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

بناء نموذج تنبؤ بحمل الرواسب المعلقة باستخدام التعلم الالي : نهر الزاب الكبير في العراق إنموذجاً

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

من قبل

لبنى جمال جاجان خورشيد

بإشراف
أ.م. بيداء سليمان بهنام

1445 هـ 2023 م

المستخلص

تحمل الأنهار الرواسب المعلقة مع تدفقها التي تترسب في أماكن مختلفة حسب تصريف النهر ومجراه ، هذه الرواسب تؤثر في الصحة البيئية والأنشطة الزراعية ومصادر المياه المحمولة. والتي تقلل من منطقة التدفق ، وبالتالي تؤثر في حركة الأحياء المائية ومما يؤدي في النهاية إلى تغيير مجرى النهر. فأن بيانات الرواسب المعلقة معلومات مهمة للجهات المعنية. تتطلب الجهات المعنية البيانات المتوقع من الرواسب المعلقة في النهر لتشغيل الهياكل الهيدروليكية المختلفة بشكل صحيح.

تعد دراسة التنبؤ بحمل الرواسب المعلقة Suspended Sediment Load (SSL) أمر بالغ الأهمية لغرض المراقبة لتقليل الخسائر والضرر المحتمل. ومن أجل الحصول على تنبؤ دقيق لقيم حمل الرواسب المعلقة أقرب ما تكون لقيم حمل الرواسب المعلقة الحقيقية، تم في هذه الرسالة استخدام خوارزمية التعلم الآلي الغابة العشوائية Random Forest (RF) وشبكات التعلم العميق شبكة الذاكرة طويلة قصيرة المدى Long short Term Memory (LSTM) ، وشبكة الذاكرة طويلة قصيرة المدى ثنائية الإتجاه Bidirectional Long short Term Memory (Bi_LSTM)، وحدة البوابات المتكررة Gated Recurrent Unit (GRU) للتنبؤ بحمل الرواسب المعلقة الشهرية SSL على مجموعة بيانات مستجمعات مياه روافد نهر الزاب الكبير في العراق للمدة من 1981–2008.

تم تدريب النماذج المقترحة وإختبارها بأستخدام سبعة سيناريوهات تضمنت مدخلات مختلفة لهذه البيانات بهدف إيجاد أفضل نموذج تنبؤي وأدقه بحمل الرواسب المعلقة لهذا النهر. وتم إجراء سلسلة من المعالجات المسبقة على البيانات عن طريق تطبيعها وتحضيرها للنماذج وبعد ذلك التنبؤ بحمل الرواسب المعلقة للشهر الحالي وتم حساب مقاييس التقييم (R^2، MAE، MSE، RMSE ، RAE، RSE، CE) بالاعتماد على قيم حمل الرواسب المعلقة المتنبأ بها وقيمها الحقيقية والمقارنة بين هذه النماذج لإيجاد أفضل نموذج.

حيث تم اجراء مقارنتين اعتماداً على هذه المقاييس لإيجاد أكفاء وأفضل نموذج للتنبؤ بحمل الرواسب المعلقة وهما :

- المقارنة الأولى : تمت بين نماذج كل خوارزمية على حدا اعتماداً على السيناريوهات السبعة لايجاد أفضل إنموذج لتلك الخوارزمية وكان أفضل إنموذج للخوارزميتين Bi_LSTM و GRU في السيناريو الخامس (مدخلاته قيم التصاريف للشهر الحالي، قيم التصاريف للشهر السابق، قيم التصاريف قبل شهرين). أما LSTM فقد كان في السيناريو السادس (مدخلاته قيم التصاريف للشهر الحالي، قيم التصاريف للشهر السابق، قيم التصاريف قبل شهرين، قيم الرواسب للشهر السابق). وأخيراً أفضل إنموذج لخوارزمية RF كان في السيناريو الرابع (مدخلاته قيم التصاريف للشهر الحالي، قيم التصاريف للشهر السابق، قيم الرواسب للشهر السابق).
- المقارنة الثانية : تمت بين أفضل إنموذج لكل خوارزمية من الخوارزميات الأربعة، فقد تبين أن أفضل نموذج تنبؤي وادقه بحمل الرواسب المعلقة هو نموذج Bi_LSTM في السيناريو الخامس، حيث حصل على أعلى معامل تحديد R^2=0.885، وأدنى MAE = 0.024، RMSE= 0.040 ،MSE=0.001 ،RAE = 0.26 ،RSE = 0.11، وأعلى CE= 0.89 يليه نموذج GRU ثم LSTM وأخيراً RF .

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

Building a Model to Predict Suspended Sediments Load Using Machine Learning: The Great Zab River in Iraq as a Sample

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

By

Lubna Jamal Chachan Khurshid

Supervised by

Asst. Prof. Baydaa Sulaiman Bahnam

Assistant Professor

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

Abstract

Rivers carry suspended sediments with their flow that are deposited in different places depending on the river's drainage and course. These sediments affect environmental health, agricultural activities, and portable water sources. Which reduces the flow area, thus affecting the movement of aquatic organisms and ultimately changing the course of the river. Suspended sediment data is important information for concerned parties. Concerned authorities require the expected data of suspended sediments in the river to operate the various hydraulic structures properly.

Suspended Sediment Load (SSL) prediction study is crucial for monitoring purpose to reduce potential losses and damage. In order to obtain an accurate prediction of the suspended sediment load values that are as close as possible to the real suspended sediment load values In this thesis, the machine learning algorithm Random Forest (RF) and deep learning networks were used, the Long Short Term Memory (LSTM) network, the Bidirectional Long Short Term Memory (Bi_LSTM) network, and the Gated Recurrent Unit. (GRU) to predict monthly suspended sediment load (SSL) on a dataset of tributary watersheds of the Greater Zab River in Iraq for the period 1981–2008.

The proposed models were trained and tested using seven scenarios that included different inputs to this data, with the aim of finding the best and most accurate predictive model for the suspended sediment load for this river. A series of pre-processing was performed on the data by normalizing it and preparing it for the models, after which the suspended sediment load was predicted for the current month. Evaluation metrics (R2, MAE, MSE, RMSE, RAE, RSE, CE) were calculated based on the predicted suspended sediment load values and their real values. And compare these models to find the best model.

Two comparisons were made based on these metrics to find the most efficient and best model for predicting the load of suspended sediments:

• The first comparison: It was made between the models of each algorithm separately based on the seven scenarios to find the best model for that algorithm. The best model for the two algorithms was Bi_LSTM and GRU

in the fifth scenario (Its inputs are current month discharge values, previous month discharge values, two months ago discharge values). As for the LSTM, it was in the sixth scenario (its inputs are the discharge values for the current month, the discharge values for the previous month, the discharge values two months ago, and the sediment values for the previous month). Finally, the best model for the RF algorithm was in the fourth scenario (its inputs are the current month's discharge values, the discharge values for the previous month, and the sediment values for the previous month).

• The second comparison: It was made between the best model for each of the four algorithms. It turned out that the best and most accurate model predicting suspended sediment load is the Bi_LSTM model in the fifth scenario, as it obtained the highest determination coefficient R2 = 0.885, the lowest MAE = 0.024, RMSE = 0.040, MSE = 0.001, RAE = 0.26, RSE = 0.11, CE = 0.89 followed by the GRU model, then LSTM and finally RF.