أكثر عملية لصياغة السياسات والتخطيط الاستراتيجي، وذلك باستخدام البرامج الإحصائية ولغات البرمجة، أهمها في هذا الصدد: R, MATLAB, Python (Wang, X., Lin, X., & Dang, X., 2020)،(Liao, Zhifang, et al., 2021).

٩) مواصلة تطوير وتحديث نماذج التنبؤ؛ لضمان دقة أكبر وقدرة على التكيف مع المتغيرات ومرونتها في مواجهة التحديات المستقبلية. مواصلة تطوير وتحديث نماذج التنبؤ؛ لضمان دقة أكبر وقدرة على التكيف مع المتغيرات ومرونتها في مواجهة التحديات المستقبلية.

المراجع

أولا: المراجع باللغة العربية:

١) خضاري، حنان(2023) ، "مقارنة الشبكاتُ العصبية ذاتُ الذاكرة قصيرةِ المدى المطولة ونماذج بوكس وجينكنز."، المجلة العلمية للبحوث التجارية (جامعة المنوفية)، ٤٩(٢)، ٥١٧-٥٤٢.

٢) سهاد علي شهيد التميمي، رواء صالح محمد (2020)، "النماذج الوبائية ذو منظومة المعادلات التفاضلية المتعددة."، Journal of Administration and Economics 126, 358-362.

٣) الشرقاوي، ولاء واخرون (2021)، **"نهاية الموجة الثالثة لفيروس كورونا في مصر "دراسة استكشافية (الإصدار الثالث)**، الجهاز المركزي للتعبئة العامة والإحصاء في مصر.

٤) الشوادفى، جمال أحمد & حجاج، عبد الوهاب السيد (2013). "الذكاء الاصطناعي وتحليل السلاسل الزمنية." المجلة العلمية لقطاع كليات التجارة بجامعة الأزهر (1)10، 42-1.

ثانيا: المراجع باللغة الإنجليزية:

1) Mohajan, H. (2022)."Mathematical analysis of SIR model for COVID-19 transmission.", 1-18.
2) Jakhar, M., Ahluwalia, P. K., & Kumar, A. (2020). COVID-19 epidemic forecast in different states of India using Sir model. Medrxiv, 2020-05.
3) Wang, X., Ren, H., & Wang, A. (2022). Smish: A novel activation function for deep learning methods. Electronics, 11(4), 540.
4) Paul, Amit, et al. "An introduction to Compartmental Model and its use in Stata." (2022).
5) Katris, C. (2021). A time series-based statistical approach for outbreak spread forecasting: Application of COVID-19 in Greece. Expert systems with applications, 166, 114077.
6) Randolph, H. E., & Barreiro, L. B. (2020). Herd immunity: understanding COVID-19. Immunity, 52(5), 737-741.
7) Nguyen, M. M. (2024). Upper bounds on overshoot in SIR models with nonlinear incidence. npj Complexity, 1(1), 11.
8) Sebbagh, A., & Kechida, S. (2022). "EKF-SIRD model algorithm for predicting the coronavirus (COVID-19) spreading dynamics." Scientific Reports 12(1), 13415.
9) Azeem, Muhammad, et al. (2023). "Neural Networks for the Detection of COVID-19 and Other Diseases: Prospects and Challenges." Bioengineering, 10(7), 850.

10) Cao, H., Wu, H., & Wang, X. (2020). Bifurcation analysis of a discrete SIR epidemic model with constant recovery. Advances in Difference Equations, 2020(1), 49.

11) Liao, Zhifang, et al. (2021)."SIRVD-DL: A COVID-19 deep learning prediction model based on time-dependent SIRVD." Computers in Biology and Medicine, 138, 104868.