# پیش بینی برش آب در چاه های یک میدان نفتی با رویکرد یادگیری عمیق شبکه عصبی حافظه کوتاه مدت بلند و داده های تاریخچه تولید

**عارف الطاف[۱]، سیدجواد ایرانبان‌فرد[۲] *، عباس خمسه[۳]، رضا رادفر[۱]**

۱- گروه مدیریت فناوری اطلاعات، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

۲- گروه مدیریت صنعتی، واحد شیراز، دانشگاه آزاد اسلامی، شیراز، ایران

۳- گروه مدیریت صنعتی، واحد کرج، دانشگاه آزاد اسلامی، کرج، ایران

airanban@yahoo.com

Aref.altaf@gmail.com



## چکیده

در مخازن نفتی با آبده قوی که در نیمه دوم عمر تولیدی خود هستند، افزایش برش آب[۱] معمولاً نه به‌صورت یک پدیده ناگهانی، بلکه به‌عنوان نتیجه تجمعی پیشروی آب، ناهمگنی لایه‌ها و مداخلات بهره‌برداری ظاهر شده و در عمل به مهم‌ترین عامل محدودکننده تولید و بسته‌شدن چاه‌ها تبدیل می‌شود؛ به همین دلیل، پیش‌بینی واقع‌بینانه و مبتنی بر رفتار مخزن، یکی از کلیدی‌ترین ابزارهای مهندسی برای مدیریت تولید و تصمیم‌گیری عملیاتی در این‌گونه مخازن به شمار می‌آید. در این پژوهش، رویکرد یادگیری عمیق با استفاده از شبکه عصبی حافظه کوتاه‌مدت بلند[۲] برای پیش‌بینی برش آب در یک میدان نفتی بزرگ با ۱۸۰ حلقه چاه، پشتیبانی آبده قوی و خصوصیات سنگ شناسی[۳] لایه ای پیچیده ارائه می شود. مجموعه داده با کیفیت شامل ۳۳,۶۹۷ نمونه برچسب دار پس از اجرای فرآیند پیش پردازش، به مجموعه ورودی شامل ۱۷ پارامتر پویا و ایستا مؤثر بر رفتار مخزن از جمله موقعیت چاه (x, y)، تعداد لایه های مشبک شده، طول و عمق مشبک، تراوایی مؤثر، نرخ تولید نفت و زمان تولید منتج شد. با توجه به ماهیت سری‌زمانی داده‌های تولید و به منظور اطمینان کافی از اعتبار نتایج پیش بینی مدل، ۸۰ درصد اول داده های ثبت شده برای آموزش و ۲۰ درصد باقیمانده (بدون هم‌پوشانی) برای آزمون انتخاب شد. معماری LSTM ابتدا با جستجوی شبکه‌ای بهینه سازی شد، در چهار ناحیه متمایز مخزن آموزش دید و به صورت بخش به بخش مورد ارزیابی قرار گرفت. نتایج ارزیابی نشان دهنده عملکرد استثنایی مدل بود به طوری که ضریب همبستگی پیرسون(R) بین ۰٫۹۵۷ تا ۰٫۹۷۶ (در فاز آموزش) و ۰٫۸۸۹ تا ۰٫۹۲۰ (در فاز آزمون) بود. همچنین، توزیع خطای نسبی مطلق در همه بخش‌ها نزدیک به صفر بود. نکته مهم آن که، این مدل در فاز آزمون موفق شد به دقت رویدادهای عملیاتی کلیدی مانند مسدود کردن لایه‌های آب‌ده، تکمیل مجدد و تغییرات برش آب ناشی از کاهش نرخ تولید، را پیش بینی کند. اعتبارسنجی‌های بصری و آماری تأییدکنندهٔ قابلیت این مدل در نشان دادن روندهای بلندمدت و رفتارهای گذرا با صحت بالا است. یافته‌های این پژوهش نشان می‌دهد که LSTM ابزاری مؤثر، مقیاس‌پذیر و قابل‌اعتماد برای مدیریت

[1] Water cut
[2] Long Short-Term Memory (LSTM)
[3] Lithology

داده محور آب، پیش‌بینی تولید و تصمیم گیری راهبردی در مخازن نفتی بزرگ و ناهمگن با مکانیزم رانش آب محسوب می شود.

**کلمات کلیدی: پیش‌بینی برش آب، یادگیری عمیق، مدل‌سازی سری‌زمانی، مدل LSTM، تحلیل رفتار مخازن، پیش‌بینی داده‌محور، مدل‌های هوشمند مهندسی مخزن**

## مقدمه

تولید آب در میادین نفتی بالغ، به‌ویژه در مخازنی با مکانیزم رانش آب قوی، بخشی اجتناب‌ناپذیر از رفتار طبیعی مخزن در طول زمان به شمار می‌آید که با تداوم تولید تشدید می‌شود. در چنین شرایطی، پیشروی تدریجی آب سازندی به نواحی نفت‌دار سبب افزایش سهم آب در جریان تولیدی شده و در نتیجه، برش آب روندی صعودی پیدا می‌کند. افزایش برش آب، افزون بر محدودکردن نرخ تولید نفت، هزینه‌های عملیاتی مرتبط با جمع‌آوری، تفکیک و دفع آب تولیدی را به‌طور قابل‌توجهی افزایش داده و در موارد حاد، عملاً به عامل تعیین‌کننده در توقف تولید و رهاسازی چاه‌ها تبدیل می‌شود. بر این اساس، دستیابی به برآوردی دقیق و قابل اتکا از رفتار برش آب، نقشی کلیدی در بهینه‌سازی راهبردهای تولید، مدیریت سیستم‌های فرازآوری مصنوعی و ارزیابی عملکرد آتی میدان ایفا می‌کند.

روش‌های سنتی پیش‌بینی برش آب، چه در قالب مدل‌های تحلیلی مانند باکلی-لورِت[1] و چه در قالب شبیه‌سازی عددی مخزن (Ahmed, 2019) ، با وجود اتکا به مبانی فیزیکی معتبر، به داده‌های زمین‌شناسی و پتروفیزیکی گسترده و محاسبات پیچیده نیاز دارند و همین موضوع موجب می‌شود در بازنمایی رفتار مخازن ناهمگن با تاریخچهٔ تکمیل پویا، دقت و کارایی محدودی داشته باشند. از سوی دیگر، روش‌های تجربی نظیر چان-فریک[2] و نسبت آب به نفت، با وجود اینکه ساده و کم‌هزینه هستند، مبنای مکانیکی کافی و تعمیم‌پذیری خوبی در مورد رفتارهای مختلف چاه ندارند.

در سال‌های اخیر، ظهور روش‌های داده محور، به ویژه یادگیری ماشین و یادگیری عمیق، افق جدیدی برای پیش‌بینی تولید گشوده است. این رویکردها روابط غیرخطی میان ورودی‌ها و خروجی‌ها را مستقیما از داده ها و بدون نیاز به توصیف صریح معادلات فیزیکی استخراج می‌کنند. شبکه‌های عصبی مصنوعی، ماشین بردار پشتیبان و جنگل تصادفی با موفقیت نسبی برای پیش‌بینی برش آب مورد استفاده قرار گرفته‌اند(Ahmadi et al., 2019; Chen et al., 2022; Das et al., 2025)، اما این مدل‌ها اغلب داده‌های تولید را به عنوان مشاهدات ایستا یا مستقل در نظر گرفته، از وابستگی‌های زمانی حاکم بر پویایی مخزن صرفه نظر می‌کنند. شبکه‌های بازگشتی و به‌ویژه LSTM با قابلیت حفظ وابستگی‌های زمانی بلندمدت و رفع مشکل محوشوندگی گرادیان(Hochreiter & Schmidhuber, 1997; Sheikhoushaghi et al., 2022)، در سال‌های اخیر در پیش‌بینی نرخ تولید(Tripathi et al., 2024)، فشار ته‌چاهی و عملکرد چاه(Jin & Emami-Meybodi, 2025) و تشخیص نابهنجاری و خرابی تجهیزات(Al-Mamoori et al., 2025) به نحو موفقی مورد استفاده قرار گرفته است. پژوهش‌های اخیر مانند( Li et al., 2025; WANG et al., 2020; Zhang et al., 2024) نیز کاربرد LSTM و مدل‌های هیبریدی را در پیش‌بینی تولید نشان داده‌اند، هرچند پیچیدگی محاسباتی برخی رویکردها و نیاز به پردازش گسترده سیگنال، استفاده آنها را در میادین عظیم

[1] Buckley–Leverett displacement theory
[2] Chan–Frick

با صدها چاه محدود می‌کند. در مطالعات داخلی نیز نشان داده شده است که ادغام داده‌های زمین‌شناسی، پتروفیزیکی، لرزه‌ای و چاهی با استفاده از روش‌های هوشمند (مانند شبکه عصبی) می‌تواند در بازنمایی ناهمگنی‌های مخزنی (مانند تراکم شکستگی) و تحلیل رفتار میدان نقش مؤثری داشته باشد (Bayat et al., 2015).

با وجود پیشرفت‌های اخیر، چالش‌هایی مانند امکان سنجی بکارگیری روش‌های هوشمند در پیش بینی عملکرد مخازن بزرگ و بسیار پیچیده با تعداد چاه‌های زیاد، توجه کمتر به کیفیت داده‌ها و مراحل دقیق پیش‌پردازش، تقسیم‌بندی نادرست داده‌های سری زمانی و نبود اعتبارسنجی در برابر رخدادهای عملیاتی واقعی همچنان در بسیاری از مطالعات مشاهده می‌شود. از سوی دیگر، با توجه به محدودیت های مدل های ارائه شده پیشین در تخمین برش آب برای میادین عظیم و با پیچیدگی زمین شناسی فراوان، نوآوری و تمایز اصلی این پژوهش در ارائه و اعتبارسنجی یک چارچوب عملیاتی مبتنی بر یادگیری عمیق برای پیش‌بینی برش آب در یکی از بزرگ‌ترین مخازن نفتی دنیا است که ضمن رفع محدودیت های ذکر شده، سه پیشرفت کلیدی را محقق می‌سازد: (۱) توانایی مدلسازی همزمان روندهای بلندمدت و رویدادهای عملیاتی گذرا (مانند تغییر لایه تولیدی یا نرخ تولید) تنها از داده‌های تاریخچه تولید؛ (۲) مقیاس‌پذیری و تطبیق‌پذیری با آموزش ناحیه‌ای مدل بر روی یک میدان واقعی بزرگ با ۱۸۰ چاه و ساختار زمین‌شناسی پیچیده؛ و (۳) قابلیت تعمیم قوی که از طریق یک روش تقسیم‌بندی زمانی واقع‌بینانه و بهینه‌سازی سیستماتیک معماری LSTM حاصل شده است. برخلاف روش‌های متعارف که اغلب به داده‌های ورودی پرهزینه وابسته‌اند یا در مدلسازی وابستگی‌های غیرخطی و زمانی پیچیده محدودیت دارند، مدل پیشنهادی ما داده‌های معمول تولید را به یک ابزار شبه-شبیه‌ساز قدرتمند تبدیل می‌کند که می‌تواند به عنوان یک سیستم پشتیبانی تصمیم‌گیری راهبردی برای مدیریت تولید و افزایش برداشت نفت در مخازن بالغ به کار رود. در این چارچوب، پیش‌پردازش کامل شامل حذف داده‌های پرت، جایگزینی مقادیر مفقوده، هموارسازی، کنترل سازگاری فنی و تحلیل حساسیت مونت‌کارلو انجام شد و بیش از ۳۳ هزار رکورد از داده های تاریخچه تولید با ۱۷ ورودی دارای تفسیر فیزیکی روشن به منظور طراحی و ارزیابی مدل هوشمند مورد استفاده قرار گرفت. مدل نهایی علاوه بر ارزیابی آماری، در برابر رخدادهای عملیاتی واقعی مانند تکمیل مجدد و تغییر نرخ تولید نیز آزموده شد و بدین ترتیب قابلیت کاربرد عملی آن تأیید گردید.

به‌منظور تبیین جایگاه پژوهش حاضر در میان مطالعات پیشین، جدول ۱ خلاصه‌ای از پژوهش‌های شاخص در زمینه پیش‌بینی برش آب و عملکرد تولید را از منظر برخی معیارهای کلیدی در پژوهش حاضر ارائه می‌کند.

**جدول ۱: مقایسه مطالعات منتخب در زمینه پیش‌بینی برش آب و عملکرد تولید و جایگاه پژوهش حاضر**

| مرجع | هدف پژوهش | روش مورد استفاده | داده‌های مورد استفاده | دستاورد اصلی | محدودیت نسبت به مسئله این پژوهش | جنبه توسعه‌یافته در پژوهش حاضر |
|---|---|---|---|---|---|---|
| Ahmadiو همکاران (2019) | پیش‌بینی برش آب و شوری سیال تولیدی | Data-driven + AR/VAR | داده‌های واقعی بهره‌برداری از دو گروه چاه نفت | ارائه چارچوب داده‌محور برای پیش‌بینی برش آب و شوری بدون نیاز | عدم استفاده همزمان از ویژگی‌های زمین‌شناسی، | ادغام همزمان اطلاعات مخزن، تکمیل چاه و تاریخچه تولید |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | به شبیه‌سازی عددی. | مشخصات مخزن و تکمیل چاه | |
| **Chenو همکاران (2022)** | پیش‌بینی برش آب چاه‌های نفت | Random Forest Regression | داده‌های واقعی تولید چاه‌های نفت | مدل‌سازی مؤثر روابط غیرخطی و بهبود دقت نسبت به روش‌های رگرسیونی کلاسیک. | عدم مدل‌سازی وابستگی‌های زمانی بلندمدت و وابستگی به مهندسی ویژگی‌ها | مدل‌سازی وابستگی‌های زمانی بلندمدت با استفاده از LSTM |
| Wangو همکاران (2020) | پیش‌بینی تولید در مرحله برش آب بسیار بالا | LSTM | داده‌های واقعی تولید یک میدان ماسه‌سنگی با تراوایی متوسط تا زیاد تحت توسعه با آب‌تزریقی | برتری LSTM در پیش‌بینی سری‌های زمانی تولید نسبت به FCNN و روش‌های متداول. | تمرکز بر پیش‌بینی تولید و عدم مدل‌سازی مستقیم برش آب | پیش‌بینی مستقیم برش آب به‌جای نرخ تولید |
| **Liو همکاران (2025)** | پیش‌بینی تولید چاه‌های با برش آب بالا | CEEMDAN–SR–BiLSTM | داده‌های واقعی تولید چاه‌های نفت با برش آب بالا | افزایش دقت و پایداری پیش‌بینی با استفاده از مدل هیبریدی CEEMDAN–SR–BiLSTM. | پیچیدگی محاسباتی بالا و تمرکز بر پیش‌بینی نرخ تولید | کاهش پیچیدگی مدل همراه با حفظ قابلیت کاربرد در مدیریت مخزن |

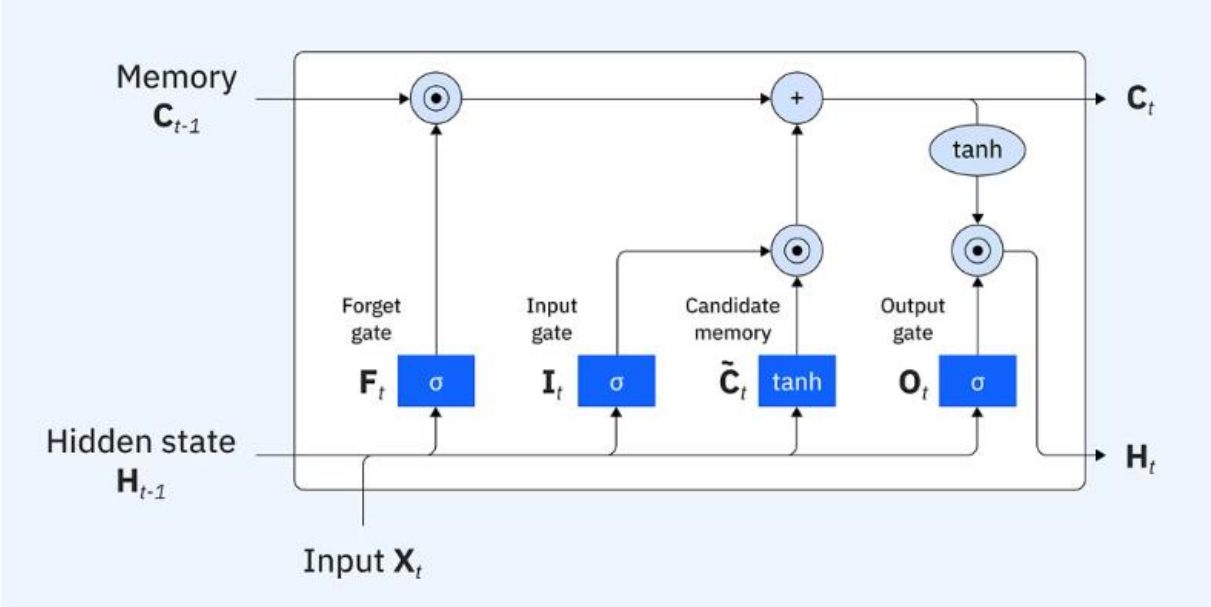
همان‌گونه که جدول ۱ نشان می‌دهد، مطالعات پیشین هر یک بخشی از مسئله پیش‌بینی برش آب یا عملکرد تولید را پوشش داده‌اند؛ با این حال، ادغام همزمان اطلاعات زمین‌شناسی، ویژگی‌های تکمیل چاه، تاریخچه تولید و مدل‌سازی وابستگی‌های زمانی بلندمدت در یک چارچوب واحد کمتر مورد توجه قرار گرفته است. پژوهش حاضر با تمرکز بر این خلأ، چارچوبی مبتنی بر LSTM برای پیش‌بینی مستقیم برش آب ارائه می‌دهد.

## تئوری شبکه حافظه کوتاه مدت بلند

شبکه‌های عصبی بازگشتی[1]، خانواده‌ای از مدل‌های عصبی هستند که با هدف مدل‌سازی داده‌های سری زمانی توسعه یافته‌اند و از طریق حفظ یک حالت پنهان در نقش «حافظه» در طول توالی، امکان درنظرگرفتن اثر اطلاعات پیشین را در پردازش ورودی‌های جاری فراهم می‌کنند. در مدل استاندارد RNN، حالت پنهان $h_t$ در هر گام زمانی بر اساس ورودی جاری $x_t$ و حالت پنهان قبلی $h_{t-1}$ محاسبه می‌شود و معمولاً از تابع فعال‌سازی tanh بهره می‌گیرد؛ با این حال، این ساختار هنگام آموزش با روش پس‌انتشار در طول زمان[2] با مشکل محوشدن[3] یا افزایش ناگهانی گرادیان‌ها[4] مواجه می‌شود (Bengio et al., 1994; Hochreiter, 2001)، این پدیده مانع از یادگیری وابستگی‌های زمانی بلندمدت می‌شود، زیرا گرادیان‌هایی که

[1] Recurrent Neural Network (RNN)
[2] Backpropagation Through Time (BPTT)
[3] Vanishing Gradient Problem
[4] Exploding Gradient Problem

برای به‌روزرسانی وزن‌ها در فرایند آموزش به‌کار می‌روند، یا به‌صورت نمایی تضعیف می‌شوند و یا به‌طور کنترل‌نشده رشد می‌کنند؛ در نتیجه، اثر اطلاعات مربوط به گام‌های زمانی دوردست بر خروجی لحظه‌ای شبکه عملاً از بین می‌رود. برای غلبه بر این چالش، شبکه حافظه کوتاه‌مدت بلند توسط هوخریتر و اشمیدهوبر معرفی شد (Hochreiter & Schmidhuber, 1997). ویژگی متمایز معماری LSTM، وجود یک سلول حافظه با حلقه بازخورد داخلی و سازوکار دروازه‌ای است که به‌صورت انتخابی ورود، نگهداشت و خروج اطلاعات را تنظیم می‌کند و بدین‌ترتیب امکان انتقال پایدار گرادیان‌ها را در طول توالی زمانی فراهم می‌سازد؛ ازاین‌رو، برخلاف RNNهای متعارف که در هر گام زمانی حالت پنهان را به‌طور کامل بازنویسی می‌کنند، LSTM قادر است الگوهای معنادار شناسایی‌شده در مراحل اولیه توالی را در بازه‌های زمانی طولانی حفظ کرده و وابستگی‌های زمانی بلندمدت را با دقت بالاتری مدل‌سازی کند (شکل ۱).




**شکل۱ طرح‌واره‌ای از شبکه LSTM**

## ویژگی‌ها و خصوصیات کلیدی شبکه‌های LSTM

توانایی شبکه LSTM در مدل‌سازی وابستگی‌های زمانی بلندمدت، مستقیماً از ساختار متمایز آن نسبت به RNN های متعارف ناشی می‌شود؛ به‌گونه‌ای که وجود حلقه خطای ثابت[1] باعث می‌شود حالت سلول $C_t$ به‌صورت جمعی و خطی به‌روزرسانی شده و گرادیان‌ها در فرایند BPTT، بدون تضعیف نمایی منتقل شوند (Hochreiter & Schmidhuber, 1997). افزون بر این، سازوکار دروازه‌ای با حذف انتخابی اطلاعات غیرمرتبط، کنترل ورود داده‌های معنادار و جداسازی حالت سلول از خروجی لحظه‌ای، امکان نگهداشت اطلاعات را در بازه‌های زمانی طولانی فراهم می‌کند (Gers et al., 2000). این معماری، در کنار ماهیت جمعی به‌روزرسانی و تفکیک صریح میان $C_t$ و $h_t$، به پایداری فرآیند یادگیری کمک کرده و با بهره‌گیری از توابع فعال‌سازی سیگموید و تانژانت هایپربولیک، کنترل مؤثر جریان اطلاعات در شبکه را امکان‌پذیر می‌سازد. در نهایت،

[1] Constant Error Carousel (CEC)

انعطاف‌پذیری ذاتی LSTM و توسعه گونه‌هایی مانند اتصالات پیپ‌هول[1] و واحدهای بازگشتی دروازه‌دار[2]، قابلیت استخراج الگوهای زمانی و مکانی پیچیده را به‌طور قابل‌توجهی تقویت کرده است (Graves & Schmidhuber, 2005).

## توصیف مخزن مورد مطالعه

مجموعه‌داده‌ای که در این مطالعه مورد استفاده قرار گرفته است، متعلق به یکی از بزرگ‌ترین میادین نفتی خاورمیانه با سابقه طولانی تولید است. این مخزن کاملاً ناهمگن است، به‌گونه‌ای که در طول زمان، سطوح مختلفی از تماس آب-نفت در سراسر میدان مشاهده شده است. این مخزن دارای ۱۸۰ چاه تولیدی است و مهمترین مکانیزم تولیدی از آن رانش آبده است و این آبده تا حدی قوی و فعال است که تغییرات اندکی در فشار مخزن ناشی از تولید ایجاد می‌شود. روند فشار مخزن و محاسبات موازنه مواد، وجود آبده قوی را تأیید کرده اند. این مخزن به لحاظ لیتولوژی بسیار پیچیده است و غالباً ماسه‌ای، کربناته دولومیتی و کلسیتی و شیل ماسه‌ای است، که به ۴ سکتور و ۱۰ لایه تقسیم بندی شده است. مخزن مورد مطالعه غالبا شکافدار در هر دو ناحیه ماسه ای و کربناته است. نواحی ماسه ای با متوسط تخلخل ۲۲-۲۴ درصد، حدود ۳۵ تا ۴۰ درصد از کل مخزن را تشکیل می دهند و سایر مخزن مورد مطالعه را نواحی کربناته با متوسط تخلخل ۱۲-۱۵ درصد تشکیل می دهد. یکی از چالش‌های اصلی در مورد این مخزن عدم قطعیت در عملکرد و توان تولیدی آن است و ارائه روشی که بتواند این عملکرد را پیش بینی و شبیه سازی کند بسیار حائز اهمیت است. از سویی، میزان تولید آب از این مخزن بالا است و بسیاری از چاه‌ها نرخ بالایی از برش آب را تجربه کرده‌اند و می‌توان گفت دلیل اصلی تعمیر و یا بسته شدن چاه‌های این مخزن مقدار بالای برش آب است و افت فشار مخزن و یا تولید گاز نقش زیادی در این موضوع ندارد. مجموع این عوامل، این نتیجه را می‌دهد که پیش بینی برش آب تحت شرایط و عوامل تولیدی و عملیاتی مختلف می‌تواند تخمینی از عملکرد مخزن و توان بالقوه آن در سال‌های آتی بدست دهد. برخی از این عوامل تولیدی شامل نوع لایه تکمیل شده، طول و عمق مشبک‌کاری، نرخ تولید نفت و ... می باشد که محدوده این پارامترها برای سکتورهای مختلف این مخزن در جدول ۲ ارائه شده است.

**جدول۲ ویژگی‌های آماری متغیرهای تأثیرگذار مربوط به ۴ بخش مخزن مورد مطالعه**

| شماره سکتور مخزن | پارامتر | تراوایی (md) | طول مشبک‌کاری (m) | عمق مشبک‌کاری (m) | نرخ تولید نفت (STBD) | برش آب (%) |
|---|---|---|---|---|---|---|
| ۱ | حداقل | ۲.۷۰ | ۲.۰۰ | ۲۵۷۹.۰۰ | ۱۲.۴۷ | ۰.۰۴ |
| | حداکثر | ۱۰۳۷.۸۰ | ۱۹.۰۰ | ۲۹۸۶.۵۰ | ۹۸۰۰.۰۰ | ۵۴.۰۰ |
| | میانگین | ۳۷۹.۸۸ | ۷.۶۷ | ۲۷۱۲.۹۱ | ۲۸۹۵.۷۴ | ۹.۹۱ |
| | انحراف معیار | ۳۲۹.۰۱ | ۴.۱۲ | ۷۹.۷۷ | ۱۷۴۴.۲۶ | ۱۰.۱۴ |
| ۲ | حداقل | ۸.۷۰ | ۲.۰۰ | ۲۴۷۳.۳۰ | ۲۱.۲۰ | ۰.۰۰ |
| | حداکثر | ۱۴۳۸.۷۰ | ۲۵.۰۰ | ۲۹۷۴.۰۰ | ۲۰۶۶۹.۶۷ | ۵۰.۶۴ |
| | میانگین | ۵۷۰.۷۵ | ۸.۰۲ | ۲۶۳۰.۵۰ | ۲۵۸۰.۳۲ | ۱۰.۳۶ |

[1] peephole
[2] Gated Recurrent Unit (GRU)

| | | | | | | |
|---|---|---|---|---|---|---|
| | انحراف معیار | ۴۶۵.۶۳ | ۴.۸۱ | ۱۰۵.۰۳ | ۲۳۶۷.۹۲ | ۱۱.۸۸ |
| ۳ | حداقل | ۸.۷۰ | ۲.۰۰ | ۲۵۰۶.۵۰ | ۱۸.۵۲ | ۰.۰۱ |
| | حداکثر | ۱۴۳۸.۷۰ | ۲۰.۰۰ | ۲۸۲۱.۰۰ | ۱۰۸۶۴.۵۳ | ۶۴.۵۴ |
| | میانگین | ۴۹۴.۷۱ | ۶.۷۹ | ۲۶۶۶.۸۱ | ۲۶۶۴.۷۳ | ۱۱.۱۶ |
| | انحراف معیار | ۵۰۹.۷۸ | ۳.۸۴ | ۷۶.۹۶ | ۱۸۸۰.۱۵ | ۱۲.۹۲ |
| ۴ | حداقل | ۸.۷۰ | ۲.۰۰ | ۲۴۵۳.۸۰ | ۳۶.۰۳ | ۰.۰۲ |
| | حداکثر | ۱۴۳۸.۷۰ | ۲۴.۰۰ | ۲۷۹۱.۳۰ | ۱۸۳۶۵.۸۷ | ۵۶.۰۲ |
| | میانگین | ۴۲۸.۹۷ | ۸.۰۶ | ۲۶۱۸.۷۸ | ۲۴۷۴.۹۴ | ۱۲.۱۲ |
| | انحراف معیار | ۴۲۸.۵۸ | ۴.۵۳ | ۷۲.۲۵ | ۱۹۳۰.۴۷ | ۱۳.۸۶ |

## آماده‌سازی و پیش‌پردازش داده‌ها

### حذف داده‌های تکراری و ویژگی‌های کم‌واریانس

در این مرحله، داده‌های تکراری و ویژگی‌های کم‌واریانس از پایگاه داده پالایش شدند. داده‌های تکراری شامل نمونه‌هایی بود که در آن‌ها مقادیر پارامترهای مستقل و متغیر پاسخ به‌طور کامل یکسان بودند که در این حالت، تنها نخستین نمونه نگه داشته شد. در مقابل، نمونه‌هایی که علی‌رغم یکسان‌بودن پارامترهای مستقل، مقادیر متفاوتی برای متغیر پاسخ داشتند، به‌دلیل عدم امکان تشخیص مقدار معتبر، به‌طور کامل از مجموعه داده حذف شدند. همچنین، ویژگی‌هایی که تغییرپذیری بسیار ناچیزی در میان داده‌ها داشتند و عملاً اطلاعات مفیدی برای فرآیند یادگیری فراهم نمی‌کردند، پس از نرمال‌سازی داده‌ها در بازه [۰,۱] و با استفاده از آستانه واریانس کمتر از ۰.۰۰۵ شناسایی و از مجموعه ورودی مدل کنار گذاشته شدند.

### جایگزینی مقادیر مفقود

در بسیاری از مطالعات کاربردی، مجموعه داده‌ها به‌صورت کامل در دسترس نبوده و برخی از پارامترهای مؤثر با مقادیر مفقود مواجه هستند. هرچند حذف سطرها یا ستون‌های ناقص ساده‌ترین راهکار به شمار می‌آید، اما این رویکرد با از دست رفتن بخش قابل‌توجهی از اطلاعات همراه بوده و برای مدل‌های داده‌محور، که کارایی آن‌ها به حجم داده وابسته است، گزینه مناسبی محسوب نمی‌شود. ازاین‌رو، جایگزینی مقادیر مفقود با روش‌هایی نظیر استفاده از شاخص‌های گرایش مرکزی (میانگین، میانه یا نما) یا رویکردهای مبتنی بر مدل‌سازی، راهکاری منطقی‌تر تلقی می‌گردد. در این راستا، ابتدا میزان داده‌های مفقود هر ویژگی مورد بررسی قرار گرفت و از ابزار Missingno (Bilogur, 2018) به‌منظور تجسم الگوی فقدان داده‌ها استفاده شد. در مواردی که سهم داده‌های مفقود یک ویژگی از آستانه مشخصی فراتر می‌رفت و جایگزینی آن با عدم‌قطعیت بالایی همراه بود، حذف آن ویژگی ترجیح داده شد؛ به‌طوری‌که در این مطالعه، آستانه ۵۰ درصد به‌عنوان معیار حذف ویژگی‌های دارای فقدان داده قابل‌توجه در نظر گرفته شد

### مقیاس‌بندی داده‌ها

وجود متغیرهایی با مقیاس‌های عددی متفاوت می‌تواند فرآیند یادگیری مدل را دچار اختلال کرده و به کاهش دقت پیش‌بینی منجر شود؛ ازاین‌رو، کلیه پارامترهای پایگاه داده در یک بازه یکنواخت مقیاس‌بندی شدند. این اقدام، علاوه بر

تسهیل و تسریع فرآیند بهینه‌سازی، اثرات نامطلوب ناشی از اختلاف مقیاس میان متغیرها را حذف کرده و امکان مشارکت متوازن آن‌ها در فرآیند یادگیری مدل را فراهم می‌سازد. بدین منظور، از رابطه (۱) برای نرمال‌سازی داده‌ها و نگاشت آن‌ها به یک بازه استاندارد استفاده شد. در این مطالعه، پارامترهای نرمال‌سازی صرفاً بر اساس داده‌های آموزش محاسبه گردید و پس از طراحی مدل، همان تبدیل بدون هیچ‌گونه تغییر بر مجموعه داده‌های آزمون اعمال شد تا از نشت اطلاعات جلوگیری شود. در رابطه (۱)، $X_i$ مقدار هر پارامتر بوده و Mean(X) و STD(X) به‌ترتیب میانگین و انحراف معیار آن پارامتر را نشان می‌دهند.

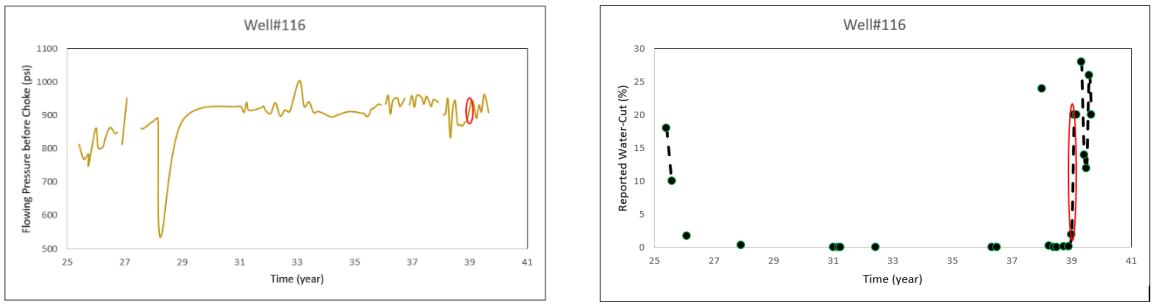
$$X_i^{Normalized} = \frac{X_i - Mean(X)}{STD(X)} \quad (۱)$$

**حذف داده‌های پرت**

وجود داده‌های پرت در فرآیند مدل‌سازی می‌تواند موجب ناپایداری یادگیری و کاهش دقت پیش‌بینی، به‌ویژه در نواحی از فضای داده شود که تراکم نمونه‌ها در آن‌ها محدود است. به‌طور مشخص، حضور نقاط داده‌ای دورافتاده (Outliers) باعث تغییر شکل توزیع داده‌ها و انحراف آن از حالت نرمال شده و در نتیجه، عملکرد مدل را تحت تأثیر قرار می‌دهد. ازاین‌رو، حذف این دسته از داده‌ها به‌عنوان رویکردی مؤثر برای افزایش پایداری و دقت مدل در نظر گرفته می‌شود. در این مطالعه، به‌منظور شناسایی و حذف داده‌های پرت، آستانه‌ای معادل ۹ برابر انحراف معیار نسبت به مقدار میانگین اعمال گردید. شایان ذکر است که عبور داده از آستانه ۹ برابر انحراف معیار به تنهایی ملاک حذف نبود؛ بلکه روندهای افزایشی یا کاهشی مربوط به رخنه آب به صورت دستی و با در نظر گرفتن سازگاری فیزیکی با رویدادهای واقعی مخزن (مانند رخنه آب، تغییر لایه یا تغییر نرخ تولید) بررسی شد. تنها نقاطی که از نظر فیزیکی غیرممکن یا ناسازگار بودند، حذف گردیدند و آن‌هایی که با رفتار فیزیکی مخزن (مانند شروع ناگهانی افزایش برش آب پس از مدتی تولید با برش کم) همخوانی داشتند، حذف نشدند.

**ناسازگاری فنی میان پارامترها**

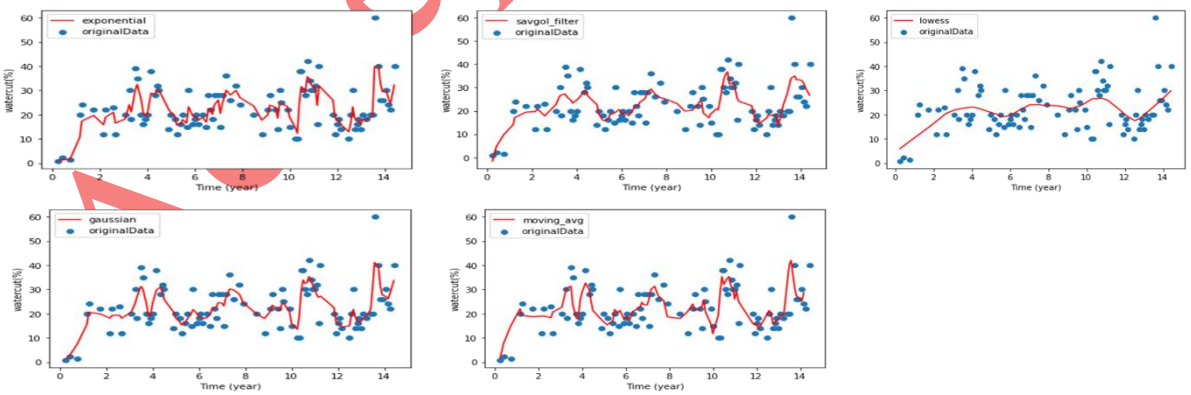
یکی از ملاحظات اساسی در فرآیند پردازش داده‌های ورودی، اطمینان از یکنواختی و سازگاری فنی میان پارامترهای مورد استفاده در مدل است. در این مطالعه، یکی از ناسازگاری‌های شناسایی‌شده در داده‌های برش آب به عدم همخوانی میان روند تغییرات برش آب و فشار جریانی چاه بازمی‌گردد؛ به‌طوری‌که در برخی بازه‌ها، افزایش برش آب هم‌زمان با افزایش فشار جریانی پیش از کاهنده مشاهده شده است. از آنجا که چنین رفتاری از منظر مهندسی تولید و هیدرودینامیکی قابل توجیه نیست، این موارد به‌عنوان داده‌های ناسازگار فنی تلقی شده و وجود خطا در ثبت یا پردازش داده‌ها را نشان می‌دهد. شکل۲ داده‌های برش آب و فشار جریانی قبل از کاهنده برای چاه شماره ۱۱۶ در میدان مورد مطالعه را نشان می‌دهد. نقاط قرمزرنگ نمونه‌هایی از عدم‌سازگاری میان مقادیر گزارش‌شده برش آب و فشار جریانی پیش از کاهنده هستند.




**شکل۲ داده‌های فشار جریانی قبل از کاهنده و برش آب برای چاه شماره ۱۱۶**

## هموارسازی داده‌ها

یکی از مراحل کلیدی در آماده‌سازی داده‌ها پیش از اجرای هرگونه مدل‌سازی داده‌محور، هموارسازی داده‌ها است؛ به‌ویژه در مواردی که داده‌ها از یکنواختی کافی برخوردار نبوده و تحت تأثیر نویزهای اندازه‌گیری یا عملیاتی قرار دارند. برای این منظور، روش‌های متعددی در ادبیات موجود است که از جمله مهم‌ترین آن‌ها می‌توان به هموارسازی نمایی، فیلتر ساویتسکی– گلای، هموارسازی لاوس، هموارسازی گوسی و روش میانگین متحرک اشاره کرد. در این مطالعه، تمامی روش‌های یادشده به‌صورت مقایسه‌ای بر روی داده‌ها اعمال گردید و نتایج آن‌ها از نظر پایداری روند، حفظ رفتار فیزیکی داده‌ها و کاهش نوسانات ناخواسته مورد ارزیابی قرار گرفت. بر اساس این بررسی‌ها، روش میانگین متحرک عملکرد مناسب‌تری نسبت به سایر روش‌ها نشان داد و به‌عنوان روش نهایی هموارسازی انتخاب شد. در این مطالعه، روش میانگین متحرک ساده (با وزن های مساوی) با پنجره ای به اندازه ۵ نقطه داده متوالی به کار گرفته شد. برای نقاط ابتدا و انتهای سری زمانی، از پنجره‌های نامتقارن استفاده گردید. انتخاب پنجره ۵ نقطه‌ای پس از مقایسه با پنجره‌های ۳، ۷ و ۹ نقطه‌ای انجام شد و بهترین تعادل را بین کاهش نویز و حفظ رویدادهای فیزیکی ناگهانی (مانند رخنه آب و تغییر لایه تولیدی) ارائه داد. نمونه‌ای از تأثیر به‌کارگیری روش‌های مختلف هموارسازی بر داده‌های یکی از چاه‌های مخزن مورد مطالعه در شکل ۳ ارائه شده است.

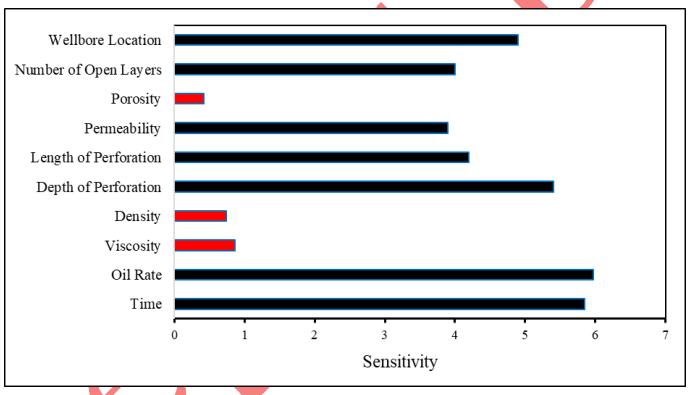



**شکل۳ نتایج حاصل از به‌کارگیری روش‌های مختلف هموارسازی داده بر روی یکی از چاه‌های مخزن مورد مطالعه**

## انتخاب ویژگی‌ها

در این مطالعه، در گام نخست و پیش از نهایی‌سازی مجموعه ورودی‌های مدل، کلیه داده‌های در دسترس مرتبط با مخزن شامل موقعیت چاه، تعداد و لیتولوژی لایه‌های مشبک‌کاری‌شده (ماسه‌ای، کربناته و شیلی)، ویژگی‌های سنگ نظیر تخلخل

و تراوایی، خواص سیال از جمله گرانروی و چگالی، عمق و طول مشبک‌کاری، نرخ تولید نفت و زمان تولید به‌عنوان متغیرهای ورودی و برش آب به‌عنوان متغیر خروجی در نظر گرفته شد. با توجه به تعداد زیاد این پارامترها و پیچیدگی فرآیندهای مؤثر بر تولید آب، افزایش ابعاد فضای ورودی می‌توانست احتمال بیش‌برازش مدل را به‌طور قابل‌توجهی افزایش دهد؛ ازاین‌رو، کاهش تعداد متغیرهای ورودی و انتخاب ویژگی‌های مؤثرتر به‌عنوان یک گام ضروری مدنظر قرار گرفت. بدین منظور، از تحلیل حساسیت مونت‌کارلو برای شناسایی پارامترهای کم‌اثر بر رفتار برش آب استفاده شد. افزون بر این، به‌منظور جلوگیری از افزایش غیرضروری ابعاد مدل ناشی از تعدد لایه‌ها و تفاوت لیتولوژی آن‌ها، به‌جای ورود مستقل مشخصات هر لایه، یک ویژگی فیزیکی یکپارچه نظیر تراوایی به‌عنوان نماینده رفتار لایه‌ها در ساختار ورودی مدل لحاظ گردید. اهمیت تراوایی به‌عنوان یکی از متغیرهای کلیدی در توصیف رفتار مخزن در مطالعات داده‌محور داخلی نیز مورد تأکید قرار گرفته است؛ به‌طوری‌که مرادی و همکاران (مرادی و همکاران، ۲۰۲۲ et al.,) با استفاده از شبکه عصبی پرسپترون چندلایه، تراوایی را از داده‌های لاگ چاه تخمین زده و سپس توزیع آن را در مقیاس میدان شبیه‌سازی کردند. نتایج تحلیل حساسیت (شکل ۴) نشان داد که پارامترهایی نظیر تخلخل، چگالی و گرانروی کمترین تأثیر را بر مقدار برش آب دارند؛ نتیجه‌ای که با رفتار واقعی مخزن نیز همخوانی دارد، زیرا به‌واسطه پشتیبانی قوی آبده، فشار مخزن در طول زمان تقریباً ثابت باقی مانده و در نتیجه، تغییرات خواص سیال نقش تعیین‌کننده‌ای در روند برش آب ایفا نمی‌کنند.




**شکل ۴ نتایج تحلیل حساسیت مونت کارلو جهت تعیین ورودی‌های مدل پیشنهادی**

بر اساس یافته‌های تحلیل حساسیت، پارامترهای کلیدی شامل مختصات چاه، تعداد لایه‌های باز، تراوایی، طول و عمق مشبک‌کاری، نرخ تولید نفت و زمان به‌عنوان ورودی‌های نهایی مدل انتخاب شدند و برش آب نیز به‌عنوان خروجی تعیین گردید. مطابق بررسی کل داده‌های میدان برای برخی از چاه‌ها، چندین لایه به طور همزمان مشبک‌کاری انجام شده لذا به منظور اعمال این رویداد، در مدل LSTM طراحی شده، حداکثر چهار لایه باز برای هر چاه لحاظ شد؛ درنتیجه، مدل به چهار مقدار تراوایی، چهار طول مشبک‌کاری و چهار عمق مشبک‌کاری نیاز دارد. با این ساختار، مدل نهایی شامل ۱۷ ورودی و یک خروجی است. برای چاه‌هایی که تعداد لایه‌های باز کمتر از چهار لایه است، از روش Zero Padding برای لایه‌های غیرفعال استفاده شده است (مقادیر تراوایی، طول مشبک و عمق مشبک برابر صفر قرار داده می‌شوند). ترتیب لایه‌ها در ورودی مدل بر اساس عمق از کم عمق به عمیق مرتب شده است. به عنوان مثال، کم عمق‌ترین لایه در موقعیت اول و عمیق‌ترین لایه در موقعیت چهارم قرار می‌گیرد.

مجموعه داده نهایی شامل ۳۳,۶۹۷ نمونه از ۱۸۰ چاه تولیدی است. بنابراین، میانگین تعداد نمونه به ازای هر چاه حدود ۱۸۷ نمونه است. دامنه تغییرات این تعداد از کمتر از ۵۰ نمونه (برای چاه‌های با تاریخچه تولید کوتاه یا تازه حفاری‌شده) تا بیش از ۴۰۰ نمونه (برای چاه‌های قدیمی با تاریخچه تولید طولانی) متغیر است. بازه تولیدی از چاه ها در این میدان از ۵۰ سال تا حدود ۵ سال متغیر است و داده های ثبت شده مربوط به سال ۱۳۷۷ تا انتهای ۱۴۰۳ می باشد. داده های مربوط به تمامی ۱۸۰ چاه تولیدی در این میدان، در روند مدلسازی مورد استفاده قرار گرفتند. پایگاه داده خام اولیه شامل ۳۴,۷۵۰ رکورد بود. پس از اعمال مراحل مختلف پیش‌پردازش، در نهایت ۳۳,۶۹۷ رکورد برای مدل‌سازی باقی ماند که معادل ۹۷.۰ درصد داده‌های اولیه است. در مجموع، ۱,۰۵۳ رکورد (حدود ۳.۰ درصد) حذف گردید. تأثیر هر مرحله از پیش‌پردازش بدین شرح است: ۱. حذف داده‌های تکراری: در این مرحله، رکوردهای کاملاً یکسان و همچنین رکوردهایی که پارامترهای مستقل یکسان ولی مقادیر متفاوت برش آب داشتند (غیرقابل اعتماد) حذف شدند. تعداد ۵۲۰ رکورد (حدود ۱.۵ درصد از داده‌های اولیه) در این مرحله حذف گردید. ۲. حذف داده‌های پرت: با استفاده از آستانه ۹ برابر انحراف معیار نسبت به میانگین، داده‌های پرت شناسایی و حذف شدند. این مرحله منجر به حذف ۳۱۰ رکورد (حدود ۰.۹ درصد از داده‌های اولیه) شد. ۳. حذف ناسازگاری فنی: موارد نادر عدم همخوانی بین روند تغییرات برش آب و فشار جریانی پیش از کاهنده (مانند شکل ۲ برای چاه ۱۱۶) بررسی و حذف شدند. در این مرحله، ۲۲۳ رکورد (حدود ۰.۶ درصد از داده‌های اولیه) حذف گردید. شایان ذکر است که حذف داده‌ها صرفاً بر اساس معیارهای آماری و فنی و با هدف افزایش کیفیت داده‌های ورودی انجام شده است و داده‌های مربوط به رویدادهای فیزیکی واقعی (مانند رخنه آب) در این فرآیند حفظ شده‌اند.

## نتایج و بحث

در این پژوهش، از شبکه LSTM برای پیش‌بینی رفتار برش آب در چاه‌های تولیدی یک مخزن نفتی عظیم استفاده شده است. مخزنی که به‌واسطه ظرفیت بالای تولید آب و پشتیبانی قوی آبده، با وجود ثابت‌ماندن نسبی فشار ایستای مخزن و نبود چالش‌های مرتبط با تولید گاز، به‌طور فزاینده‌ای تحت تأثیر افزایش برش آب قرار دارد. تجربه بهره‌برداری میدان نشان می‌دهد که در چنین شرایطی، افزایش برش آب عملاً به عامل اصلی محدودکننده تولید و بسته‌شدن چاه‌ها تبدیل شده است. بر این اساس، دستیابی به پیش‌بینی دقیق و قابل‌اتکا از نرخ تولید آب می‌تواند مبنای مناسبی برای ارزیابی پتانسیل تولیدی چاه‌ها و تحلیل واقع‌بینانه عملکرد مخزن فراهم کند.

به‌منظور انجام مدل‌سازی، ۸۰ درصد ابتدایی داده‌های سری‌زمانی هر یک از چاه‌ها به‌عنوان مجموعه آموزش و ۲۰ درصد پایانی به‌عنوان داده‌های آزمون در نظر گرفته شد تا از نشت اطلاعات جلوگیری شود. داده‌های آزمون در هیچ‌یک از مراحل طراحی و تنظیم مدل مورد استفاده قرار نگرفتند و صرفاً برای ارزیابی عملکرد نهایی مدل به‌کار گرفته شدند. معماری مدل توسعه‌یافته شامل یک لایه ورودی دنباله‌ای، یک لایه LSTM، یک لایه کاملاً متصل، یک لایه Dropout و در نهایت یک لایه رگرسیون بود. ابرپارامترهای کلیدی مدل با استفاده از الگوریتم جستجوی شبکه‌ای (Liashchynskyi & Liashchynskyi, 2019) مطابق مشخصات ارائه‌شده در جدول ۳ تنظیم گردید. در این مطالعه، هیچ طول پنجره ثابتی در مرحله پیش‌پردازش اعمال نشده است. هر چاه با کل تاریخچه تولید خود (با طول متغیر) به عنوان یک دنباله ورودی به مدل LSTM معرفی

شده است. در متلب، از آپشن 'SequenceLength', 'longest' در تابع trainingOptions استفاده شد که دنباله‌های کوتاه‌تر را در هر مینی‌بچ تا طول بلندترین دنباله با صفر پر می‌کند. مدل به صورت پیش‌بینی یک گام به جلو (one-step ahead) عمل می‌کند. مدل LSTM در متلب نسخه ۲۰۲۳ پیاده‌سازی شد. پیش‌پردازش با اسکریپت‌های سفارشی متلب انجام گردید. بهینه‌سازی ابرپارامترها با روش جستجوی شبکه‌ای صورت گرفت و درنهایت آموزش بر روی سرور با GPU NVIDIA Tesla T4 انجام شد. در این پژوهش، هر یک از بخش‌های مخزن به‌صورت مستقل مدل‌سازی شده و تحلیل نتایج نیز برای هر بخش به‌طور جداگانه انجام گرفت. لازم به ذکر است که به دلیل وجود چهار سکتور طبیعی مجزا (تفکیک شده توسط گسل‌های اصلی و تفاوت‌های سنگ‌شناسی) و همچنین حجم بالای داده‌ها (۱۸۰ چاه و بیش از ۳۳ هزار رکورد)، مدل‌سازی مجزای هر سکتور نسبت به یک مدل یکپارچه برای کل میدان، دقت پیش‌بینی بهتری ارائه می دهد. از آنجا که مخزن به چهار سکتور مجزا تقسیم شده است، اطمینان از یکپارچگی فیزیکی برای چاه‌های مستقر در نواحی مرزی ضروری است. در این مطالعه، ۲۷ چاه مرزی شناسایی شدند. برای هر یک از این چاه‌ها، مدل LSTM سکتور مجاور نیز اجرا گردید و نتایج مقایسه شد. در بیش از ۹۵٪ موارد، مدل سکتوری که چاه بر اساس داده‌های تکمیل در آن قرار داشت، دقت بالاتری ارائه داد. همچنین، همبستگی خطاهای پیش‌بینی بین چاه‌های دو سوی مرز محاسبه گردید که عدم وجود همبستگی منفی معنادار، تأییدکننده یکپارچگی فیزیکی مدل‌های مستقل است. به‌منظور ارزیابی عملکرد مدل، معیار ریشه میانگین مربعات خطا به‌عنوان شاخص اصلی در نظر گرفته شد که هدف فرآیند آموزش، کمینه‌سازی این معیار بود.

**جدول ۳ مشخصات به‌کاررفته برای ساخت شبکه**LSTM

| پارامتر | مقدار / ویژگی‌ها |
|---|---|
| تعداد واحدهای پنهان | ۶۴ |
| نرخ حذف تصادفی | ۰.۱ |
| بهینه‌ساز | Adam |
| حداکثر تعداد دوره‌های آموزشی | ۵۰۰ |
| نرخ یادگیری | ۰.۰۰۱ |
| حداقل اندازه دسته | ۱۰۲۴ |
| Verbose | ۰ |

به منظور بررسی و ارزیابی عملکرد مدل، پارامترهای آماری زیر مورد استفاده قرار گرفتند:

$$ARE\% = \left|\frac{O_i^{pred} - O_i^{exp}}{O_i^{exp}}\right| \times 100 \quad (۲)$$

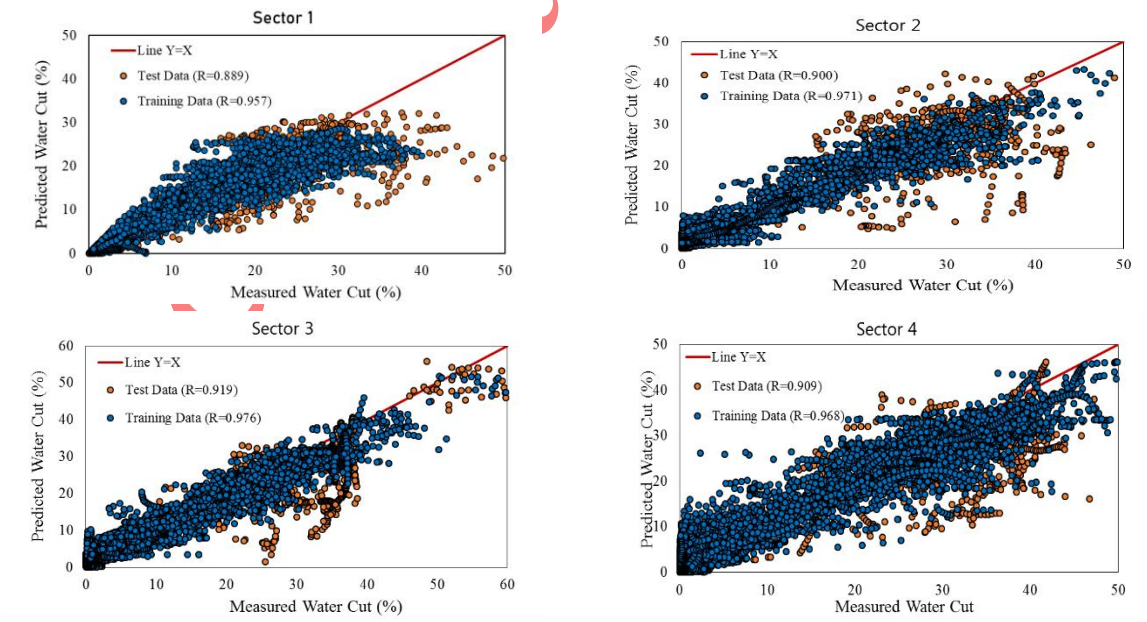
$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(O_i^{pred} - O_i^{exp}\right)^2}{N}} \quad (۳)$$

$$R = \frac{\sum_{i=1}^{n}\left(O_i^{exp} - \overline{O^{exp}}\right)\left(O_i^{pred} - \overline{O^{pred}}\right)}{\sqrt{\sum_{i=1}^{n}\left(O_i^{exp} - \overline{O^{exp}}\right)^2} - \sqrt{\sum_{i=1}^{n}\left(O_i^{pred} - \overline{O^{pred}}\right)^2}} \quad (۴)$$

در معادلات فوق، زیرنویس i نشان‌دهنده نقطه‌داده‌ای است که مورد ارزیابی قرار می‌گیرد، در حالی‌که مقادیر pred و exp به‌ترتیب بیانگر مقادیر پیش‌بینی‌شده توسط مدل و مقدار برش آب به‌دست‌آمده از داده‌های تجربی هستند. پارامتر N نیز تعداد کل نقاط موجود در مجموعه‌داده را نشان می‌دهد.

پس از ساخت مدل، عملکرد آن در فازهای آموزش و آزمون برای هر یک از بخش‌های مخزن ارزیابی شد. نمودارهای پراکنش در شکل۵ ارائه شده‌اند. نقاط داده که خوشه‌ای فشرده در امتداد خط ۴۵ درجه تشکیل می‌دهند، نشان‌دهنده مدل‌هایی با بالاترین دقت پیش‌بینی هستند. مدل توسعه‌یافته در تمامی ۴ بخش مخزنی، عملکرد و توان تعمیم‌دهی بسیار مطلوبی از خود نشان داد. ضریب همبستگی پیرسون (R) برای داده‌های آموزش در بازه ۰.۹۵۷ تا ۰.۹۷۶ قرار گرفت که بیانگر توان بالای مدل در یادگیری روابط پیچیده و غیرخطی است. نکته مهم این است که مدل در داده‌های آزمون نیز، که مستقل از فرآیند آموزش هستند، دقت پیش‌بینی بسیار مطلوبی ارائه داد و مقادیر R در بازه ۰.۸۸۹ تا ۰.۹۲۰ به‌دست آمد. این اختلاف اندک بین عملکرد مدل در فاز آموزش و آزمون نشان می‌دهد که مدل به‌طور مؤثر از بیش‌برازش -که یکی از چالش‌های اصلی در مدل‌سازی داده‌محور پیچیده است- اجتناب کرده است. یکنواختی نتایج آزمون نیز بسیار قابل‌توجه است. تنوع بسیار اندک $R_{test}$ در تمامی بخش‌های مخزن، نشان‌دهنده پایداری، استحکام و توانایی مدل در ثبت رفتارهای بنیادی مخزن است؛ رفتارهایی که در کل میدان یکسان هستند. بخش ۳ بالاترین دقت پیش‌بینی را ارائه داد ($R_{train} = 0.976$, $R_{test} = 0.920$) که نمونه‌ای برجسته از توان مدل محسوب می‌شود.

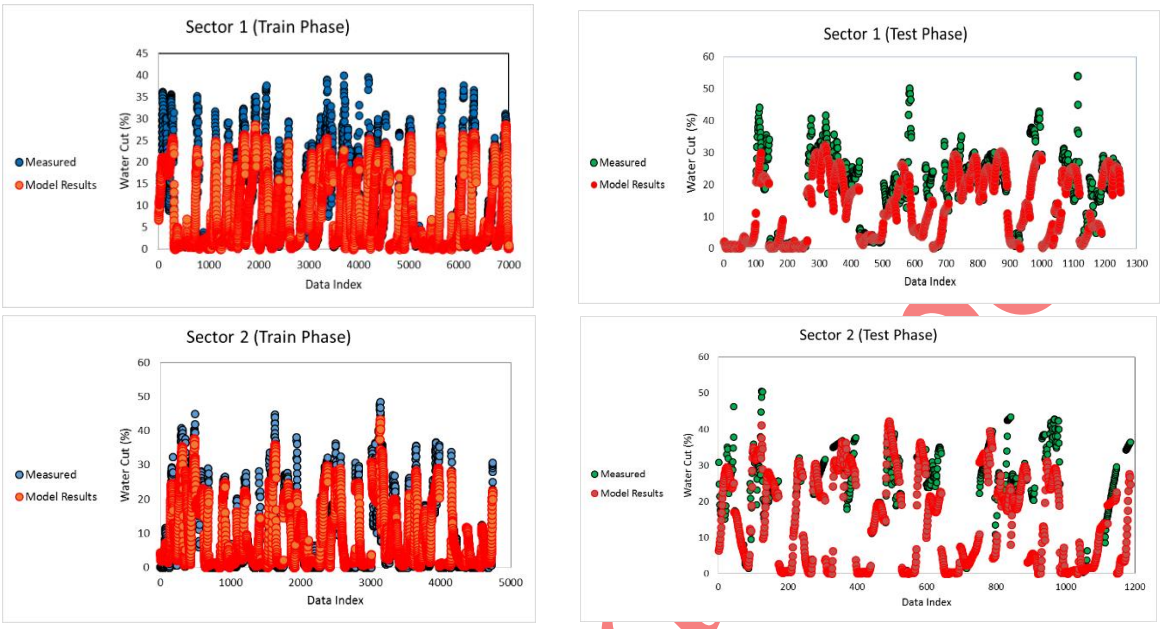
ضرایب همبستگی بالا در داده‌های آموزش و آزمون، همراه با یکپارچگی و سازگاری دقیق بین همه بخش‌ها، نشان می‌دهد که مدل توسعه‌یافته ابزاری بسیار قابل‌اعتماد، پایدار و کارآمد برای پیش‌بینی برش آب است. توان تعمیم‌دهی قوی مدل، ارزش ویژه‌ای در تصمیم‌گیری‌های مرتبط با راهبرد تولید و مدیریت مخزن در سطح کل میدان ایجاد می‌کند.




**شکل۵ نمودار تقاطعی مقادیر واقعی و پیش‌بینی برش آب توسط مدل LSTM**

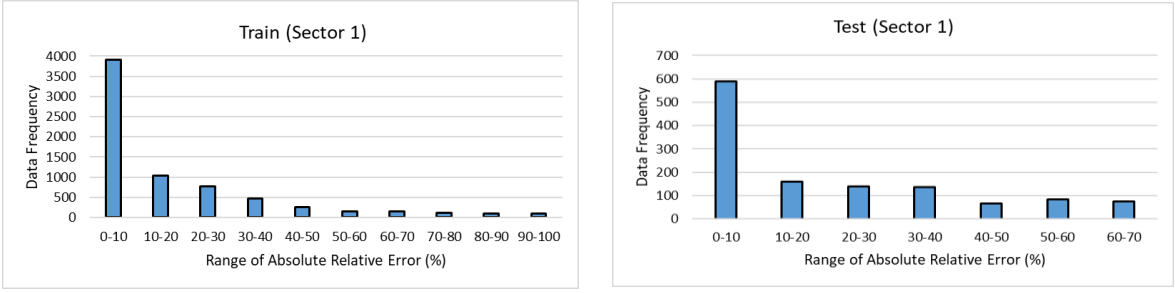
علاوه بر موارد بالا، برای ارزیابی عملکرد مدل، از نمودار پراکنش ترتیبی شامل مقادیر واقعی و پیش‌بینی‌شده برش آب برحسب شاخص داده استفاده شد. همان‌گونه که در شکل۶ برای ۲ بخش از مخزن نشان داده شده است، پیش‌بینی‌های

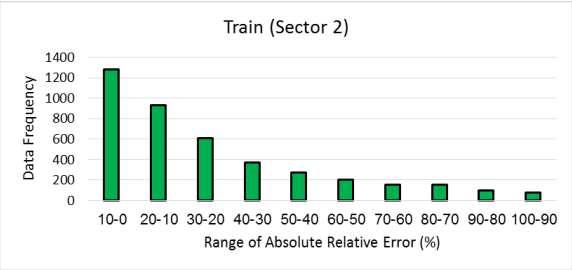
مدل در هر دو فاز آموزش و آزمون تطابقی چشمگیر با داده‌های واقعی دارند. این تحلیل بصری دو نقطه قوت اصلی مدل را آشکار می‌سازد: نخست، توانایی مدل در بازنمایی رفتار دینامیکی سامانه و بازتولید دقیق روندهای بلندمدت و نوسانات کوتاه‌مدت برش آب؛ دوم، حفظ دقت نقطه‌به‌نقطه در کل دامنه داده‌ها که بیانگر توانایی مدل در پیش‌بینی مقادیر با انحراف حداقلی در هر گام زمانی است.

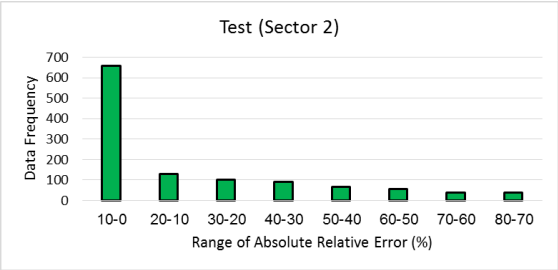



**شکل۶ پیش‌بینی برش آب در فاز آموزش و آزمون. محور افقی 'شاخص داده' نشان‌دهنده شماره ترتیبی زمانی هر نمونه پس از پیش‌پردازش است (از قدیم به جدید).**

برای ارزیابی دقت و پایداری مدل، توزیع خطای نسبی مطلق در بخش‌های مختلف مخزن و در هر دو مجموعه آموزش و آزمون مورد بررسی قرار گرفت. همان‌گونه که در شکل۷ برای دو بخش از مخزن نشان داده شده است، توزیع ARE در تمامی بخش‌ها عمدتاً در پایین‌ترین بازه‌های خطا متمرکز بوده که بیانگر عملکرد مدل در ناحیه خطای بسیار کم برای بخش عمده پیش‌بینی‌ها است. نکته قابل‌توجه، شباهت چشمگیر الگوی توزیع خطا میان بخش‌های مختلف در فاز آموزش است؛ روندی که در فاز آزمون نیز مشاهده می‌شود و نشان‌دهنده رفتار پایدار مدل و توانایی آن در تعمیم مؤثر به داده‌های دیده‌نشده می‌باشد. البته همان‌گونه که انتظار می‌رود، توزیع آزمون اندکی گسترده‌تر شده و یک دنباله با احتمال کم در بازه خطاهای بزرگ‌تر دیده می‌شود؛ امری طبیعی هنگام گذار از داده‌های آموزش به داده‌های جدید. در مجموع، نتایج توزیع ARE نشان می‌دهد که مدل نه‌تنها پیش‌بینی‌های دقیقی ارائه می‌دهد، بلکه از پایداری و قابلیت‌اعتماد بالا در کل میدان برخوردار است و احتمال بروز خطاهای بزرگ بسیار پایین می‌باشد.





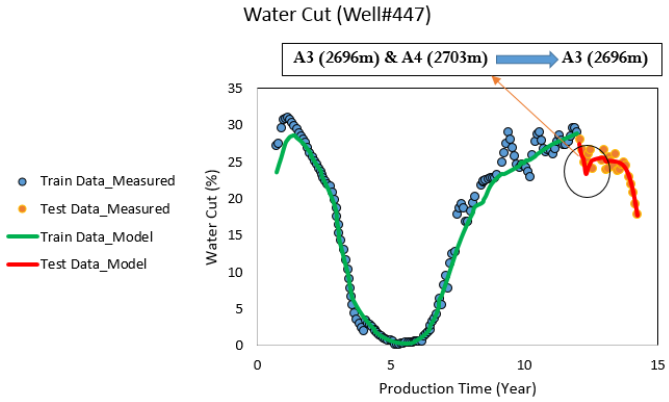

**شکل۷ فراوانی داده در برابر خطای نسبی مطلق**

لازم به ذکر است که برای ارزیابی تأثیر حجم داده، مدل LSTM با درصدهای مختلف داده آموزش (۲۰٪ تا ۱۰۰٪) آموزش داده شد. نتایج نشان داد که افزایش حجم داده از ۸۰٪ به ۱۰۰٪ تنها ۰.۰۱ بهبود در ضریب همبستگی داده های تست (میانگین) ایجاد می‌کند که بیانگر کفایت حجم داده فعلی است. همچنین، ارزیابی تأثیر مراحل پیش‌پردازش نشان داد که اعمال کامل پیش‌پردازش، میانگین ضریب همبستگی داده های تست را از ۰.۶۵ (داده خام) به ۰.۹۰ افزایش می‌دهد که حاکی از تأثیر بسیار بیشتر کیفیت داده نسبت به کمیت آن است. مقادیر معیارهای مختلف خطا برای هر چهار بخش مخزن در جدول ۴ ارائه شده است.

**جدول۴ مقادیر معیارهای خطا برای داده‌های آموزش و آزمون در چهار بخش مخزن**

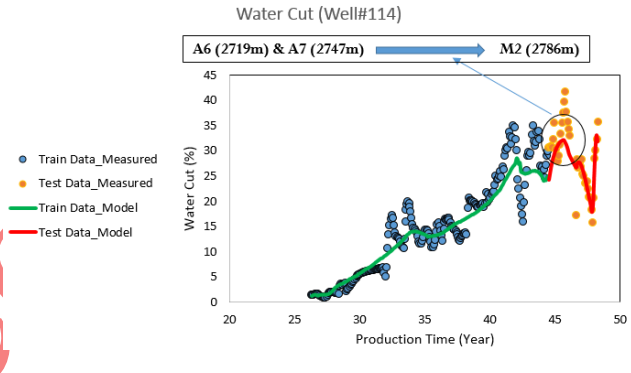
| Sector | Phase | RMSE (%) | MAE (%) | MAPE (%) | R |
|---|---|---|---|---|---|
| Sector 1 | آموزش | ۱۱.۲۴ | ۱۰.۱۸ | ۱۴.۸ | ۰.۹۵۷ |
| | آزمون | ۱۲.۵۶ | ۱۱.۳۱ | ۱۷.۴ | ۰.۸۸۹ |
| Sector 2 | آموزش | ۱۰.۸۵ | ۹.۸۷ | ۱۳.۶ | ۰.۹۷۱ |
| | آزمون | ۱۲.۳ | ۱۱.۰۵ | ۱۶.۷ | ۰.۹ |
| Sector 3 | آموزش | ۱۰.۴۵ | ۹.۶۲ | ۱۲.۸ | ۰.۹۷۶ |
| | آزمون | ۱۱.۶۸ | ۱۰.۵۴ | ۱۵.۲ | ۰.۹۱۹ |
| Sector 4 | آموزش | ۱۰.۹۵ | ۹.۹۵ | ۱۴ | ۰.۹۶۸ |
| | آزمون | ۱۲.۲ | ۱۱.۱ | ۱۶.۸ | ۰.۹۰۹ |

عملکرد مدل در بخش‌های مختلف تطابق تاریخچه و پیش‌بینی، برای چندین چاه به‌صورت جداگانه ارزیابی شد که نتایج آن در شکل‌های ۸ تا ۱۱ ارائه شده است. در شکل ۸، مربوط به چاه ۴۴۷، مشاهده می‌شود که مدل در فاز آزمون و در زمان تولید ۱۳ سال، تغییرات لایه‌ای (تغییر از A3 و A4 به A3) و اثر آن بر مقدار برش آب را با دقت بالا پیش‌بینی کرده است. همچنین در فاز آموزش نیز مدل توانسته است مقادیر برش آب را با دقت قابل‌توجهی بازتولید کند.



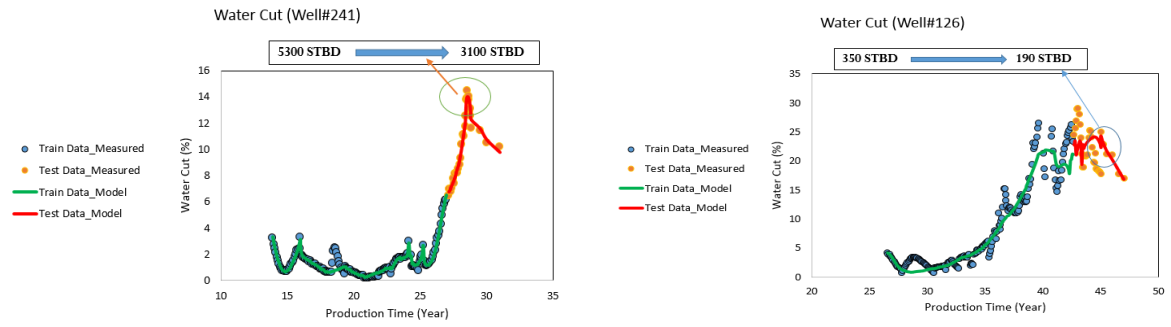

**شکل۸ مقایسه نتایج مدل LSTM با داده‌های واقعی برای چاه ۴۴۷**

در شکل ۹، مربوط به چاه ۱۱۴، برآورد مدل برای برش آب تولیدی نمایش داده شده است. در این چاه، مدل در فاز آزمون- که داده‌های آن در فرآیند طراحی مدل مورد استفاده قرار نگرفته‌اند- توانسته است اثر تغییر لایه از A6 و A7 به لایه عمیق‌تر M2 در عمق ۲۷۸۶ متری را با دقت بسیار بالا پیش‌بینی کند. این نتیجه بیانگر عملکرد ممتاز مدل در شبیه‌سازی رفتار چاه طی مرحله پیش‌بینی است. نکته مهم در نمودار آن است که برای کنترل تولید آب، لایه‌های A6 و A7 بسته شده و سپس لایه عمیق‌تر M2 باز شده است. احتمال بای‌پس‌شدن چاه وجود داشت، اما بررسی CDR مرتبط با این چاه نشان می‌دهد که بای‌پس رخ نداده است. همچنین با توجه به رفتار لایه‌ای مخزن، مشاهده برش آب کمتر در لایه‌های عمیق‌تر برای این چاه کاملا قابل‌توجیه است.



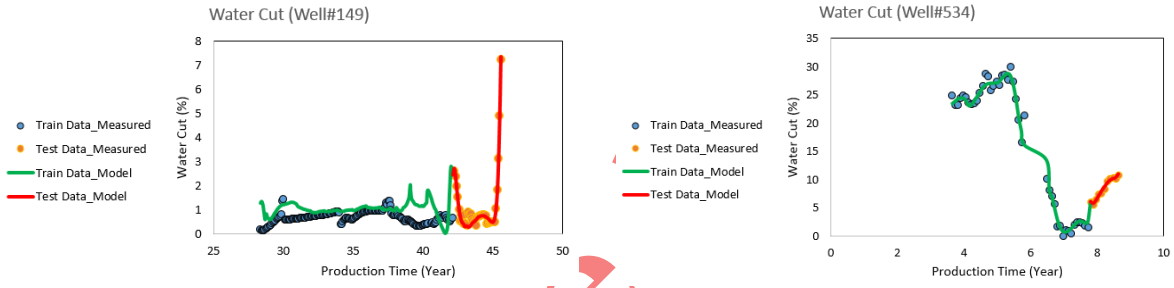

**شکل۹ مقایسه نتایج مدل LSTM با داده‌های واقعی برای چاه ۱۱۴**

یکی از عوامل مؤثر بر مقدار برش آب، نرخ تولید نفت است که معمولاً به‌عنوان ابزاری برای کنترل تولید آب در چاه‌ها به‌کار می‌رود. مدل LSTM توسعه‌یافته قادر است اثر تغییرات نرخ تولید نفت بر حجم آب تولیدی را به‌طور دقیق -حتی در فاز آزمون که داده‌های آن در فرآیند آموزش مدل هوش مصنوعی استفاده نشده‌اند- پیش‌بینی کند. این قابلیت مهم در شکل ۱۰ برای چاه‌های ۲۴۱ و ۱۲۶ به روشنی مشاهده می‌شود. در این چاه‌ها، طی فاز آزمون، نرخ تولید نفت به‌ترتیب از ۵۳۰۰ به ۳۱۰۰ بشکه در روز برای چاه ۲۴۱ و از ۳۵۰ به ۱۹۰ بشکه در روز برای چاه ۱۲۶ کاهش یافته است. مدل توانسته است اثر این کاهش بر مقدار برش آب را با دقت بسیار بالا بازتولید کند، که بیانگر توانایی مدل در درک روابط دینامیکی میان نرخ تولید و رفتار آبدهی چاه‌ها است.




**شکل۱۰ مقایسه نتایج مدل LSTM با داده‌های واقعی برای چاه ۲۴۱ و ۱۲۶**

شکل ۱۱ مقایسه‌ای بین داده‌های واقعی و نتایج مدل LSTM برای چاه‌های ۱۴۹و ۵۳۴ ارائه می‌کند. همان‌گونه که مشاهده می‌شود، نتایج مدل تطابق قابل‌قبولی با داده‌های مشاهده‌شده داشته و هم در برآورد مقادیر نقطه‌ای برش آب و هم در پیش‌بینی روند کلی تغییرات، عملکرد رضایت‌بخشی از خود نشان داده است.




**شکل۱۱ مقایسه نتایج مدل LSTM با داده‌های واقعی برای چاه ۵۳۴**

## نتیجه‌گیری

نتایج این پژوهش نشان داد که شبکه LSTM با طراحی مناسب و استفاده از داده های پیش‌پردازش شده، توان بالایی در پیش‌بینی برش آب در میادین ناهمگن با مکانیزم رانش آبی دارد. استفاده از مجموعه‌داده‌ای جامع با پیش‌پردازش دقیق و در نظر گرفتن ۱۷ ورودی دارای تفسیر فیزیکی روشن، منجر به خلق مدلی شد که بطور دقیق تعاملات پیچیده میان ویژگی‌های مخزن، شرایط تکمیل چاه و رفتار تولید را بازسازی می‌کند.

تفکیک زمانی داده‌ها (۸۰ درصد اولیه برای آموزش و ۲۰ درصد انتهایی برای آزمون) امکان ارزیابی واقع‌بینانه قابلیت تعمیم مدل به شرایط آینده و نادیده را فراهم نمود. معماری LSTM پس از بهینه‌سازی از طریق تنظیم سیستماتیک ابرپارامترها، در هر ۴ بخش مخزن عملکردی فوق‌العاده مطلوب از خود نشان داد؛ به‌طوری‌که ضرایب همبستگی پیرسون برای داده‌های آموزشی بین ۰.۹۵۷ تا ۰.۹۷۶ و برای داده‌های آزمون بین ۰.۸۸۹ تا ۰.۹۲۰ به‌دست آمد. فاصله ناچیز میان عملکرد آموزشی و آزمون، همراه با توزیع خطای نسبی مطلق متمرکز در نزدیکی صفر، نشان‌دهنده توان تعمیم‌دهی بالای مدل است. افزون بر این موارد، در فاز آزمون و بر روی داده‌های کاملاً نادیده، مدل توانست پیامدهای عملیاتی کلیدی را با دقت بالا پیش‌بینی کند؛ از جمله تکمیل مجدد لایه‌ها (مانند تغییر لایه‌های تولیدی چاه ۱۱۴ از لایه‌های با نرخ برش آب بالا A6 و A7 به لایه عمیق‌تر M2) و نیز تغییر نرخ تولید (مانند چاه‌های ۲۴۱ و ۱۲۶) و در نتیجه، تغییرات برش آب ناشی از این مداخلات را

به‌درستی بازتولید نمود. این نتایج مؤید آن است که مدل نه‌تنها الگوهای مربوط به تاریخچه تولید را پیش بین کرده، بلکه روابط فیزیکی قابل‌تعمیم و حاکم بر پدیده تولید آب را به‌خوبی فراگرفته است.

سازگاری بالای عملکرد مدل در چاه‌های مختلف، همراه با توانایی بازتولید دقیق روندهای بلندمدت و نوسانات کوتاه‌مدت برش آب، نشان‌دهنده استحکام و قابلیت به‌کارگیری عملی بالای این مدل است. با توجه به اینکه پدیده نفوذ یا شکست آب[1] مهم‌ترین عامل محدودکننده عمر تولیدی چاه‌ها در این مخزن به شمار می‌رود، چارچوب پیشنهادی مبتنی بر LSTM ابزاری قابل‌اتکا برای تخمین پتانسیل تولید باقی‌مانده، اولویت‌بندی مداخلات مهندسی مخزن و درنهایت تصمیمات راهبردی در سطح میدان ارائه می‌دهد.

مدل پیشنهادی در این پژوهش، هر بخش از مخزن را به‌طور مستقل مورد تحلیل قرار داده و داده‌های فشار–با توجه به ثبات فشار مخزن–در آن لحاظ نشده است. مطالعات آتی می‌تواند شامل توسعه مدل‌های یکپارچه برای کل میدان، ادغام داده‌های برخط، یا تلفیق رویکردهای یادگیری عمیق با اصول شبیه‌سازی مخزن باشد. با این‌وجود، این پژوهش نشان می‌دهد که پیش‌بینی داده‌محور در مخازن بالغ با مکانیزم رانش آب، هنگامی که مدل بر پایه طراحی و آماده سازی سیستماتیک داده قرار گیرد، می‌تواند دقت پیش‌بینی و ارزش کاربردی بالایی فراهم سازد. اگرچه مدل LSTM استاندارد عملکرد قابل قبولی در تمام سکتورها داشت، اما خطا در سکتور ۱ اندکی بالاتر بود. برای کاهش خطا در چنین بخش هایی، روش‌های پیشرفته‌تری مانند مدل‌های ترکیبی CNN-LSTM، شبکه‌های Transformer یا Temporal Convolutional Networks (TCN) و همچنین تکنیک‌های افزایش داده می‌توانند مورد استفاده قرار گیرند. ارزیابی این روش‌ها در سکتورهای ناهمگن می‌تواند موضوع پژوهش‌های آتی باشد. علاوه بر این، اگرچه تمرکز این مطالعه بر پیش‌بینی برش آب بوده است، در مطالعات آتی می‌توان مدل‌های مشابه یادگیری عمیق را با خروجی مستقیم دبی آب تولیدی برای کاربردهای مدیریت تأسیسات سطح‌الارضی و سیستم‌های جمع‌آوری و دفع آب توسعه داد.

## مراجع


1. Ahmadi, R., Shahrabi, J., & Aminshahidy, B. (2019). Water cut/salt content forecasting in oil wells using a novel data-driven approach. *Oil & Gas Science and Technology – Revue d'IFP Energies Nouvelles*, *74*, 68. https://doi.org/10.2516/OGST/2019040
2. Ahmed, T. (2019). Reservoir Engineering Handbook-Tarek Ahmed-5th Edition. *Oxford: Elsevier Inc.*, *5*(4), 53.
3. Al-Mamoori, H. N., Tian, J., & Ma, H. (2025). Stuck Pipe Detection in Oil and Gas Drilling Operations Using Deep Learning Autoencoder for Anomaly Diagnosis. *Applied Sciences 2025, Vol. 15, Page 5042*, *15*(9), 5042. https://doi.org/10.3390/APP15095042
4. Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning Long-Term Dependencies with Gradient Descent is Difficult. *IEEE Transactions on Neural Networks*, *5*(2), 157–166. https://doi.org/10.1109/72.279181
5. Bilogur, A. (2018). Missingno: a missing data visualization suite. *Journal of Open Source Software*, *3*(22), 547. https://doi.org/10.21105/JOSS.00547
6. Chen, Y., Zhu, Y., Li, Y., Hu, D., Shi, S., Chen, Y., Li, Q., & Gu, F. (2022). Data-Driven Prediction Method of Water Cut Based on Random Forest Regression Model. *Society of Petroleum Engineers - ADIPEC 2022*. https://doi.org/10.2118/211408-MS


[1] Water Breakthrough

7. Das, A., Sudhanshu, R., Shukla, G., & Chaturvedi, N. D. (2025). Artificial intelligence-based approach for water-cut prediction in crude oil processing. *Engineering Applications of Artificial Intelligence*, *142*, 109892. https://doi.org/10.1016/J.ENGAPPAI.2024.109892
8. Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to Forget: Continual Prediction with LSTM. *Neural Computation*, *12*(10), 2451–2471. https://doi.org/10.1162/089976600300015015
9. Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, *18*(5–6), 602–610. https://doi.org/10.1016/J.NEUNET.2005.06.042
10. Hochreiter, S., Bengio, Y., Frasconi, P., & Schmidhuber, J. (n.d.). *Gradient Flow in Recurrent Nets: the Difficulty of Learning Long-Term Dependencies*.
11. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/NECO.1997.9.8.1735
12. Jin, M., & Emami-Meybodi, H. (2025). Deep-Learning Prediction of Flowing Bottomhole Pressure in Gas-Lifted Unconventional Wells. *SPE Annual Technical Conference and Exhibition*. https://doi.org/10.2118/227983-MS
13. Li, Z., Qian, Q., Guo, H., Wu, T., Cui, H., & Zhu, B. (2025). A Hybrid Framework for Production Prediction in High-Water-Cut Oil Wells: Decomposition-Feature Enhancement-Integration. *Processes 2025, Vol. 13, Page 1467*, *13*(5), 1467. https://doi.org/10.3390/PR13051467
14. Liashchynskyi, P., & Liashchynskyi, P. (2019). *Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS*. https://arxiv.org/pdf/1912.06059
15. Sheikhoushaghi, A., Gharaei, N. Y., & Nikoofard, A. (2022). Application of Rough Neural Network to forecast oil production rate of an oil field in a comparative study. *Journal of Petroleum Science and Engineering*, *209*, 109935. https://doi.org/10.1016/J.PETROL.2021.109935
16. Tripathi, B. K., Kumar, I., Kumar, S., & Singh, A. (2024). Deep Learning–Based Production Forecasting and Data Assimilation in Unconventional Reservoir. *SPE Journal*, *29*(10), 5189–5206. https://doi.org/10.2118/223074-PA
17. WANG, H., MU, L., SHI, F., & DOU, H. (2020). Production prediction at ultra-high water cut stage via Recurrent Neural Network. *Petroleum Exploration and Development*, *47*(5), 1084–1090. https://doi.org/10.1016/S1876-3804(20)60119-7
18. Zhang, K. ; ;, Wan, H. ;, Yao, W. ;, Zuo, Y. ;, Lin, J. ;, Liu, P. ;, Zhang, L. ;, Yang, Y. ;, Yao, J. ;, Federico, V. Di, Feng, G., Zhang, K., Wan, H., Yao, W., Zuo, Y., Lin, J., Liu, P., Zhang, L., Yang, Y., … Liu, C. (2024). Enhancing Oil–Water Flow Prediction in Heterogeneous Porous Media Using Machine Learning. *Water 2024, Vol. 16, Page 1411*, *16*(10), 1411. https://doi.org/10.3390/W16101411
19. کدخدایی, ع. (۲۰۲۲). تخمین تراوایی و شبیه‌سازی آن به منظور تعیین &., مرادی, م., رحیم پور بناب, ح ویژگی‌های مخزنی سازند شوریجه در یکی از مخازن شمال شرق ایران. *پژوهش نفت* *32*(1401–5), 3–18. https://doi.org/10.22078/PR.2022.4660.3094



## Using Deep Learning and Historical Production Data for Accurate Prediction of Water-Cut in Oil Wells: Application of LSTM

Accurate water-cut forecasting is essential in mature water-drive reservoirs, where excessive water production is the primary cause of well shutdowns and reduced field productivity. This study develops an advanced deep-learning framework using Long Short-Term Memory (LSTM) networks to predict water cut in a giant Iranian oil field characterized by strong aquifer support, heterogeneous lithology, and multi-layer completions. A high-resolution dataset comprising 33,697 water-cut measurements was comprehensively preprocessed through duplicate removal, outlier filtering, feature-variance analysis, technical consistency checks, smoothing, and missing-value imputation. After Monte Carlo sensitivity analysis, seventeen static and dynamic features—such as well coordinates, perforation geometry,

permeability, oil-rate history, and production time—were selected as model inputs. A time-consistent data split was applied, where 80% of early-time data were used for training and the remaining 20% for testing to avoid information leakage. The final LSTM architecture, optimized using grid-search hyperparameter tuning, was trained independently for four reservoir zones. The model achieved excellent performance, with Pearson correlation coefficients of 0.957–0.976 on training data and 0.889–0.920 on testing data, while Absolute Relative Error distributions remained tightly centered near zero. The LSTM successfully reproduced long-term production trends, short-term fluctuations, and operational interventions such as recompletions and flow-rate modifications. These results demonstrate that the proposed LSTM-based framework provides a robust and scalable tool for water-cut prediction and supports informed reservoir-management decisions in large heterogeneous water-drive fields

Keywords: Deep Learning–Based Water-Cut Forecasting, LSTM Production Time-Series Modeling, Water-Drive Reservoir Performance Analysis, Data-Driven Well Productivity Prediction, Intelligent Reservoir Engineering Models