گزارش کوتاه علمی

نشریه پژوهش‌های حفاظت آب و خاک
جلد بیست و پنجم، شماره پنجم، ۱۳۹۷
http://jwsc.gau.ac.ir
DOI: 10.22069/jwsc.2018.12334.2691

پیش‌بینی فرم بستر رودخانه‌های ماسه‌ای با استفاده از روش درخت تصمیم

مسعود کرباسی[1]* و نگین میرمرسلی[2]

[1]دانشیار گروه مهندسی آب، دانشگاه زنجان، [2]دانشجوی کارشناسی‌ارشد مهندسی سازه‌های آبی، دانشگاه زنجان

تاریخ دریافت: ۹۶/۶/۱۲ ؛ تاریخ پذیرش: ۹۷/۵/۲۹

چکیده

سابقه و هدف: فرم بستر یا به‌عبارتی ناهمواری‌های بستر به شکل‌های مختلف در بستر رودخانه اطلاق می‌شود که در اثر حرکت جریان به وجود می‌آید و تأثیر مستقیم و مهمی روی زبری بستر و در نتیجه مقاومت در مقابل جریان و تأثیر روی پروفیل سطح آب را در پی دارد. از آن‌جا که محاسبات دبی- اشل رودخانه و سرعت جریان کاملاً تحت‌تأثیر زبری قرار دارد، بنابراین پیش‌بینی دقیق شکل بستر از اهمیت زیادی برخوردار است. به‌دلیل تأثیر پارامترهای مختلف در شکل‌گیری فرم بستر رودخانه‌ها، تعیین معادلات حاکم بر آن مشکل بوده و مدل‌های ریاضی نیز از دقت کافی برخوردار نیستند. امروزه استفاده از سیستم‌های هوش مصنوعی به‌عنوان راهکاری جدید در تحلیل مسائل آبی، گسترش یافته است. هدف این پژوهش، معرفی روشی است که با استفاده از آن بتوان فرم بستر رودخانه‌های ماسه‌ای را با دقت بالایی پیش‌بینی نمود.

مواد و روش‌ها: در پژوهش حاضر، برای به‌دست آوردن نتایج بهتر و کاهش پراکندگی داده‌ها، داده‌ها به‌طور تصادفی به دو بخش آموزش (۷۰ درصد) که شامل ۱۶۴۷ داده آزمایشگاهی و آزمون (۳۰ درصد) که شامل ۵۶۰ داده آزمایشگاهی است تقسیم شدند. روش هوشمند درخت تصمیم بر روی داده‌های بخش آزمون در محیط برنامه‌نویسی وکا کدنویسی شد و در نهایت با استفاده از الگوریتم‌های (Random Forest) و (Random Tree) واسنجی بر روی داده‌ها انجام گردید. سپس روش‌های تجربی وان‌راین، انگلند هانسن و سیمونز و ریچاردسون بر روی داده‌های بخش آزمون اجرا گردید.

یافته‌ها: ارزیابی نتایج به‌دست آمده با استفاده از معیارهای آماری مجذور میانگین مربعات خطا (RMSE)، نرخ دسته‌بندی صحیح (CCI) و مساحت زیر منحنی (ROC Area) انجام شده است. نتایج نشان داد که الگوریتم (Random Forest) برای داده‌های آزمایشگاهی با معیارهای آماری CCI=٪۸۵ درصد، RMSE=۰/۱۷، ROC=۰/۹۷ دارای بهترین عملکرد است. از سوی دیگر با بررسی نتایج روش‌های تجربی مشخص شد که برای داده‌های

* مسئول مکاتبه: m.karbasi@znu.ac.ir

آزمایشگاهی، روش وان‌راین با نتایج CCI=٦٤٪ درصد، RMSE=۱/۰۷ دارای عملکرد بهتری می‌باشد. بین متغیرهای مختلف محیطی دبی، عرض، عمق، شیب، قطر متوسط ذرات رسوبی و دما برای داده‌های آزمایشگاهی دارای بیش‌ترین اهمیت در پیش‌بینی فرم‌های بستر بودند.

نتیجه‌گیری: در این پژوهش برتری مدل‌های محاسباتی نرم در مدل‌سازی و پیش‌بینی فرم بستر مشهود بوده و مدل‌های اجرا شده در محیط وکا عملکرد بهتری داشتند. اصولاً از آن‌جا که در شکل‌گیری فرم بستر رودخانه‌ها، عوامل متعددی دخالت دارند و همچنین به‌دلیل ماهیت پیچیده آن، پیش‌بینی این پدیده بسیار دشوار و گاهی کم‌دقت است. از آن‌جا که روش‌های هوش مصنوعی برای تحلیل مسائلی به‌کار می‌روند که شناخت و توصیف صریح از ماهیت مسأله وجود ندارد، بنابراین بسیاری از مسائل مربوط به فرم‌های بستر را می‌توان با این روش‌ها حل نمود.

واژه‌های کلیدی: پیش‌بینی، درخت تصمیم، رودخانه‌های ماسه‌ای، فرم بستر، هوش مصنوعی

مقدمه

یکی از مهم‌ترین مسائل در بحث انتقال رسوب نحوه ایجاد شکل بستر در رودخانه است. ایجاد و توسعه شکل بستر ناشی از اثر متقابل و پیچیده سه عامل ساختار تلاطم و آب‌شستگی جریان، انتقال رسوب و مورفولوژی بستر جریان می‌باشد (۱). پیش‌بینی دقیق هندسه فرم‌های بستر، یک جزء ضروری برای برآورد مقاومت در برابر جریان و شرایط جریان است (۷).

فرم بستر ابتدا توسط گیلبرت در سال ۱۹۱٤ عنوان شده و پس از آن مطالعات زیادی در این خصوص انجام شده است (۱۱). نیل (۱۹٦۹) دریافت که تغییرات بستر رودخانه در یک بازه کوتاهی از رودخانه در فواصل زمانی طولانی حاصل شده‌اند (۸). ون‌شن و همکاران (۱۹۹۰) دریافتند که مجموع مقاومت جداره و مقاومت بستر برابر مقاومت کل است. از سوی دیگر، می‌توان مقاومت بستر را از روی اشکال فرم‌های بستر پیش‌بینی نمود (۱۲).

غلامی و همکاران (۲۰۱۵) به ارزیابی و مقایسه مدل‌های ماشین بردار پشتیبان و معادلات رگرسیونی برای پیش‌بینی هندسه بستر رودخانه‌های رسوبی پرداختند. مقایسه مدل‌ها بیانگر عملکرد بهتر مدل (SVM) بر روی معادلات رگرسیونی بوده و باعث کاهش میانگین خطای نسبی مطلق می‌گردد (٤). جوادی و همکاران (۲۰۱۵) به برآورد ابعاد بستر رودخانه با استفاده از مدل‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان پرداختند. نتایج حاصل از این دو روش با روش تجربی وان‌راین مقایسه شدند. مقادیر بالای پارامترهای آماری بیانگر عملکرد بهتر مدل (SVM) با کرنل (RBF) برای پیش‌بینی فرم بستر نسبت به سایر روش‌ها می‌باشد. هم‌چنین مدل (SVM) نسبت به مدل (ANN) توانایی بالایی در تخمین و شبیه‌سازی ارتفاع فرم بستر دارد (٦).

بررسی منابع نشان داد که تاکنون مطالعه‌ای در مورد استفاده از روش درخت تصمیم برای پیش‌بینی فرم‌های بستر در رودخانه‌های ماسه‌ای انجام نشده است. از این‌رو پژوهش حاضر با هدف ارزیابی دقت تکنیک الگوریتم‌های (Random Forest) و (Random Tree) در پیش‌بینی فرم‌های بستر انجام گرفت.

مواد و روش‌ها

داده‌های مورد استفاده: داده‌های مورد استفاده در پژوهش حاضر، داده‌های براونلی (۱۹۸۱) است (۲). برای به‌دست آوردن نتایج بهتر و کاهش پراکندگی داده‌ها، داده‌ها به‌طور تصادفی به دو بخش آموزش و آزمون که به‌ترتیب ۷۰ درصد و ۳۰ درصد داده‌ها را شامل می‌شوند تقسیم شدند. جهت بررسی کارایی مدل‌های به‌کار گرفته شده در این پژوهش از نرم‌افزار داده‌کاوی (WEKA) جهت پیش‌بینی فرم بستر استفاده گردید. پارامترهای ورودی در تمامی مدل‌ها مشابه بوده و از داده‌های بخش آزمون برای مقایسه نتایج مدل‌ها استفاده شده است. جدول ۱ متغیرهای مورد استفاده و پارامترهای آماری آن را برای داده‌های آموزش و آزمون نشان می‌دهد.

جنگل تصادفی (RF)[1]: در روش (RF) به جای رشد یک درخت، صدها یا هزاران درخت طبقه‌بندی تولید می‌شوند. در این روش یک‌سوم داده‌های اصلی در ایجاد هر درخت استفاده نمی‌شوند که تحت عنوان نمونه‌های بیرون از سبد (OOB)[2] نامیده می‌شوند. چون این نمونه‌ها در آموزش درخت استفاده نشده‌اند در آزمون مورد استفاده قرار می‌گیرند و بنابراین نیازی به اعتبارسنجی مستقل در این روش نمی‌باشد (۱۰).

درخت تصادفی (RT)[3]: درخت تصادفی (RT) یک درخت تصادفی از مجموعه درختان ممکن با (K) ویژگی تصادفی در هر گره از درخت است. منظور از تصادفی بودن این است که هر درخت شانس مساوی نمونه‌برداری را در مجموعه درختان دارد که دارای توزیع یکنواخت است (۵).

روش‌های تجربی: سیمونز و ریچاردسون (۱۹۶۶) قدرت جریان را در مقابل قطر میانه سقوط دانه‌های رسوبی برای داده‌های فلوم آزمایشگاهی و تعدادی کانال به صورت نموداری ارائه دادند (۱۱). انگلند و هانسن (۱۹۶۶) فرم بستر را تابعی از عدد فرود $Fr = \frac{V}{\sqrt{gD}}$ و نسبت $\frac{V}{U_*}$ ارائه دادند. در روش وان‌راین حرکت بار بستر توسط پارامتر بی‌بعد پارامتر ذره D_* و پارامتر مقدار حرکت T مشخص می‌شود (۱۱).

شاخص‌های آماری: به‌منظور مقایسه روابط و مدل‌های مختلف مورد بررسی از معیارهای آماری، میانگین مربعات خطا (RMSE)[4]، دقت یا نرخ دسته‌بندی (CCI)[5] و مساحت زیر منحنی (ROC Area) استفاده گردید.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(O_i - P_i)^2}{N}} \qquad (۱)$$

که در آن، N بیانگر تعداد پارامترها، O_i مقادیر مشاهداتی و P_i مقادیر برآوردی از روش‌های مختلف می‌باشند.

مهم‌ترین معیار برای تعیین کارایی یک الگوریتم دسته‌بندی دقت یا نرخ دسته‌بندی (CCI) است که این معیار دقت کل یک دسته‌بندی را محاسبه می‌کند. مودار (ROC) روشی برای بررسی کارایی دسته‌بندها می‌باشد. هرچه مقدار (ROC) مربوط به یک دسته‌بندی بزرگ‌تر باشد کارایی نهایی دسته‌بند مطلوب‌تر ارزیابی می‌شود.

1- Random Forest
2- Out Of Bag
3- Random Tree
4- Root Mean Squared Error
5- Correctly Classified Instances

جدول ۱- مقادیر عددی پارامترهای ورودی در مدل‌های مورد استفاده.

Table 1. Numerical values of input parameters in used models.

انحراف معیار Standard Deviation	میانگین Mean	حداقل Minimum	حداکثر Maximum	متغیر Variable
0.261	0.125	0.001	2.214	دبی Discharge ($\frac{m^3}{s}$)
0.671	0.858	0.076	2.438	عرض Width (m)
0.116	0.156	0.015	0.86	عمق Depth (m)
0.002	0.002	0.0001	0.021	شیب Slop (m/m)
0.0003	0.0005	0.00004	0.001	قطر متوسط ذره $\mathbf{D_{50}}$ (m)
0.240	1.40	1.05	2.42	درجه‌بندی Gradation
4.650	20.91	1.67	48	دما Temperature (°C)
4.851	4.006	0.131	42.154	نسبت بدون بعد $\mathbf{\frac{Q}{g^{1/2}D^{5/2}}}$
6.954	7.584	0.344	51.217	نسبت بدون بعد $\mathbf{\frac{W}{D}}$
0.006	0.006	0.0002	0.048	نسبت بدون بعد $\mathbf{\frac{d_{50}}{D}}$
10.389	16.068	1.182	39.975	قطر بدون بعد $\mathbf{d_{gr}}$

جدول ۲- پارامترهای ورودی در مدل‌های مورد استفاده.

Table 2. Input parameters in used models.

	ورودی Input
A	(فرم بستر، دما، درجه‌بندی، قطر متوسط ذرات، شیب، عمق، عرض، دبی) تابع = فرم بستر $\text{Bed Form} = f(\text{discharge}, \text{width}, \text{depth}, \text{slop}, D_{50}, \text{gradation}, T, \text{bedform})$
B	(فرم بستر، درجه‌بندی، شیب، نسبت بدون بعد، نسبت بدون بعد، نسبت بدون بعد) تابع = فرم بستر $\text{Bed Form} = f(\frac{Q}{g^{1/2}D^{5/2}}, \frac{W}{D}, \frac{d_{50}}{D}, \text{slop}, \text{gradation}, \text{bedform})$

نتایج و بحث

داده‌های ورودی به مدل‌ها: یک از مهم‌ترین مراحل در مدلسازی انتخاب داده‌های ورودی و ترکیب آن‌ها است. جهت به‌دست آوردن ترکیب مناسب ورودی‌های مدل، چندین ترکیب مختلف از داده‌های ورودی مورد بررسی قرار گرفت. جدول ۲ دو ترکیب مختلف از داده‌های ورودی را نشان می‌دهد. ترکیب‌های زیر به‌عنوان ورودی برای الگوریتم‌های (Random Forest) و (Random Tree) مورد استفاده قرار گرفتند. ترکیب ورودی (A) شامل پارامترهای دارای بعد و ترکیب ورودی (B) شامل پارامترهای بدون بعد می‌باشند. جدول ۳ نتایج الگوریتم‌های (Random Forest) و (Random Tree) را در پیش‌بینی نوع فرم بستر تشکیل یافته نشان می‌دهد. با توجه به جدول ۳ در هر دو الگوریتم مورد استفاده، ترکیب ورودی (A) (پارامترها دارای بعد) به‌طورکلی نتایج بهتری را نسبت به استفاده از ترکیب (B) (پارامترهای بدون بعد) ارائه کرده‌اند. در الگوریتم (RF)، تغییر پارامتر مدل که در جدول ۲ داخل پرانتز نوشته شده است، بین صفر و یک، تأثیر چندانی بر دقت پیش‌بینی مدل‌ها ندارد. بهترین نتیجه در الگوریتم (RF) با ترکیب ورودی (A) و پارامتر صفر، دقت دسته‌بندی ۸۵ درصد و مساحت زیر منحنی (ROC)، ۰/۹۷ به‌دست آمده است. در الگوریتم (RT)، استفاده از پارامتر صفر RT(0) نتایج بهتری را نسبت به مدل RT(1) ارائه کرده است. بهترین نتیجه در این الگوریتم با ترکیب ورودی A، دقت دسته‌بندی ۷۸ درصد و مساحت زیرمنحنی (ROC)، ۰/۸۵ به‌دست آمده است. مقایسه دو الگوریتم بررسی شده نشان‌دهنده الگوریتم (Random Forest) است.

جدول ۳- عملکرد روش درخت تصمیم برای پیش‌بینی فرم بستر با ورودی‌های متفاوت برای داده‌های آزمایشگاهی.

Table 3. Performance of Decision Tree method for predicting bed forms with different inputs for laboratory data.

معیارهای آماری Statistical criteria	الگوریتم Algorithm							
	RF (0)		RF (1)		RT (0)		RT (1)	
	مجموعه ورودی Input selection							
	A	B	A	B	A	B	A	B
دقت دسته‌بندی (٪) Correctly Classified Instances (%)	85	83	85	83	78	77	76	74
ریشه میانگین مربعات خطا Root Mean Squared Error	0.17	0.19	0.18	0.19	0.25	0.25	0.26	0.27
مساحت زیر منحنی ROC Area	0.97	0.95	0.97	0.95	0.85	0.83	0.83	0.82

به‌منظور بررسی دقت بهترین مدل پیشنهادی در قسمت قبل RF(0) با ترکیب ورودی (A)، مساحت زیر منحنی (ROC) مدل فوق برای هر یک از انواع فرم بسترها در جدول ٤ ارائه شده است. با توجه به جدول ٤، مدل پیشنهادی توانسته است شرایط آستانه حرکت و امواج ایستاده را با بالاترین دقت (ROCArea=۱) پیش‌بینی کند. کم‌ترین دقت نیز مربوط به شرایط انتقالی بوده است (ROCArea=۰/۹۳). مدل (RF) توانسته است فرم‌های بستر دیون و ریپل را نیز با دقت بالایی پیش‌بینی کند. با توجه به این‌که فرم‌های بستر دیون و ریپل دارای تأثیرات زیادی بر مقاومت جریان در رودخانه‌ها هستند، دقت پیش‌بینی این فرم‌های بستر دارای اهمیت زیادی است.

جدول ٤- مقادیر مربوط به انواع فرم بستر در روش درخت تصمیم برای داده‌های آزمایشگاهی.

Table 4. The values for the bed forms in the Decision Tree method for laboratory data.

معیارهای آماری Statistical criteria	الگوریتم Algorithm						
	جنگل تصادفی – ترکیب ورودی (A) Random Forest – Input composition (A)						
کلاس‌های طبقه‌بندی Classification classes	کلاس ۱ Class 1	کلاس ۲ Class 2	کلاس ۳ Class 3	کلاس ٤ Class 4	کلاس ٥ Class 5	کلاس ٦ Class 6	کلاس ۷ Class 7
مساحت زیر منحنی ROC Area	1	0.97	0.97	0.93	0.95	1	0.96
فرم بستر Bed form	آستانه حرکت Incipient Motion	ریپل Ripple	دیون Dune	انتقالی Transition	بستر صاف Plane bed	امواج ایستاده Standing Wave	آنتی‌دیون Anti Dune

به‌منظور مقایسه مدل پیشنهادی با روش‌های تجربی پرکاربرد، سه روش تجربی (ون‌راین، انگلوند و هانسن و سیمویز و ریچاردسون) مورد مقایسه قرار گرفت. جدول ٤ نتایج مربوط به دسته‌بندی روش‌های تجربی را نشان می‌دهد. با توجه به جدول ٥ روش ون راین با دقت دسته‌بندی ٦٤ درصد بهترین نتایج را ارائه کرده است. ضعیف‌ترین نتیجه مربوط به روش انگلاند- هانسن با دقت دسته‌بندی ٤١ درصد می‌باشد. همچنین مقایسه بین الگوریتم‌های درخت تصمیم و روش‌های تجربی نشان می‌دهد که الگوریتم‌های درخت تصمیم توانسته‌اند با دقت بالاتری انواع فرم‌های بستر تشکیل شده را پیش‌بینی نمایند.

جدول ٥- عملکرد روش‌های محاسباتی مختلف در پیش‌بینی فرم بستر برای داده‌های آزمایشگاهی.

Table 5. The performance of different computational methods to predicting bed forms for laboratory data.

معیارهای آماری Statistical criteria	روش‌ها Methods		
	سیمونز و ریچاردسون Simons & Richardson	انگلوند و هانسن Engelund & Hansen	وان‌راین Van Rijn
دقت دسته‌بندی (٪) Correctly Classified Instances (%)	47	41	64
ریشه میانگین مربعات خطا Root Mean Squared Error	1.62	1.56	1.07

تحلیل حساسیت: به‌منظور بررسی تأثیر متغیرهای ورودی بر مدل، تحلیل حساسیت صورت پذیرفت. بدین‌منظور در هر اجرای مدل یکی از متغیرها حذف و دقت مدل مورد ارزیابی قرار گرفت. متغیرهایی که تأثیر بیش‌تری بر دقت طبقه‌بندی داشتند به‌عنوان متغیرهایی که مدل حساسیت بیش‌تری به تغییرات آن‌ها دارد مورد ارزیابی قرار گرفتند. جدول ٦ نتایج تحلیل حساسیت مدل (RF) را به متغیرهای ورودی نشان می‌دهد. با توجه به جدول ٦ پارامتر دبی جریان بیش‌ترین تأثیر را بر کاهش دقت مدل دارد. متغیرهای شیب آبراهه و قطر متوسط ذرات (D_{50}) در رتبه‌های بعدی قرار می‌گیرند. همچنین جدول ٦ نشان می‌دهد که پارامترهای دما (T)، درجه‌بندی رسوب (Gradation) و عمق آب تأثیر چندانی بر دقت طبقه‌بندی فرم‌های بستر ندارند.

جدول ٦- تحلیل حساسیت پارامترهای ورودی در مدل جنگل تصادفی.

Table 6. Sensitivity Analysis of Input Parameters on Random Forest Model.

متغیر حذف‌شده Omitted Variable	دقت طبقه‌بندی Classification Accuracy	رتبه Rank
دبی Discharge	80.5%	1
عرض Width	84.3%	4
عمق Depth	85.9%	7
شیب Slope	81.8%	2
متوسط قطر ذرات D_{50}	83.7%	3
درجه‌بندی رسوب Sediment Gradation	84.8%	5
دما Temperature	85.5%	6

بحث

تعیین ضریب زبری رودخانه‌ها یکی از مراحل مهم در طراحی پروژه‌های مهندسی رودخانه، مدلسازی و مطالعه رودخانه‌ها می‌باشد. محاسبات عمق، سرعت و افت‌های جریان بر اساس ضریب زبری می‌باشد. فرم‌های بستر ایجاد شده در کف رودخانه‌ها تأثیر بسیار چشمگیری بر مقدار ضریب زبری دارند. تنوع انواع فرم‌های بستر و تأثیر متفاوت هر یک بر ضریب زبری، پیش‌بینی نوع فرم بستر ایجاد شده در کف رودخانه را به مسأله مهمی تبدیل کرده است. در پژوهش حاضر مدل‌های هوشمند درخت تصمیم با روش‌های تجربی مرسوم در پیش‌بینی فرم بستر در رودخانه‌های با بستر ماسه‌ای مورد مقایسه قرار گرفتند. نتایج بیانگر برتری مدل‌های درخت تصمیم بود، چنان‌چه برترین مدل تجربی وان‌راین دارای دقت دسته‌بندی ٦٤ درصد بود، در حالی‌که در بهترین مدل هوشمند الگوریتم (Random Forest) دقت دسته‌بندی به ٨٥ درصد رسیده است. بنابراین نتایج پژوهش فوق می‌تواند موجب بهبود پیش‌بینی نوع فرم بستر و در نتیجه تعیین میزان زبری گردد.

مقایسه دو نوع درخت تصمیم (Random tree) و (Random Forest) نیز نشان داد که مدل (Random Forest) دارای دقت بالاتری است. چنان‌چه در بهترین مدل (Random tree) دقت دسته‌بندی ۷۸ درصد است، در حالی‌که مدل (Random Forest) دارای دقت دسته‌بندی ۸۵ درصد است. برتری مدل (RF) به (RT) در پژوهش‌های چن و همکاران (۲۰۱۷) و یوسف و همکاران (۲۰۱۵) که در مسائل پیش‌بینی زمین‌لغزش از روش‌های مختلف درخت تصمیم استفاده کرده‌اند نیز گزارش شده است (۳ و ۱۳). قابلیت بالای روش (RF) در طبقه‌بندی را می‌توان به داشتن مزایایی مانند یادگیری روابط غیرخطی و داشتن حساسیت کم‌تر در برابر گیر افتادن در کمینه محلی ارتباط داد (۹).

نتیجه‌گیری کلی

نظر به اهمیت پیش‌بینی فرم‌های بستر در رودخانه‌های ماسه‌ای و تأثیر آن‌ها بر ضریب زبری جریان، در پژوهش حاضر عملکرد دو الگوریتم درخت تصمیم (Random Forest) و (Random tree) در پیش‌بینی نوع فرم بستر تشکیل یافته در رودخانه‌های با بستر ماسه‌ای مورد بررسی قرار گرفت و نتایج با مدل‌های تجربی مقایسه گردید. برتری الگوریتم‌های درخت تصمیم در مدل‌سازی و پیش‌بینی فرم بستر نسبت به روش‌های تجربی مشهود بوده و روش‌های تجربی به‌کار گرفته شده در این پژوهش نتایج ضعیف‌تری را ارائه می‌کنند. مقایسه دو الگوریتم درختی مورد استفاده در این تحقیق نشان دهنده برتری الگوریتم (Random Forest) بود. همچنین ترکیب ورودی (A) نسبت به ترکیب ورودی (B) دقت بالاتری را در پیش‌بینی فرم‌های بستر نشان می‌دهد. پیشنهاد می‌گردد سایر روش‌های طبقه‌بندی هوشمند نیز برای پیش‌بینی تشکیل انواع فرم‌های بستر در رودخانه‌ها مورد استفاده قرار بگیرند.

منابع

1. Amiri, S.M., and Hekmatzadeh, A.A. 2014. A Finite Volume Numerical Model for Simulating Dune Bed Form. Babool, 8th National Civil Engineering Congress, Pp: 211-220. (In Persian)
2. Brownlie, W.R. 1981. Compilation of alluvial channel data: laboratory and Field.
3. Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D.T., Duan, Z., and Ma, J. 2017. A comparative study of logistic model tree, random forest and classification and regression tree models for spatial prediction of landslide susceptibility. Catena, 151: 147-160.
4. Gholami, A., Bonakdari, H., Ebtehaj, I., and Fenjan, S. 2015. Assessment and comparing of support vector machines model and regression equation for predicting alluvial channel geometry. Mesopotamia Environ. J. 1: 56-64.
5. Gupta, Sh., Abraham, S.K., Sugumaran, V., and Amarnath, M. 2016. Fault Diagnostics of a Gearbox via Acoustic Signal using Wavelet Features, J48 Decision Tree and Random Tree Classifier. Indian, J. Sci. Technol. 33: 9. 1-8.
6. Javadi, F., Ahmadi, M.M., and Qaderi, K. 2015. Estimation of Bedform Dimension Using Artificial Neural Network (ANN) and Support Vector Machine (SVM). J. Agr. Sci. Tech. 17: 859-868.
7. Karamisheva, R.D., Lyness, J.F., Myers, W.R.C., and Sullivan, J.O. 2005. Prediction of Bed Form Height in Straight and Meandering Compound Channels. J. Water Resour. Manage.III. 80: 311-320.
8. Neill, C.R. 1969. Bed Forms in the Lower Red Deer River, ALBERTA. Amsterdam, J. Hydrol. 7: 58-85.

9. Norozi, H., Nadiri, A., Asghari, A., and Gharekhani, M. 2017. Prediction of Transmissivity of Malekan Aquifer using Random Forest Method. J. Water Soil Sci. 27: 2. 61-75. (In Persian)
10. Pahlavan Rad, M.R., Khormali, F., Toomanian, N., Kiani, F., and Komaki, Ch.B. 2015. Digital Soil Mapping Using Random Forest Model in Golestan Province. J. Water Soil Cons. 21: 6. 73-93. (In Persian)
11. Shafaee Bajestan, M. 1999. Sediment Hydraulic. Shahid Chamran University. Press. 305p. (In Persian)
12. Shen, H.W., Fehlman, H.M., and Mendoza, C. 1990. Bed Form Resistances in Open Channel Flows. California, J. Hydr. Engin. 116: 6. 799-815.
13. Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., and Al-Katheeri, M.M. 2016. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides, 13: 5. 839-856.

Gorgan University of Agricultural Sciences and Natural Resources

J. of Water and Soil Conservation, Vol. 25(5), 2019
http://jwsc.gau.ac.ir
DOI: 10.22069/jwsc.2018.12334.2691

Short Technical Report

Prediction of sand rivers bed form using decision tree

***M. Karbasi[1] and N. Mirmorsali[2]**
[1]Associate Prof., Dept. of Water Engineering, University of Zanjan,
[2]M.Sc. Student, Dept. of Water Structures Engineering, University of Zanjan
Received: 09.03.2017; Accepted: 08.20.2018

Abstract

Background and Objectives: The bed forms or in the other word, bed irregularities are structures that form due to stream flow and they have direct impact on roughness and flow resistance in sand bed rivers. Bed forms have different shapes and forms in sand bed rivers. Since river discharge and flow velocity are totally affected by roughness, accurate prediction of the shape of the bed is of great importance. Due to the influence of different parameters in the formation of the river bed form, it is difficult to determine the governing equations and the mathematical models with sufficient precise. Today, the use of artificial intelligence systems has expanded as a new way of analyzing water resources issues. The objective of this research is to introduce a method that can be used to predict the shape of the river bed with high precision.

Materials and Methods: In the present study, in order to obtain better results and to reduce the dispersion of data, the data were randomly divided into two parts of the training (70%) including 1647 laboratory data and test (30%) containing 560 laboratory data. The decision trees were coded on the data of the test section in the WEKA programming environment and finally calibration was performed on the data by using Random Forest and Random Tree algorithms. Then, the experimental methods of Van Rijn, Engelund and Hansen and Simons and Richardson were implemented on test data.

Results: Evaluation of the results were done using root mean square error (RMSE), Correctly Classified Instances and Roc area under curve. The results showed that the best performance reached by Random Forest algorithm for experimental data with statistical criteria of CCI=85 (%), RMSE=0.17, ROC=0.97. On the other hand, by examining the results of empirical methods, it was determined that for laboratory data, van Rijn method has better performance with the results of CCI=64 (%), RMSE=1.07. Among different environmental variables of discharge, width, depth, slope of the channel, average diameter of sediment particles and temperature for laboratory data were the most important parameters in predicting bed forms.

Conclusion: In this research, the superiority of soft computing models was evident compared to the empirical methods in modeling and prediction of the bed form and the models performed in the VKA environment were better. Basically, because of the formation of the river bed form is depended on several factors and also because of its complex nature, the prediction of this phenomenon is very difficult and sometimes with high errors. Since artificial intelligence methods are used to analyze issues that do not explicitly describe the nature of the problem, so many of the issues of bed form can be solved with these methods.

Keywords: Artificial intelligence, Bed form, Decision tree, Prediction, Sand rivers

* Corresponding Author; Email: m.karbasi@znu.ac.ir