علوم و مهندسی آبیاری
جلد ٤١، شماره‌ی۱، بهار ۹۷، ص. ١٦-١

بررسی کارآیی الگوریتم M5 در محاسبه حداکثر عمق چاله آبشستگی اطراف تکیه‌گاه پل

جواد ظهیری[1*] و سید محمود کاشفی‌پور[2]

۱*- نویسنده مسئول، استادیار گروه مهندسی آب دانشگاه علوم کشاورزی و منابع طبیعی رامین خوزستان. j.zahiri@ramin.ac.ir

۲- استاد گروه سازه‌های آبی، دانشکده مهندسی علوم آب، دانشگاه شهید چمران اهواز.

تاریخ دریافت: ۹۴/۲/۱۶ تاریخ پذیرش: ۹۵/۶/۲۱

چکیده

آبشستگی صورت گرفته در اطراف تکیه‌گاه‌ها و پایه‌ها پدیده‌ای است که منجر به تخریب تعداد زیادی از پل‌ها گردیده است. در این تحقیق برای شبیه‌سازی حداکثر عمق چاله آبشستگی اطراف تکیه‌گاه پل از الگوریتم درختی M5 استفاده شده است. از مزایای این الگوریتم دقت بالا، سادگی و قابل فهم بودن آن است. برای ساخت و آموزش مدل از اطلاعات آزمایشگاهی منابع معتبر استفاده شد. پارامترهای مورد استفاده برای برآورد عمق آبشستگی اطراف تکیه‌گاه شامل فاکتور شکل و طول تکیه‌گاه، عمق و سرعت جریان و اندازه متوسط رسوبات بستر بوده است که به‌صورت عدد فرود ذره، ضریب شکل تکیه‌گاه، ضریب عمق- طول تکیه‌گاه و ضریب شدت جریان به مدل درختی معرفی شدند. در مدل ساخته شده توسط الگوریتم M5 و پس از عملیات هرس کردن، ضریب عمق- طول تکیه‌گاه و ضریب شدت جریان بیشترین کاربرد را در ساختار درختی داشته و باعث تقسیم فضای مسأله به سه زیردامنه متمایز و ارائه رابطه رگرسیونی برای هر زیردامنه گشته است. برای بررسی کارآیی مدل درختی از تحلیل‌های آماری متفاوت و مقایسه آن با معادلات ارائه شده توسط محققین مختلف استفاده شد. نتیجه این مقایسه نشان‌دهنده کارآیی بالای الگوریتم M5 در محاسبه عمق آبشستگی می‌باشد. برای بررسی اهمیت هر کدام از پارامترهای ورودی بر میزان آبشستگی از تحلیل حساسیت استفاده گردید. بر این اساس ضریب عمق- طول تکیه‌گاه بیشترین تأثیر را بر میزان آبشستگی داشته و حذف این پارامتر از مدل درختی باعث افزایش مجموع مربعات خطا از ٤/٣٧ به ١٩/٤١ گشته است.

کلید واژه‌ها: تکیه‌گاه پل، الگوریتم M5 ، هرس کردن، ضریب عمق-طول تکیه‌گاه، تحلیل حساسیت.

مقدمه

طراحی و پایداری پل‌ها از جمله موضوعات مهمی بوده که در طول تاریخ ذهن بشر را همواره درگیر کرده است. از آنجا که تخریب پل‌ها می‌تواند خسارات مالی و جانی زیادی را به دنبال داشته باشد، قسمت مهمی از مطالعات مهندسی رودخانه در این بخش متمرکز شده است. یکی از عوامل مهم تخریب پل‌ها آبشستگی اطراف تکیه‌گاه‌ها و پایه‌های پل بوده که این پدیده به مرور زمان اتفاق افتاده و در مواقع سیلابی تشدید شده و ممکن است باعث تخریب کلی پل گردد. بر همین اساس تخمین دقیق میزان آبشستگی اطراف تکیه‌گاه و پایه‌های پل می‌تواند به مهندسین طراح کمک کند تا در برای مقابله با تخریب پل اقدامات لازم را انجام دهند. تاکنون مطالعات متعددی در زمینه تخمین حداکثر عمق آبشستگی در اطراف تکیه‌گاه پل صورت گرفته است که از آن جمله می‌توان به Laursen (1962 و 1963)، Gill (1970 و 1972)، Melville (1992 و 1997)، Froehlich (1989) و Richardson و Davis (1995) اشاره کرد (Melville and Coleman, 2000). از جمله تحقیقات جدید صورت گرفته در این زمینه می‌توان به Dey، Chiew و Kadam (2008)، Ballio، Teruzzi و Radice (2009) و Yorozuya و Ettema (2015) اشاره کرد. Dey، Chiew و Kadam (2008) در آزمایشات خود که در بستر متحرک انجام شد، از تکیه‌گاه‌های کوچک (طول تکیه‌گاه به عمق جریان<۱) استفاده کردند. با وجود فرم بستر ایجاد شده در طول کانال، حداکثر عمق آبشستگی در اطراف تکیه‌گاه اتفاق افتاد که این محققین برای برآورد میزان آبشستگی صورت گرفته از معادلات متداول در این زمینه استفاده کردند. از میان معادلات مورد استفاده معادله Kandasamy و Melville (1998) بیشترین دقت را دارا بوده است. Ballio، Teruzzi و Radice (2009) نیز برای بررسی تأثیر طول تکیه‌گاه بر میزان آبشستگی از شرایط آب زلال استفاده کردند. نتایج آزمایشات صورت گرفته نشان دهنده تأثیر بالای طول تکیه‌گاه بر میزان آبشستگی بوده، هر چند تأثیر این پارامتر با گذشت زمان کاهش می‌یابد. Yorozuya و Ettema (2015) نیز به بررسی آزمایشگاهی میزان آبشستگی اطراف تکیه‌گاه پل پرداختند. این محققین انواع آبشستگی صورت گرفته در اطراف تکیه‌گاه پل در مقاطع مرکب را به سه دسته تقسیم کردند که شامل آبشستگی در مقطع اصلی، آبشستگی اطراف تکیه‌گاه در دشت‌های سیلابی و آبشستگی محل تکیه‌گاه در زمان شکست آن بوده است. تحلیل نتایج آزمایشگاهی نشان داد که دو نوع اول آبشستگی مربوط به فشردگی‌های کم و نوع سوم مربوط به تکیه‌گاه‌هایی با طول زیاد می‌باشد. با این وجود ماهیت سه بعدی جریان و گردابه‌های متعدد در محل چاله آبشستگی باعث شده است که اکثر روابط ارائه شده به‌صورت تجربی یا نیمه تجربی و از

طریق تحلیل ابعادی استخراج شوند. از طرفی هر کدام از روابط ارائه شده بر اساس شرایط خاصی استخراج شده و تنها در آن شرایط قابل کاربرد هستند. بر همین اساس در سال‌های اخیر از مدل‌های داده کاوی برای برآورد میزان آبشستگی استفاده شده است که از جمله این تحقیقات می‌توان به Najafzadeh، Barani و Hessami-Kermani (2015) اشاره کرد. این محققین برای برآورد میزان آبشستگی اطراف تکیه‌گاه پل از مجموعه‌ای از مدل‌های داده کاوی از قبیل الگوریتم بهینه‌سازی جستجوی گرانشی (Gravitational Search Algorithm) و الگوریتم بهینه‌سازی ازدحام ذرات (Particle Swarm Optimization) استفاده کردند. اندازه ذرات رسوبی، مشخصات هندسی تکیه‌گاه پل و شرایط جریان نزدیک شونده به‌عنوان پارامترهای مؤثر بر مکانیزم آبشستگی در نظر گرفته شد. بر اساس مقایسه صورت گرفته میان نتایج مدل‌های مورد استفاده و معادلات موجود مشخص شد که مدل‌های داده کاوی می‌توانند با دقت بالاتری میزان آبشستگی را برآورد نمایند(Najafzadeh, Barani and Hessami-Kermani, 2015). مدل‌های داده کاوی از قبیل شبکه‌های عصبی، منطق فازی و الگوریتم‌های فراکاوشی هر چند در حل بسیاری از مسایل مهندسی و از جمله مدل‌سازی آبشستگی، مورد استفاده قرار می‌گیرند ولی به دلیل ساختار این مدل‌ها و عدم ارائه معادله برای حل مسأله مورد نظر نمی‌توان به‌عنوان جایگزین مناسب برای حل مسایل کاربردی از آن‌ها بهره جست. بر همین اساس در سال‌های اخیر استفاده از مدل‌های درختی در شاخه‌های مختلف مهندسی و از جمله مهندسی رودخانه رو به افزایش بوده است. از میان الگوریتم‌های درختی مدل M5 جزء الگوریتم‌های نوین بوده که مزایای عمده‌ای نسبت به دیگر الگوریتم‌های موجود دارد. از جمله این مزایا می‌توان به سادگی و قابل فهم بودن در کنار دقت بالای آن در مسایل کاربردی اشاره کرد. Bhattacharya و Solomatine (2006) از الگوریتم M5 برای شبیه‌سازی رسوبات در بندر رتردام استفاده کردند. بر اساس نتایج به‌دست آمده می‌توان از نتایج مدل ذکر شده برای تصمیمات کاربردی سود جست. Etemad-Shahidi و Mahjoobi (2009) از این الگوریتم برای شبیه‌سازی ارتفاع امواج رودخانه استفاده کردند. آنها علاوه بر مدل درختی از شبکه عصبی نیز برای مقایسه کارآیی مدل درختی استفاده کردند و نشان دادند که مدل درختی با دقت بیشتری می‌تواند ارتفاع امواج را شبیه‌سازی کند. Etemad-Shahidi و Ghaemi (2011) برای بررسی کارآیی الگوریتم M5 در محاسبه عمق آبشستگی اطراف مجموعه پایه‌های پل از اطلاعات آزمایشگاهی و میدانی استفاده کردند. نتیجه این تحقیق دقت بالاتر مدل درختی در مقایسه با معادلات رایج از قبیلSumer ، Fredsøe و Christiansen (1992) و Bayram و Laursen (2000) بوده است.

در این تحقیق برای شبیه‌سازی عمق چاله آبشستگی اطراف تکیه‌گاه پل از مدل درختی M5 استفاده گردید. بر همین اساس برای ساخت مدل درختی از پارامترهای عدد فرود ذره، ضریب شکل تکیه‌گاه، ضریب عمق– طول تکیه‌گاه و ضریب شدت جریان به عنوان پارامترهای ورودی برای برآورد میزان عمق آبشستگی در اطراف تکیه‌گاه استفاده شد. صفاتی که در تشکیل زیردامنه‌ها توسط مدل M5 مورد استفاده قرار گرفت، ضریب عمق جریان–طول تکیه‌گاه و ضریب شدت جریان بوده که فضای مسأله توسط مدل درختی به سه زیردامنه تقسیم شده و برای هر زیردامنه رابطه رگرسیونی ارائه گردید. برای بررسی کارآیی مدل درختی مورد استفاده از معادلات Laursen (1963)، Froehlich (1989)، Richardson و Davis (1995) و Melville و Coleman (2000) استفاده گردید(Melville and Coleman, 2000).

مدل درختی M5

درخت‌های تصمیم‌گیری نسل جدیدی از روش‌های داده‌کاوی هستند که امروزه در علوم مختلف برای استخراج قوانین و روابط حاکم بر یک پدیده مورد استفاده قرار می‌گیرند. درخت تصمیم‌گیری به شکل یک درخت وارونه بوده که ریشه آن در بالاترین قسمت و برگ‌ها در پایین آن قرار می‌گیرند. الگوریتم M5 یکی از الگوریتم‌های درختی رگرسیونی بوده که اولین بار توسط Quinlan (1992) ابداع شد و پس از آن توسط Wang و Witten (1996) توسعه و بهبود یافت. این الگوریتم فضای مسأله را به زیردامنه‌هایی تقسیم کرده و برای هر کدام از زیردامنه‌ها یک مدل خطی چند متغیره ارائه می‌دهد که تنها برای همان زیردامنه قابل کاربرد است. این الگوریتم مانند CART یک مدل درختی بوده با این تفاوت که در برگ‌های مدل رگرسیونی M5 به جای مقادیر عددی از مدل‌های خطی چند متغیره استفاده می‌شود که می‌توان از آنها به‌عنوان توابع خطی استفاده کرد. بر این اساس می‌توان M5 را به‌صورت مجموعه‌ای از توابع خطی چند ضابطه‌ای در نظر گرفت. از دیگر تفاوت‌های این مدل با مدل CART کوچک‌تر و قابل فهم‌تر بودن مدل M5 و دقت بهتر آن در مثال‌های مورد استفاده است(Quinlan, 1992). در این الگوریتم از پارامتر انحراف معیار مقادیر متغیر هدف به‌عنوان معیار اندازه‌گیری خطا در هر گره برای ایجاد شاخه در آن گره استفاده می‌شود. بدین صورت که صفتی که موجب کاهش بیشتر انحراف معیار در هر گره می‌شود به‌عنوان صفت مورد نظر برای ایجاد شاخه انتخاب می‌شود. کاهش انحراف استاندارد (Standard Deviation Reduction (SDR)، تابع خطا در الگوریتم M5 بوده و به‌صورت زیر به‌کار می‌رود:

$$\Delta error = SDR = sd(T) - \sum \frac{|T_i|}{|T|} \times sd(T_i) \qquad (۱)$$

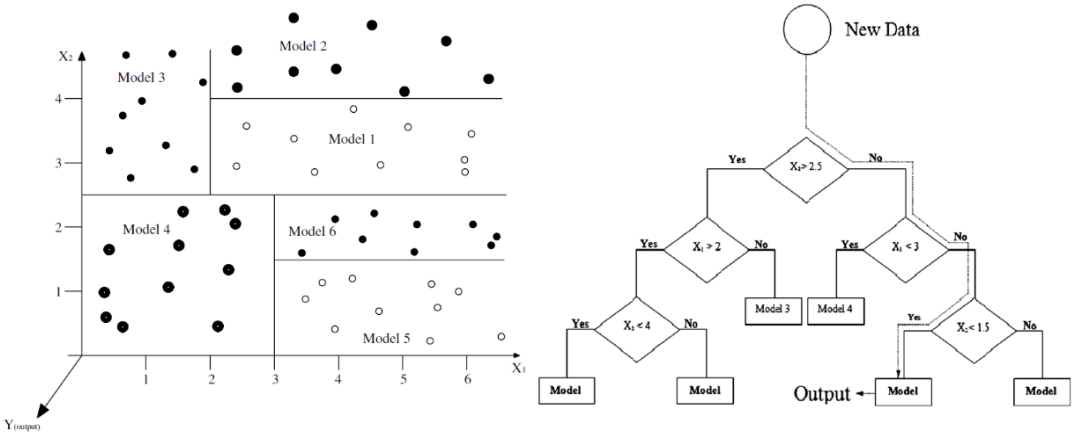
در رابطه (۱)، T شامل تمام نمونه‌هایی است که به گره مورد نظر رسیده‌اند، T_i ها نمونه‌هایی هستند که از تقسیم کردن در گره

مورد نظر بر اساس صفت انتخاب شده به دست آمده‌اند. sd نیز معرف انحراف معیار می‌باشد (Wang and Witten, 1996). الگوریتم M5 تمامی حالتهای مختلف برای ایجاد شاخه بر اساس صفت خاص را بررسی کرده و در نهایت حالتی را انتخاب می‌کند که بتواند بیشتر از حالت‌های دیگر تابع خطای فوق را افزایش دهد. پس از تکمیل درخت، برای نمونه‌های موجود در هر زیر شاخه گره داخلی، یک مدل رگرسیون خطی چند متغیری برازش داده می‌شود. این مدل، براساس داده‌های وابسته به آن گره و تمام صفاتی که در زیردرخت (Subtree) تا ریشه آن گره مورد استفاده قرار گرفته‌اند، ایجاد می‌شود. شکل(۱) نحوه تقسیم فضای مسأله به اجزای کوچکتر و کاربرد مدل‌های خطی را برای هر کدام از زیر دامنه‌ها با توجه به ساختار درختی نشان می‌دهد(Etemad-Shahidi and Taghipour, 2012).

در طی ساخت مدل درختی توسط الگوریتم M5 به مانند الگوریتم CART، فرآیند جداسازی در گره‌های تقسیم ممکن است بارها تکرار شده و در نتیجه درخت بسیار بزرگی ایجاد شود. در این حالت دقت مدل برای داده‌های آموزشی بالا رفته ولی ممکن است برای داده‌های صحت‌سنجی کاهش یابد. در این حالت مدل دچار بیش برازش (Overfitting) شده که از طریق هرس کردن (Pruning) می‌توان این مشکل را رفع کرد. در بعضی حالت‌ها هرس کردن شاخه‌های ضعیف‌تر می‌تواند به بهبود قدرت پیش‌بینی مدل کمک نماید. به عبارت دیگر هرس کردن زمانی اتفاق می‌افتد که خطای محاسبه شده برای مدل خطی در منشاء زیر درخت کوچکتر یا مساوی خطای مورد نظر در زیر درخت باشد. در این حالت زیر درخت مورد نظر به برگ تبدیل می‌شود. بر اساس تحقیقات Wang و Witten (1996) خطای پیش‌بینی برابر است با میانگین قدر مطلق مقادیر پیش‌بینی شده توسط مدل و مقادیر مشاهداتی برای هر کدام از داده‌های آموزشی که به گره مورد نظر می‌رسند. بر این اساس عمل هرس کردن باعث کاهش خطای مورد انتظار برای داده‌های غیر آموزشی می‌شود. پس از آن خطای مورد انتظار در $(n+\nu)/(n-\nu)$ ضرب می‌شود که n تعداد نمونه‌های آموزشی در گره مورد نظر بوده و ν تعداد پارامترهای موجود در مدل خطی مورد استفاده در آن گره می‌باشد(Wang and Witten, 1996). پس از هرس کردن، مدل‌های خطی مورد استفاده در برگ‌های مجاور درخت هرس شده به شدت دچار ناپیوستگی می‌گردد که این امر سبب از بین رفتن پیوستگی سیستم می گردد. این امر به‌ویژه در مورد مدل‌هایی که با استفاده از داده‌های آموزش محدودی ساخته می‌شوند شدیدتر است. بر همین اساس الگوریتم M5 از مکانیزمی به نام هموارسازی (Smoothing Process) برای رفع ناپیوستگی ایجاد شده در مدل‌های خطی استفاده می کند. در این فرآیند مقدار تخمین زده شده در هر برگ با مدل‌های موجود در مسیر ریشه تا برگ مورد نظر ترکیب شده و مدل جدید در هر برگ ساخته می‌شود. چنانچه نمونه مورد نظر در شاخه S_i از زیردرخت S باشد، n_i تعداد نمونه‌های آموزشی در S_i، $PV(S_i)$ مقدار محاسبه شده در S_i و M(S) مقدار محاسبه شده توسط مدل در S باشد، آنگاه مقدار اصلاح شده(PV(S)) از رابطه زیر محاسبه خواهد شد(Quinlan, 1992):

$$PV(S) = \frac{n_i \times PV(S_i) + k \times M(S)}{n_i + k} \qquad (۲)$$

در رابطه فوق k ثابت هموارسازی بوده که به‌صورت پیش‌فرض برابر ۱۵ در نظر گرفته می‌شود. هموارسازی به‌ویژه در مواقعی که مدل‌های خطی در برگ‌های مجاور، مقادیر کاملاً متفاوتی ارائه می‌دهند و یا مدل‌هایی که براساس داده‌های آموزشی محدود ساخته می‌شوند، می‌تواند به میزان زیادی مؤثر باشد. تحقیقات انجام شده توسط Wang و Witten (1996) نشان داده است که هموارسازی می‌تواند کارآیی الگوریتم M5 را به میزان زیادی افزایش دهد.

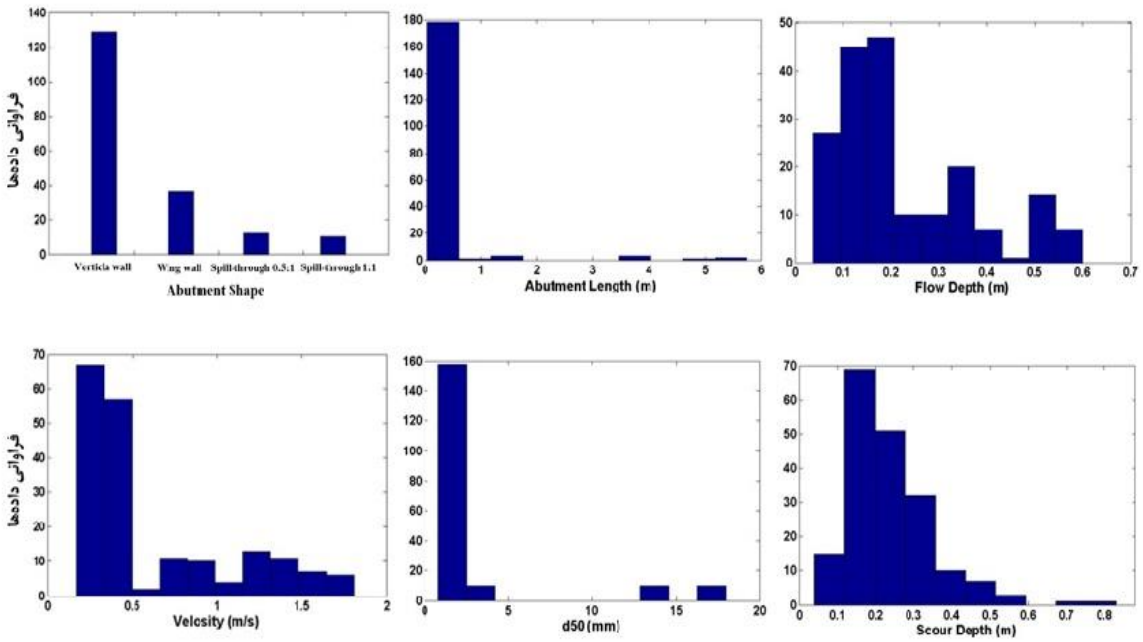


شکل ۱- تقسیم فضای مسأله و ارائه مدل‌های خطی به ازای هر کدام از فضاها توسط الگوریتم M5

مواد و روش‌ها

در این تحقیق برای ساخت و صحت‌سنجی مدل M5 از ۱۸۸ داده آزمایشگاهی که برگرفته از تحقیقات Kwan، Melville و Raudkivi (1984)، Tey (1984)، Dongol و Melville (1994)، Kuhnle ، Alonso و Shields (1999) و Coleman، Lauchlan و Melville (2003) بوده است، استفاده گردید. مجموع اطلاعات مورد استفاده شامل تکیه‌گاه‌هایی با اشکال متفاوت و شرایط هیدرولیکی متنوع بوده است. این اطلاعات مشتمل بر ۱۸۸ رکورد مربوط به عمق آبشستگی در اطراف تکیه‌گاه بوده که از منابع معتبر استخراج شده است. اطلاعات مورد استفاده شامل شکل و طول تکیه‌گاه، عمق و سرعت جریان و اندازه ذرات بستر می‌باشد. مشخصات اطلاعات آزمایشگاهی مورد استفاده در این تحقیق در ماتریس شکل (۲) ارائه شده است. در این ماتریس هیستوگرام‌های مربوط به توزیع فراوانی هر کدام از متغیرها ارائه شده است که با استفاده از آن می‌توان به توزیع متغیرها و نیز نقص موجود در داده‌ها پی برد. محور عمودی در کلیه نمودارها مربوط به فراوانی داده‌ها می‌باشد.

از مجموع اطلاعات در دسترس ۸۰٪ برای ساخت مدل‌ها و ۲۰٪ باقی‌مانده برای صحت‌سنجی مورد استفاده قرار گرفت. از آن‌جا که مدل M5 قادر به تخمین مقادیر خارج از بازه‌های معرفی شده در قسمت آموزش نیست می‌بایستی مجموعه آموزش به گونه‌ای انتخاب شود که کران بالا و پایین هر کدام از متغیرها در این مجموعه قرار گیرد. مشخصات داده‌های مورد استفاده در آموزش و ساخت مدل درختی در جدول (۱) ارائه شده است.



شکل ۲- هیستوگرام مربوط به توزیع فراوانی متغیرهای اطلاعات مورد استفاده در این تحقیق

جدول ۱- مشخصات اطلاعات مورد استفاده برای ساخت و صحت سنجی مدل M5

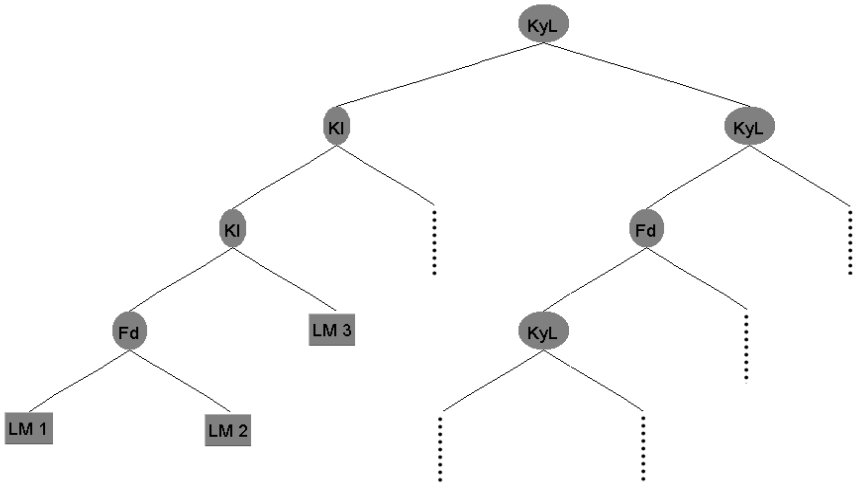
متغیر آماری	طول تکیه‌گاه (متر)	سرعت جریان (متر بر ثانیه)	عمق جریان (متر)	اندازه ذرات (میلیمتر)	عمق آبشستگی (متر)
حداکثر	۵/۷۵	۱/۸۱	۰/۶۰	۱۸/۰	۰/۸۳
حداقل	۰/۰۵	۰/۱۰	۰/۰۴	۰/۸۰	۰/۰۴
میانگین	۰/۴۴	۰/۶۲	۰/۲۳	۲/۶۵	۰/۲۴
انحراف معیار	۰/۹۰	۰/۴۵	۰/۱۵	۴/۶۱	۰/۱۲

پس از به کار بردن پارامترهای مختلف برای شبیه‌سازی عمق آبشستگی توسط مدل M5 در نهایت عدد فرود ذره (Densimetric particle Froude number (F_{rd}))، ضریب شکل تکیه‌گاه (Foundation shape factor (K_s))، ضریب عمق- طول تکیه‌گاه (Flow depth- abutment length expression (K_{yL})) و ضریب شدت جریان (Flow intensity factor (K_I)) به‌عنوان پارامترهای ورودی برای برآورد میزان عمق آبشستگی در اطراف تکیه‌گاه انتخاب شدند. عدد فرود ذره بر اساس مطالعات Hager و Oliveto (2002) نقش مهمی در حرکت ذرات رسوبی دارد، به همین دلیل به‌عنوان یکی از پارامترهای ورودی مدل درختی انتخاب شد. عدد فرود ذره از رابطه $F_{rd} = \frac{V}{\sqrt{((\rho_s - \rho)/\rho) g d_{50}}}$ محاسبه می‌شود که در آن V سرعت جریان، ρ و ρ_s به ترتیب چگالی آب و رسوبات و d_{50} اندازه رسوبات بستر می‌باشد. ضرایب K_s ، K_{yL} و K_I نیز توسط بسیاری از محققین از قبیل Melville و Coleman (2000) برای محاسبه عمق آبشستگی اطراف تکیه‌گاه مورد استفاده قرار گرفته‌اند که نحوه محاسبه این ضرایب در جدول (۲) ارائه شده است. در این جدول به دلیل پیچیدگی ضریب K_I از شکل ساده شده آن استفاده گردید.

مدل M5 تنها قادر به شبیه‌سازی مدل‌های خطی بوده که این امر با اساس معادلات متداول آبشستگی در تناقض است. بر همین اساس کلیه داده‌ها به‌صورت لگاریتم طبیعی به مدل معرفی شدند، و پس از ساخت مدل خطی از حالت لگاریتمی به توانی تبدیل شوند. مدل درختی M5 با توجه به داده‌های معرفی شده در قسمت آموزش و قبل از عملیات هرس شامل ۵۹ رابطه رگرسیونی بوده که قسمتی از مدل درختی ایجاد شده توسط این الگوریتم در شکل(۳) نشان داده شده است. پس از عملیات هرس کردن تعداد مدل‌های رگرسیونی به سه کاهش یافت که مدل درختی ایجاد شده پس از هرس کردن و خارج کردن مقادیر از حالت لگاریتمی به شکل (۴) تبدیل یافت.

جدول ۲- ضرایب مورد استفاده برای ساخت مدل درختی

پارامتر	عامل مؤثر	مقدار پارامتر			
		دیوار عمودی (Vertical wall)	دیوار بالی شکل (Wing wall)	شیب‌دار با شیب دیواره ۱ :۰/۵	شیب‌دار با شیب دیواره ۱:۱
K_s	شکل تکیه‌گاه	1.0	0.75	0.6	0.5
K_{yL}	L/y	$L/y < 1$		$K_{yL} = 2L$	
		$1 < L/y < 25$		$K_{yL} = 2\sqrt{yL}$	
		$L/y > 25$		$K_{yL} = 10y$	
K_I	V/V_c	$V/V_c < 1$		$K_I = V/V_c$	
		$V/V_c \geq 1$		$K_I = 1$	



شکل ۳- مدل درختی M5 برای برآورد عمق آبشستگی اطراف تکیه‌گاه قبل از هرس کردن

نتایج و بحث

با توجه به مدل ساخته شده توسط الگوریتم M5، همان‌گونه که در شکل(۴) نیز نشان داده شده است، مدل‌های (۱) تا (۳) روابط خطی بوده که توسط الگوریتم درختی برای هر کدام از حالت‌های مختلف توسعه یافته است. این روابط پس از تبدیل از حالت لگاریتمی، به‌صورت معادلات (۳) تا (۵) خواهند شد.

$$Model\ 1$$
$$d_s = 0.08 F_d^{0.78} K_{yL}^{0.11} K_I^{1.23} K_s^{0.27} \quad (۳)$$

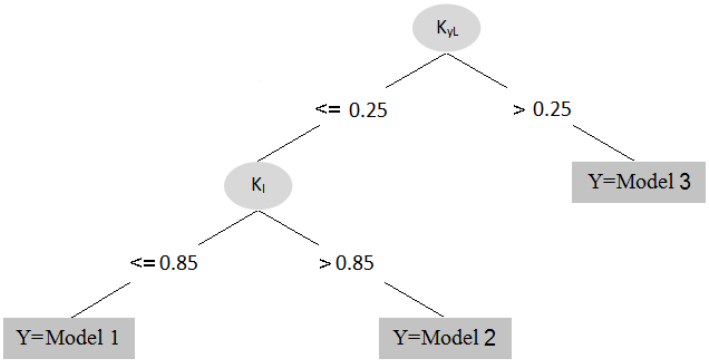
$$Model\ 2$$
$$d_s = 0.21 F_d^{0.04} K_{yL}^{0.16} K_I^{0.76} K_s^{0.27} \quad (۴)$$

$$Model\ 3$$
$$d_s = 0.67 F_d^{0.12} K_{yL}^{0.9} K_I^{1.3} K_s^{1.03} \quad (۵)$$

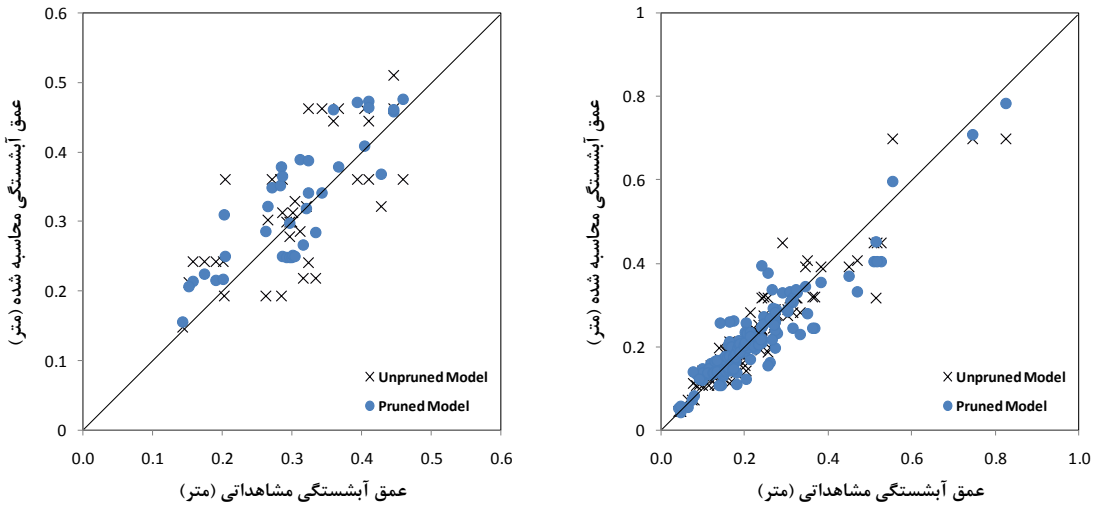
با توجه به معادلات (۳) تا (۵)، اولین متغیر برای تقسیم فضای مسأله مربوط به ضریب عمق– طول تکیه‌گاه بوده که خود تحت تأثیر نسبت طول تکیه‌گاه به عمق جریان تغییر می‌کند. از آنجا که این ضریب تأثیر هم‌زمان عمق جریان و طول تکیه‌گاه روی آبشستگی را نشان می‌دهد، بیشترین تأثیر را روی میزان آبشستگی دارد. این ضریب در اغلب مطالعات صورت گرفته به‌عنوان یکی از پارامترهای مهم و تأثیرگذار معرفی شده است (Melville and Coleman, 2000).

پس از بررسی اطلاعات مورد استفاده در قسمت آموزش مدل مشخص گردید که $K_{yL} > 0.25$ مربوط به تکیه‌گاه‌هایی با طول بین ۰/۱۵ تا ۵/۷۵ متر بوده در حالی‌که $K_{yL} \le 0.25$ مربوط به تکیه‌گاه‌هایی با طول بین ۰/۰۵ تا ۰/۳ متر می‌باشد. بنابراین مدل سه (معادله(۵))، برای تکیه‌گاه‌های با طول بزرگتر بوده و از آنجاکه تأثیر طول در این حالت روی میزان آبشستگی بیشتر است، توان K_{yL} بیشتر از دو مدل (۱) و (۲) محاسبه شده است. در مرحله بعد معیار تقسیم‌بندی بر اساس مقدار K_I بوده و مجموعه اطلاعات موجود در این قسمت به دو زیردامنه تقسیم شده و از مدل‌های (۱) و (۲) برای تخمین آبشستگی استفاده شده است. مدل (۱) که شکل حالت توانی آن به‌صورت معادله (۳) می‌باشد مربوط به داده‌های است که اندازه قطر ذرات رسوبی آن‌ها کمتر از یک میلی‌متر است در حالی‌که مدل (۲) مربوط به ذرات رسوبی با اندازه ۰/۸۲ تا ۱۸ میلی‌متر می‌باشد و از آنجا که اندازه رسوبات بستر با عمق آبشستگی رابطه عکس دارد، توان ضرایب F_d و K_I در مدل(۲) کوچک‌تر از مدل(۱) است.

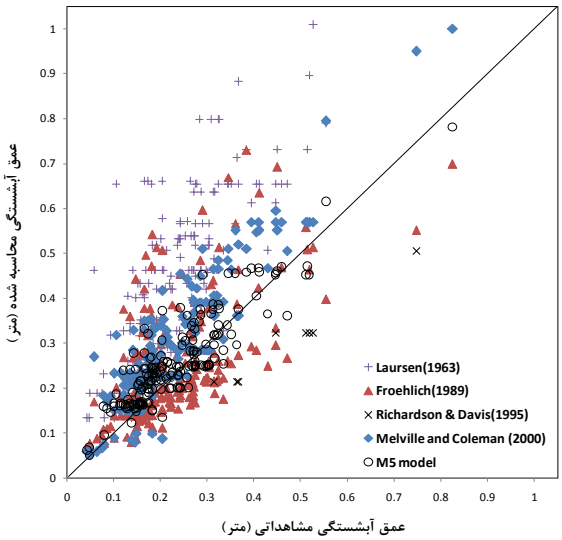
پس از ساخت مدل درختی، این مدل برای برآورد عمق آبشستگی با استفاده از داده‌های آموزشی و صحت‌سنجی مورد استفاده قرار گرفت که نتایج در شکل(۵) ارائه شده است. در این شکل نتایج مربوط به مدل M5، قبل و بعد از هرس شدن آورده شده است. همچنان‌که مشاهده می‌شود مدل درختی قبل از هرس شدن کارآیی بهتری در مرحله آموزش داشته ولی در مرحله صحت‌سنجی دقت مدل هرس شده با توجه به تراکم نقاط حول خط ۴۵ درجه بیشتر گردیده است. دلیل این امر تعداد زیاد مدل‌های مورد استفاده در الگوریتم قبل از هرس شدن می‌باشد (۵۴ مدل) که این امر احتمال بیش برازش را افزایش می‌دهد، اما در مورد مدل نهایی که تنها از سه معادله استفاده می‌کند، این احتمال کاهش یافته و مدل هرس شده نتایج بهتری را در قسمت صحت‌سنجی ارائه می‌دهد.



شکل ٤– مدل نهایی M5 برای برآورد عمق آبشستگی اطراف تکیه‌گاه



شکل ۵- همبستگی نتایج مدل درختی با مقادیر مشاهداتی در مرحله آموزش و صحت‌سنجی



شکل ٦- همبستگی نتایج معادلات سایر محققین با مقادیر مشاهداتی

برای بررسی کارآیی مدل درختی ارائه شده و مقایسه آن با روش‌های متداول از معادلات Laursen (1963)، Froehlich (1989)، Richardson و Davis (1995) و Melville و Coleman (2000) استفاده گردید. نتایج این مقایسه در شکل(۶) ارائه شده است. با توجه به شکل(۶) مدل M5 نسبت به سایر معادلات همبستگی بهتری با مقادیر آبشستگی مشاهداتی دارد. از میان معادلات سایر محققین، معادله Laursen (1963) کمترین دقت را داشته و معادله Melville و Coleman (2000) بیشترین دقت را داشته هر چند این معادله در اکثر حالات مقدار آبشستگی را بیش از مقدار واقعی برآورد کرده است. برای مقایسه بهتر روش‌های مختلف تخمین عمق آبشستگی از نمودار مستطیلی استفاده شده است. مزیت این نمودار ارائه اطلاعاتی در زمینه توزیع مقادیر با نشان دادن مقادیر میانه، دامنه تغییرات بین چارکی و بیشترین و کمترین مقادیر یک گروه از نمونه‌ها می‌باشد. این نمودار ابزاری مناسب برای شرح مکان و پراکندگی داده‌های مربوط به چند گروه مختلف است. در این تحقیق نسبت عمق آبشستگی محاسبه شده به عمق آبشستگی مشاهداتی به‌عنوان متغیر در نمودار مستطیلی در نظر گرفته شده است. برهمین اساس هرچه پراکندگی داده‌ها در اطراف مقدار یک بیشتر باشند نشان‌دهنده دقت بیشتر روش مورد نظر خواهد بود. با توجه به شکل(۷) نمودار مستطیلی مربوط به مدل درختی کوچکتر از بقیه و در اطراف یک تمرکز پیدا کرده که نشان‌دهنده دقت این الگوریتم در برآورد عمق آبشستگی می‌باشد.

روش‌های Laursen (1963) و Melville و Coleman (2000) مقادیر آبشستگی را بیش از مقدار واقعی برآورد کرده‌اند. یکی از اشکالات روش Froehlich (1989) با توجه به شکل(۷) محاسبه مقدار آبشستگی کمتر از مقدار واقعی در بسیاری از حالات می‌باشد، همچنانکه خط میانه مقادیر محاسباتی با استفاده از این روش پایین‌تر از عدد یک قرار گرفته که این امر از لحاظ مسائل اجرایی مشکلاتی را به همراه خواهد داشت. با توجه به اینکه روش Richardson و Davis (1995) تنها در شرایطی که $L/y > 25$ باشد کاربرد دارد و تنها ۸ رکورد از داده‌های مورد استفاده دارای این شرط بوده، به همین دلیل نمودار مستطیلی آن ارائه نشده است.

برای مقایسه بهتر کارآیی مدل‌های ارائه شده در این تحقیق از تحلیل‌های آماری: مجموع مربعات خطا (Sum of the Squares of the Error) خطای متوسط مطلق (Mean-Absolute Error)، جذر میانگین مربعات خطا (Root Mean Square Error) و ضریب تبیین (Determination Coefficient (R^2)) استفاده گردید. علاوه بر معیارهای ذکر شده از نسبت اختلاف (Discrepancy Ratio) که توسط White et al. (1973) پیشنهاد شده است و از معیارهای قدرتمند آماری به حساب می‌آید، استفاده گردید (Etemad-Shahidi and Taghipour, 2012).

$$SSE = \frac{\sum\left(d_s^{measured} - d_s^{computed}\right)^2}{\sum\left(d_s^{measured}\right)^2} \times 100 \qquad (۶)$$

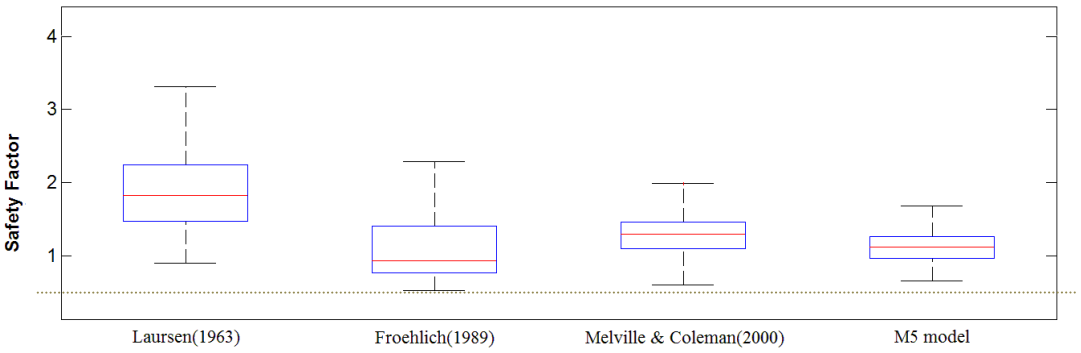
$$MAE = \frac{1}{N}\sum\left|d_s^{measured} - d_s^{computed}\right| \qquad (۷)$$

$$RMSE = \sqrt{\frac{1}{N}\sum\left(d_s^{measured} - d_s^{computed}\right)^2} \qquad (۸)$$

$$R^2 = 1 - \frac{\sum\left(d_s^{measured} - d_s^{computed}\right)^2}{\sum\left(d_s^{measured} - \overline{d_s^{measured}}\right)^2} \qquad (۹)$$

$$DR = \frac{d_s^{computed}}{d_s^{measured}} \qquad (۱۰)$$

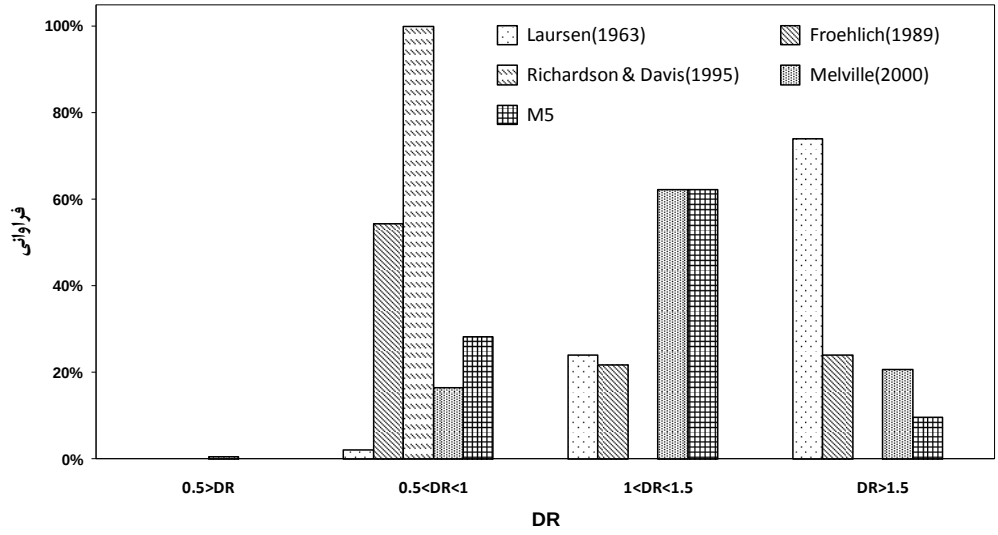
DR برابر یک نشان‌دهنده تطابق کامل نتایج مدل و مقادیر واقعی می‌باشد. چنان‌چه DR>1 باشد، آبشستگی بیشتر از مقدار واقعی برآورد شده است، در غیر اینصورت(DR<1)، آبشستگی کمتر از مقدار واقعی محاسبه شده است. در این تحقیق درصد مقادیر با DR بین ۰/۵ و ۱/۵ به‌عنوان دقت (Accuracy) هر مدل در نظر گرفته شده است که شباهت زیادی به دقت تعریف شده توسط Seo و Cheong (1998) دارد. نتایج مربوط به تحلیل آماری روش‌های مختلف تخمین عمق آبشستگی در جدول (۳) ارئه شده است. هیستوگرام فراوانی نسبت اختلاف(DR) برای تمامی روش‌ها در شکل(۸) نشان داده شده است. بر اساس تحلیل‌های صورت گرفته روش Melville و Coleman (2000) در میان روش‌های متداول بیشترین دقت را داشته هرچند معادله Richardson و Davis (1995) در بعضی از تحلیل‌ها بهتر ظاهر شده است. یکی از مشکلات اساسی روش اخیر برآورد عمق آبشستگی کمتر از مقدار واقعی است، همچنان‌که در شکل (۸) مشخص است، کلیه مقادیر محاسبه شده دارای DR<1 می‌باشند. یکی از مزایای الگوریتم درختی مورد استفاده در این تحقیق، نزدیک بودن نتایج تحلیل‌های آماری در مرحله آموزش و صحت‌سنجی است که نشان‌دهنده نحوه آموزش درست این الگوریتم می‌باشد. یکی از عوامل مؤثر در این زمینه عملیات هرس کردن است که باعث گردید مدل درختی تحت تأثیر بیش برازش قرار نگیرد. در مدل درختی مجموع مربعات خطا که در تحقیقات مربوط به آبشستگی کاربرد فراوانی دارد، در مقایسه با معادله Melville و Coleman (2000) از ۱۳/۲۲ به ۴/۳۷ کاهش یافته، علاوه بر این جذر میانگین مربعات خطا نیز از ۰/۰۹۶ به ۰/۰۵۵ کاهش یافته که این امر نشان‌دهنده کارآیی این الگوریتم در تخمین عمق آبشستگی است.



شکل ۷- نمودار مستطیلی روش‌های مختلف تخمین عمق آبشستگی اطراف تکیه‌گاه پل

جدول ۳- مقایسه کارآیی مدل‌های تخمین عمق آبشستگی اطراف تکیه‌گاه

مدل		SSE	Accuracy	MEA	RMSE	R^2
لارسن (۱۹۶۳)		۸۳/۹	٪۲۶	۰/۲۱	۰/۲۴۱	۰/۶۸
فرولیچ (۱۹۸۹)		۱۹/۰۴	٪۷۶	۰/۰۸	۰/۱۱۵	۰/۳۸
ریچاردسون و دویس (۱۹۹۵)		۱۲/۶۲	٪۱۰۰	۰/۱۷	۰/۱۷۵	۰/۹۶
ملویل و کلمن (۲۰۰۰)		۱۳/۲۲	٪۷۸	۰/۰۸	۰/۰۹۶	۰/۸۱
مدل M5	مرحله آموزش	۴/۳۱	٪۹۱	۰/۰۴۳	۰/۰۵۴	۰/۸۱
	صحت سنجی	۴/۶۱	٪۸۸	۰/۰۴۶	۰/۰۵۶	۰/۷۱
	کلیه داده‌ها	۴/۳۷	۰/۹۰	۰/۰۴۳	۰/۰۵۵	۰/۸۱



شکل ۸- مقایسه نسبت اختلاف(DR) مدل‌های مختلف

تحلیل حساسیت

برای بررسی تأثیر هر کدام از پارامترهای ورودی بر روی عمق آبشستگی در مدل درختی، پارامتر مورد نظر حذف شده و ساختار درختی بدون آن پارامتر تشکیل شد. پس از آن کارآیی مدل ساخته شده بر اساس تحلیل‌های آماری تعیین گردید. در جدول(۴) نتایج تحلیل‌های آماری مربوط به تأثیر هر کدام از پارامترهای ورودی روی خروجی مدل درختی ارائه شده است. این تغییرات به‌ویژه در مورد مجموع مربعات خطا و میانگین مربعات خطا مشخص‌تر است. بر اساس نتایج ارائه شده در این جدول ضریب K_{yL} که نشان‌دهنده تأثیر هم‌زمان طول تکیه‌گاه و عمق جریان است، بیشترین تأثیر را روی عمق آبشستگی دارد. حذف این پارامتر باعث افزایش بیش از چهار برابری مجموع مربعات خطا گردیده است. مقدار میانگین مربعات خطا نیز از ۰/۰۵۵ به ۰/۱۱۶ افزایش یافته که نشان‌دهنده کاهش بیش از حد کارآیی مدل درختی با حذف این پارامتر می‌باشد. پس از این پارامتر عدد فرود ذره و ضریب شکل تکیه‌گاه بیشترین تأثیر را روی میزان آبشستگی دارند.

جدول ٤- تحلیل حساسیت پارامترهای مورد استفاده در مدل درختی

مدل	پارامترهای آماری				
	SSE	Accuracy	MEA	RMSE	R^2
مدل M5	۴/۳۷	۰/۹۰	۰/۰۴۳	۰/۰۵۵	۰/۸۱
مدلM5 بدون F_d	۸/۷۸	۰/۸۶	۰/۰۵۸	۰/۰۷۸	۰/۵۷
مدلM5 بدون K_I	۶/۵۳	۰/۸۶	۰/۰۵۲	۰/۰۶۷	۰/۶۸
مدلM5 بدون K_s	۷/۶۲	۰/۸۶	۰/۰۵۵	۰/۰۷۳	۰/۶۳
مدلM5 بدون K_{yL}	۱۹/۴۱	۰/۷۲	۰/۰۸۶	۰/۱۱۶	۰/۰۱

نتیجه‌گیری

هدف از این تحقیق ارائه روشی بر مبنای الگوریتم‌های درختی برای محاسبه عمق آبشستگی اطراف تکیه‌گاه پل می‌باشد. از نکات مهم این تحقیق می‌توان به ترکیب روش‌های کلاسیک و نوین برای محاسبه عمق آبشستگی اشاره کرد. بر این اساس از ضرایب بدون بعد مورد استفاده در تحقیقات گذشته به‌عنوان پارامترهای ورودی الگوریتم درختی استفاده گردید. در مقایسه با روش‌های نوین داده‌کاوی از قبیل شبکه‌های عصبی و فازی، مدل M5 این مزیت را داشته که خروجی آن به‌صورت روابط رگرسیونی بوده و می‌توان در سایر مدل‌ها از آن استفاده نمود. مدل M5 مورد استفاده در این تحقیق فضای مسأله را به سه قسمت مجزا تقسیم کرده و برای هرکدام از زیردامنه‌ها یک رابطه رگرسیونی ارائه داد. ضرایب عمق جریان-طول تکیه گاه و شدت جریان در ایجاد ساختار درختی بیشترین نقش را داشته و در این بین ضریب شدت جریان در کلیه معادلات ارائه شده توسط مدل درختی بیشترین توان را داشته که نشان‌دهنده تأثیر سرعت جریان روی میزان حداکثر عمق آبشستگی در اطراف تکیه‌گاه پل می‌باشد. برای بررسی کارآیی الگوریتم درختی از تحلیل‌های آماری استفاده گردید که نتایج تحلیل‌های آماری صورت گرفته نشان‌دهنده دقت بالای الگوریتم درختی در مقایسه با روش‌های متداول می‌باشد. در بین معادلات محققین مختلف نیز معادله ارائه شده توسط Melville و Coleman (2000) دقت بالاتری داشته و همان‌گونه که اشاره شد بعضی از ضرایب این معادله برای ساخت مدل درختی مورد استفاده قرار گرفت. بر اساس تحلیل حساسیت صورت گرفته ضریب عمق جریان-طول تکیه‌گاه به‌عنوان مؤثرترین پارامتر بر روی پدیده آبشستگی تعیین گردید که با مطالعات صورت گرفته قبلی مطابقت دارد. بر اساس تحلیل صورت گرفته حذف این ضریب باعث افزایش مقادیر SSE و RMSE از ۴/۳۷ و ۰/۰۵۵ به ۱۹/۴۱ و ۰/۱۱۶ گشته است. علاوه بر این دقت مدل نیز با حذف این ضریب به میزان ۲۰ درصد کاهش یافته است.

منابع

1- Ballio, F., Teruzzi, A. and Radice, A., 2009. Constriction effects in clear-water scour at abutments. *Journal of Hydraulic Engineering*, *135*(2), pp.140-145.

2- Bayram, A. and Larson, M., 2000. Analysis of scour around a group of vertical piles in the field. *Journal of waterway, port, coastal, and ocean engineering*, *126*(4), pp.215-220.

3- Bhattacharya, B. and Solomatine, D.P., 2006. Machine learning in sedimentation modelling. *Neural Networks*, *19*(2), pp.208-214.

4- Coleman, S.E., Lauchlan, C.S. and Melville, B.W., 2003. Clear-water scour development at bridge abutments. *Journal of Hydraulic Research*, *41*(5), pp.521-531.

5- Dey, S., Chiew, Y.M. and Kadam, M.S., 2008. Local scour and riprap stability at an abutment in a degrading bed. *Journal of Hydraulic Engineering*, *134*(10), pp.1496-1502.

6- Dongol, D.M.S. and Melville, B.W., 1994. *Local scour at bridge abutments*. Department of Civil Engineering, University of Auckland.

7- Etemad-Shahidi, A. and Ghaemi, N., 2011. Model tree approach for prediction of pile groups scour due to waves. *Ocean Engineering*, *38*(13), pp.1522-1527.

8- Etemad-Shahidi, A. and Mahjoobi, J., 2009. Comparison between M5′ model tree and neural networks for prediction of significant wave height in Lake Superior. *Ocean Engineering*, *36*(15-16), pp.1175-1181.

9- Etemad-Shahidi, A. and Taghipour, M., 2012. Predicting longitudinal dispersion coefficient in natural streams using M5′ model tree. *Journal of hydraulic engineering*, *138*(6), pp.542-554.

10- Froehlich, D.C., 1989. Local scour at bridge abutments. In *Proceedings of the 1989 National Conference on Hydraulic Engineering* (pp. 13-18).

11- Gill, M.A., 1970. *Bed erosion around obstructions in rivers* (Doctoral dissertation, University of London).

12- Gill, M.A., 1972. Erosion of sand beds around spur dikes. *Journal of the Hydraulics Division*, *98*(hy9).

13- Hager, W.H. and Oliveto, G., 2002. Shields' entrainment criterion in bridge hydraulics. *Journal of Hydraulic Engineering*, *128*(5), pp.538-542.

14- Kandasamy, J.K. and Melville, B.W., 1998. Maximum local scour depth at bridge piers and abutments. *Journal of hydraulic research*, *36*(2), pp.183-198.

15- Kuhnle, R.A., Alonso, C.V. and Shields, F.D., 1999. Geometry of scour holes associated with 90 spur dikes. *Journal of Hydraulic Engineering*, *125*(9), pp.972-978.

16- Kwan, T.F., Melville, B.W. and Raudkivi, A.J., 1984. *Study of Abutment Scour: A Report Submitted to the Road Research Unit of the National Roads Board*. Department of Civil Engineering, University of Auckland.

17- Laursen, E.M., 1962. Scour at bridge crossings. *Transactions of the American Society of Civil Engineers*, *127*(1), pp.166-179.

18- Laursen, E.M., 1963. An analysis of relief bridge scour. *Journal of the Hydraulics Division*, *89*(3), pp.93-118.

19- Melville, B.W. and Coleman, S.E., 2000. *Bridge scour*. Water Resources Publication.

20- Melville, B.W., 1992. Local scour at bridge abutments. *Journal of Hydraulic Engineering*, *118*(4), pp.615-631.

21- Melville, B.W., 1997. Pier and abutment scour: integrated approach. *Journal of hydraulic Engineering*, *123*(2), pp.125-136.

22- Najafzadeh, M., Barani, G.A. and Hessami-Kermani, M.R., 2015. Evaluation of GMDH networks for prediction of local scour depth at bridge abutments in coarse sediments with thinly armored beds. *Ocean Engineering*, *104*, pp.387-396.

23- Quinlan, J.R., 1992, November. Learning with continuous classes. In *5th Australian joint conference on artificial intelligence* (Vol. 92, pp. 343-348).

24- Richardson, E.V. and Davis, S.R., 1995. *Evaluating scour at bridges: Federal Highway Administration Hydraulic Engineering Circular No. 18*. Publication FHWA-IP-90-017.

25- Seo, I.W. and Cheong, T.S., 1998. Predicting longitudinal dispersion coefficient in natural streams. *Journal of hydraulic engineering*, *124*(1), pp.25-32.

26- Sumer, B.M., Fredsøe, J. and Christiansen, N., 1992. Scour around vertical pile in waves. *Journal of waterway, port, coastal, and ocean engineering*, *118*(1), pp.15-31.

27- Tey, C.B., 1984. *Local scour at bridge abutments*. Department of Civil Engineering, University of Auckland.

28- Wang, Y. and Witten, I.H., 1996. Induction of model trees for predicting continuous classes. Proceedings of the Poster Papers of the *European Conference on Machine Learning, University of Economics, Faculty of Informatics and Statistics, Prague.*

29- White, W.R., Milli, H. and Crabbe, A.D., 1973. Sediment transport: an appraisal methods, Vol. 2: Performance of theoretical methods when applied to flume and field data. *Hydraulic Research Station Report No. IT119, Wallingford, UK*.

30- Yorozuya, A. and Ettema, R., 2015. Three abutment scour conditions at bridge waterways. *Journal of Hydraulic Engineering*, *141*(12), p.04015028.

Shahid Chamran University of Ahvaz

Irrigation Sciences and Engineering **Vol. 41, No. 1, Spring 2018, p. 1-16**

EXTENDED ABSTRACT

Predicting Maximum Scour Depth Around Bridge Abutment Using M5 Model

J. Zahiri[1*] and S. M. Kashefipour[2]

1*- Corresponding author, Assistant Professor, Department of Water Engineering, Khuzestan Ramin Agriculture and Natural Resources University, Iran *(j.zahiri@ramin.ac.ir).*

2- Professor, Department of Water Science, Shahid Chamran University of Ahvaz, Iran.

Received: 6 May 2015 Accepted:11 September 2016

Keywords: Bridge abutment, M5 model, Pruning, Flow depth-abutment length, Sensitivity analysis.

Introduction

Scour around piers and abutments leads to a lot of bridge failures. Scouring around the bridge piers and abutments is one of the most important factors in bridge failure which happens over time and increases during floods, probably causing general damage to the bridge. Accordingly, the accurate estimation of scouring around the bridge piers and abutments can help engineers to take steps to deal with bridge destructions. Recently, data-driven models have been used widely to estimate the scour depth around bridge piers and abutments. Najafzadeh, Barani and Hessami-Kermani (2015) used a set of data-driven models such as the gravitational search algorithm and the particle swarm optimization algorithm to estimate the scour around bridge abutments. Sediment particles size, geometric characteristics of bridge abutment and approach flow conditions were considered as effective parameters on scouring mechanism. The results of this research showed that data-driven models can estimate the scour depth with high accuracy. In the present study, M5 tree model has been used to predict maximum scour depth around bridge abutment. High accuracy, simplicity and understandability are considered the advantages of this algorithm. A huge amount of data from credible references have been used to build and validate the tree model.

Methodology

In this study, 188 records of scour around bridge abutments have been used to build and verify the proposed model. Densimetric particle Froude number (F_{rd}), foundation shape factor (K_s), flow depth- abutment length expression (K_{yL}) and flow intensity factor (K_I) have been used as input variables to estimate maximum scour depth around bridge abutment. These coefficients are introduced in Table 1.

Table 1- Foundation shape factor (K_s), flow depth- abutment length expression (K_{yL}) and flow intensity factor (K_I) for different abutment shapes

Parameter	*Effective factor*	*Parameter value*			
K_s	*Abutment shape*	*Vertical wall*	*Wing wall*	*Spill-through abutment*	
				1:0.5	1:1
		1.0	0.75	0.6	0.5
K_{yL}	L/y	$L/y < 1$		$K_{yL} = 2L$	
		$1 < L/y < 25$		$K_{yL} = 2\sqrt{yL}$	
		$L/y > 25$		$K_{yL} = 10y$	
K_I	V/V_c	$V/V_c < 1$		$K_I = V/V_c$	
		$V/V_c \geq 1$		$K_I = 1$	

The M5 algorithm was initially introduced by Quinlan (1992) and then developed by Wang and Witten (1997). Two main processes are considered in the algorithm: building the tree and deriving the knowledge from it. The first process involves dividing the input parameter space into a smaller subspace for which a multiple regression model is assigned. The scheme resembles an inverted tree in which the root is on the top while the leaves are at the bottom. In the second process, a data record is introduced into the root of the tree (Etemad-Shahidi and Taghipour, 2012). M5 constructs the regression tree by using standard deviation reduction (SDR) factor to split the space:

$$\Delta error = SDR = sd(T) - \sum \frac{|T_i|}{|T|} \times sd(T_i) \quad (1)$$

where T is the set of data points before splitting, T_i is the data point that results from splitting the space and falls into one subspace according to the chosen splitting parameter, and sd is the standard deviation (Wang and Witten, 1997).

Results and Discusspn

The tree structure of M5 algorithm after pruning is shown in Figure 1. The linear equations represented by this tree model have been converted from logarithmic state to power equations in (2) to (4) after conversion.

$$Model\ 1: d_s = 0.08 F_d^{0.78} K_{yL}^{0.11} K_I^{1.23} K_s^{0.27} \quad (2)$$

$$Model\ 2: d_s = 0.21 F_d^{0.04} K_{yL}^{0.16} K_I^{0.76} K_s^{0.27} \quad (3)$$

$$Model\ 3: d_s = 0.67 F_d^{0.12} K_{yL}^{0.9} K_I^{1.3} K_s^{1.03} \quad (4)$$

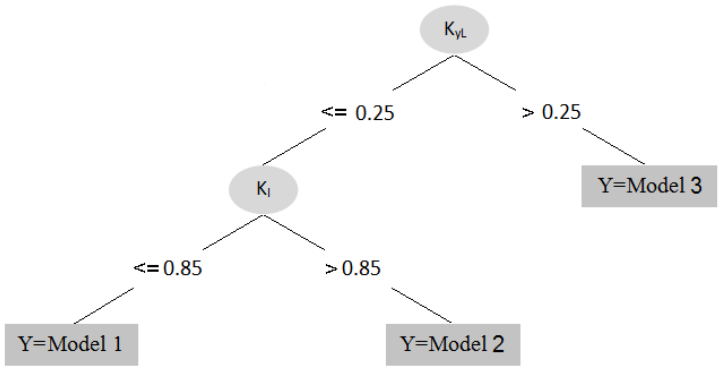


Figure 1- M5 tree structure to predict maximun scour depth around bridge abutment

It is assumed that the flow depth-abutment length and the flow intensity have the greatest impact on the tree structure. M5 model splits the input space into three subspaces and proposes a regression equation for each sub-domain. In this research, the ratio of calculated and measured scour depth was used as a variable in the box plot (Figure 2). The more scattered the data are around the value of one, the greater the accuracy will be. According to Figure 2, the box of the M5 tree model is smaller than the other equations and balanced around one, which indicates the accuracy of the tree model in scour depth estimation.

Different statistical analyses on the proposed model and the previous equations indicated that the tree model could predict scour depth with more precision. In the tree model, the sum of squared errors that are widely used in scour prediction research has decreased from 22.23 to 4.37 compared with Melville and Coleman's (2000) model. In addition, the root mean square error has changed from 0.096 to 0.555, which indicates the efficiency of this algorithm in scour depth estimation. Sensitivity analysis has been used to evaluate the effect of each parameter on scour depth. The results showed that the flow depth-abutment length has the greatest effect on scour depth and ignoring this parameter increases the sum of the squares of the error from 4.37 to 19.41.

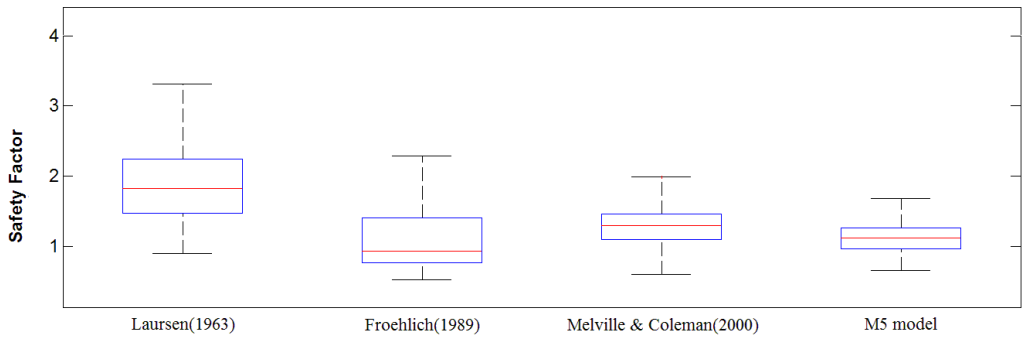


Figure 2- Box plot of different methods for abutment scour prediction

Conclusions

The M5 tree model used in this research divides the problem space into three distinct parts and provides a regression equation for each subdomain. The flow depth-abutment length expression and the flow intensity factor in the tree structure have the greatest roles. In addition, the flow intensity factor is the most powerful in all equations developed by the tree model, which indicates the effect of the flow velocity on the maximum scour depth around bridge abutment. The results of statistical analyses indicate that the tree algorithm is highly accurate compared to the

conventional methods. Among the equations introduced by various researchers, the formula provided by Melville and Coleman (2000) was more accurate and some of the coefficients of this equation were used to construct the tree model in present study. Based on sensitivity analysis, the coefficient of flow depth- abutment length was determined as the most effective parameter on the scouring phenomenon, which is consistent with the previous studies. According to the sensitivity analysis, removing this coefficient has reduced the accuracy of the model by 20%.

References

1- Etemad-Shahidi, A. and Taghipour, M., 2012. Predicting longitudinal dispersion coefficient in natural streams using M5′ model tree. *Journal of Hydraulic Engineering*, *138*(6), pp.542-554.

2- Melville, B.W. and Coleman, S.E., 2000. *Bridge scour*. Water Resources Publication.

3- Najafzadeh, M., Barani, G.A. and Hessami-Kermani, M.R., 2015. Evaluation of GMDH networks for prediction of local scour depth at bridge abutments in coarse sediments with thinly armored beds. *Ocean Engineering*, *104*, pp.387-396.

4- Quinlan, J.R., 1992, November. Learning with continuous classes. In the *5th Australian Joint Conference on Artificial Intelligence* (Vol. 92, pp. 343-348).

5- Wang, Y. and Witten, I.H., 1996. Induction of model trees for predicting continuous classes. Proceedings of the Poster Papers of the *European Conference on Machine Learning, University of Economics, Faculty of Informatics and Statistics, Prague.*