دوره هجدهم، شماره بهار ۱۴۰۴

مجله فناوری اطلاعات در طراحی مهندسی

Information Technology in Engineering Design

http://sanad.iau.ir/journal/ited

تولید داده‌های مصنوعی برای پیش‌بینی خواص بتن: نوآوری‌ها و کاربردها در مهندسی محاسباتی

سید ایمان غفوریان حیدری[۱] مجید صافحیان*[۲] فرامرز مودی[۳] شبنم شادرو[۴]

(۱) گروه مهندسی عمران، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

(۲) گروه مهندسی عمران، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران*

(۳) گروه مهندسی عمران، دانشگاه صنعتی امیرکبیر، تهران، ایران

(۴) گروه مهندسی کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران

(تاریخ دریافت: ۱۴۰۳/۰۹/۲۹ تاریخ پذیرش:۱۴۰۳/۱۲/۱۷)

چکیده

چالش‌های موجود در دسترسی به داده‌های آزمایشگاهی دقیق، استفاده از داده‌های مصنوعی را به‌عنوان راه‌حلی کارآمد برای بهبود پیش‌بینی مقاومت فشاری بلندمدت بتن تحت تأثیر مواد افزودنی شیمیایی مطرح کرده است. در این پژوهش، روش‌های پیشرفته برای تولید داده‌های مصنوعی و تحلیل آن‌ها به‌منظور پیش‌بینی مقاومت فشاری بتن بررسی شده است. داده‌های مصنوعی در سطوح مختلف تولید و ارزیابی شدند تا تعداد بهینه برای دستیابی به بهترین نتایج شناسایی شود. نتایج نشان داد که تولید داده‌های مصنوعی می‌تواند دقت مدل‌های پیش‌بینی را به‌طور چشمگیری افزایش دهد و همچنین نیاز به انجام آزمایش‌های فیزیکی را کاهش دهد.در این مطالعه از ترکیب داده‌های آزمایشگاهی و داده‌های مصنوعی استفاده شد و تأثیر مثبت آن بر بهبود عملکرد مدل‌های پیش‌بینی مورد تأیید قرار گرفت. روش پیشنهادی در این پژوهش امکان تولید داده‌های مصنوعی مرتبط با شرایط واقعی را فراهم کرده و باعث صرفه‌جویی در منابع، کاهش هزینه‌ها و کاهش تأثیرات زیست‌محیطی شده است. یافته‌ها همچنین نشان دادند که مقاومت فشاری بتن طی دوره سه‌ساله به‌طور قابل توجهی کاهش می‌یابد که می‌تواند عمر مفید سازه‌ها را کاهش داده و هزینه‌های تعمیرات را افزایش دهد.

کلمات کلیدی: تولید داده‌های مصنوعی، یادگیری ماشین، مقاومت فشاری بتن، K-Means، TGAN

*عهده‌دار مکاتبات:

مجید صافحیان

نشانی: گروه مهندسی عمران، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

پست الکترونیکی: safehian@srbiau.ac.ir

۱- مقدمه

بتن به عنوان یکی از مهم‌ترین مصالح ساختمانی، نقش کلیدی در سازه‌های مهندسی ایفا می‌کند .مقاومت فشاری بتن به عنوان یکی از ویژگی‌های حیاتی در پروژه‌های عمرانی، به طور مستقیم بر دوام، پایداری و ایمنی سازه‌ها تأثیرگذار است .مقاومت فشاری طولانی‌مدت بتن اهمیت ویژه‌ای در عملکرد بلندمدت ساختمان‌ها دارد، زیرا با افزایش عمر سازه، پایداری آن در برابر تغییرات محیطی و بارهای وارده اهمیت بیشتری پیدا می‌کند [1] [2]. با این حال، اکثر مطالعات قبلی به تأثیر کوتاه‌مدت مواد افزودنی پرداخته‌اند و پژوهش‌های کمتری به بررسی اثرات بلندمدت این مواد بر مقاومت فشاری بتن پرداخته‌اند [3]. یکی از چالش‌های مهم در این زمینه، پیش‌بینی دقیق مقاومت فشاری بتن در سنین بلندمدت است. کاهش مقاومت بتن در بلندمدت می‌تواند منجر به افزایش نیاز به تقویت ابنیه و نگهداری مکرر سازه‌ها شود، که این مسئله به افزایش هزینه‌های نگهداری و مصرف انرژی می‌انجامد. پژوهش‌های اخیر نشان داده‌اند که تکنیک‌های یادگیری ماشین می‌توانند به بهبود پیش‌بینی مقاومت فشاری بتن کمک کنند. با توجه به اینکه داده‌های آزمایشگاهی برای پیش‌بینی مقاومت بتن در سنین بالا اغلب محدود هستند، استفاده از داده‌های مصنوعی برای پر کردن این شکاف پیشنهاد شده است.

روش‌های GAN، DCGAN، WGAN-GP بیشترین استفاده را در تولید داده‌های مصنوعی به دلیل دقت و توانایی بالا در یادگیری توزیع داده‌های پیچیده دارند، در حالی که TGAN و CTGAN در تولید داده‌های جدولی مورد استفاده قرار می‌گیرند. روش‌های augmentation Data کاربرد کمتری دارند و بیشتر برای پروژه‌های ساده‌تر استفاده می‌شوند. تعدادی از مطالعات صورت گرفته در جدول ۱ ارایه شده اند. در این جدول نشان داده شده که، استفاده از تکنیک‌های تولید داده‌های مصنوعی به ویژه در پیش‌بینی مقاومت فشاری بتن، نتایج قابل توجهی به همراه داشته است. یکی از این تکنیک‌ها، الگوریتم CTGAN است که توانست با تولید ۳۵۰۸ داده مصنوعی از ۱۱۱۱ داده واقعی، دقت مدل‌های SVM و ANN را بهبود بخشد [4]. این نتایج نشان‌دهنده اهمیت استفاده از داده‌های مصنوعی در بهبود عملکرد مدل‌های یادگیری ماشین است.

با استفاده از این داده‌های تقویت شده توانست دقت بالایی را نشان دهد [5]. این مطالعه نیز به خوبی بیانگر نقش داده‌های مصنوعی در افزایش دقت مدل‌ها و امکان تعمیم‌پذیری بهتر آن‌ها است. استفاده از الگوریتم TGAN نیز توانست با تولید ۱۰۰۰ داده مصنوعی، عملکرد مدل‌های مختلف مانند CFNN و SVM را بهبود دهد [6]. این داده‌های تقویت شده نتایج بهتری در پیش‌بینی مقاومت فشاری ارائه دادند. در مطالعه‌ای دیگر، داده‌های مصنوعی تولید شده توسط رایانه باعث افزایش تعداد نمونه‌ها از ۱۰۳۰ به ۵۰،۲۰۸ شد. این داده‌ها برای پیش‌بینی مقاومت فشاری مورد

استفاده قرار گرفتند و مدل ANN توانست عملکرد قابل توجهی در این زمینه داشته باشد [7]. استفاده از GAN، DCGAN و WGAN-GP برای تولید داده‌های مصنوعی در دو مطالعه دیگر، داده‌های محدودی را به ۱۰،۰۰۰ و ۵۰۰ داده مصنوعی افزایش داد. این داده‌ها به کار گرفته شدند و مدل‌های پیشرفته مانند VGG و D CNN 1 بهبود قابل توجهی در پیش‌بینی مقاومت فشاری نشان دادند [9] [8]. همچنین، در مطالعه‌ای دیگر، تقویت داده‌های مصنوعی برای بتن FR-MCC (ترکیبی از خاکستر بادی و پوسته برنج) انجام شد که ۱۰۰۰ داده مصنوعی از ۳۲۵ داده واقعی تولید شد و مدل‌های GEP و MEP برای پیش‌بینی مقاومت فشاری عملکرد بهتری از خود نشان دادند [10]. در نهایت، استفاده از روش کوپولا گوسی (GCM) برای تقویت داده‌ها نشان داد که مدل‌های رگرسیون خطی با استفاده از داده‌های مصنوعی توانستند دقت خود را برای پیش‌بینی مقاومت فشاری افزایش دهند [11]. مطالعه نشان می‌دهد که تکنیک TGAN می‌تواند با تولید نقاط داده احتمالی متعدد به مشکل عدم وجود مجموعه داده‌های آزمایشی در مسائل مهندسی پاسخ دهد. بررسی چدول ۱ نشان می دهد ، بیشتر مطالعات به سنین کوتاه‌مدت و میان‌مدت بتن تمرکز داشته‌اند و به بررسی مقاومت فشاری در سنین بلندمدت بتن پرداخته نشده است. این محدودیت نشان می‌دهد که تحقیقات بیشتری برای ارزیابی عملکرد مدل‌ها در سنین بلندمدت مورد نیاز است. در این مطالعات، ترکیب مواد افزودنی شیمیایی مختلف در بتن به صورت گسترده بررسی نشده است. با توجه به تأثیر قابل توجه مواد افزودنی بر خواص بتن، به ویژه در ترکیب با داده‌های مصنوعی، تحقیقات جامع‌تری در این زمینه ضروری به نظر می‌رسد تا تأثیر این مواد بر مقاومت فشاری بتن به‌طور دقیق‌تر بررسی شود.

روش‌های تولید داده‌های مصنوعی مانند GAN، CTGAN، TGAN و روش کوپولا گوسی (GCM) استفاده شده است و روش مانند K-Means که در تحقیقات ارایه شده در جدول ۱ به کار گرفته نشده است. این مسئله نشان می‌دهد که تمرکز بیشتر بر تکنیک‌های پیشرفته مولد داده‌ها بوده است که به‌طور خاص برای تقویت داده‌های جدولی به کار رفته‌اند.

مطالعات قبلی شواهدی ارائه می‌دهند که روش TGANs ابزاری قدرتمند برای تولید داده‌های جدولی، به‌ویژه در مورد پیش بینی مقاومت بتن می باشد. بنابراین دراین پژوهش، از روشهای TGAN و K-means برای تولید ۳۲،۱۵۵ داده مصنوعی از ۷،۸۴۵ داده واقعی استفاده شده است. تمرکز این مطالعه بر ترکیب مواد شیمیایی افزودنی بوده که در سایر مطالعات کمتر مورد بررسی قرار گرفته است. همچنین، از داده‌های واقعی بیشتری نسبت به سایر مطالعات استفاده شده ، که دقت پیش‌بینی‌ها را افزایش داه است.در مقایسه با سایر تحقیقات،در این مطالعه از مقایسه روش‌های TGAN و K-means برای توزیع دقیق‌تر داده‌های مصنوعی بهره برده شده است، در حالی که در دیگر مطالعات بیشتر از

روش‌های CTGAN و GAN استفاده کرده‌اند. مدل‌های پیشرفته‌ای یادگیری ماشین مانند NARX و RF نیز دراین تحقیق به کار گرفته شدند، که دقت مناسب تری نسبت به مدل‌های ساده‌تری مانند SVM و ANN داشتند.

یکی از مزایای کلیدی این مطالعه، بررسی بلندمدت مقاومت فشاری بتن تا سن سه سال است، در حالی که سایر مطالعات بیشتر بر بازه‌های زمانی کوتاه‌مدت و میان‌مدت تمرکز داشتند. این تفاوت‌ها، مطالعه ما را از دیگر تحقیقات متمایز می‌کند و باعث بهبود پیش‌بینی مقاومت فشاری بتن شده است.

جدول ۱: اطلاعات تولید داده مصنوعی و مدل‌های یادگیری ماشین

No.	شماره مرجع.	تعداد داده واقعی	تعداد داده مصنوعی	روش تولید داده مصنوعی	نوع بتن	سن بتن (روز)	مدل یادگیری ماشین	سال انتشار
1	[4]	1111	3508	CTGAN algorithm	FRP-confined concrete	28	SVM, ANN	2024
2	[5]	75	675	Data augmentation	Portland Cement Concrete	28, 56, 90	DNN	2024
3	[6]	930	1000	TGAN algorithm	Geopolymer concrete	3 days to 56	CFNN, SVM, LightGBM	2024
4	[7]	1030	50208	Computer-generated	General concrete	3 days to 365	ANN	2023
5	[8]	344	10000	GAN, DCGAN, WGAN-GP	Concrete with industrial wastes	28	VGG, 1D CNN, RF, SVR	2022
6	[9]	84	500	GAN, DCGAN, WGAN-GP	Concrete with industrial wastes	3 days to 365	VGG, 1D CNN, RF, SVR	2022
8	[10]	325	1000	Dataset augmented	FR-MCC (Fly Ash, Rice Husk Ash)	28 days to 365	GEP, MEP	2024
9	[11]	815	6513	TGAN algorithm	Geopolymer concrete	3 days to 56	CFNN, SVM, LightGBM	2024
10	Study	7845	32155	TGAN algorithm and K-means	Combined Chemical Admixtures	3 days to 3 years	NARX, RBF, RF, DT	2023

۲- روش‌شناسی

این بخش الگوریتم‌های مورد استفاده در فرآیند مدل‌سازی برای پیش‌بینی مقاومت فشاری طولانی‌مدت بتن که شامل ترکیب مواد افزودنی شیمیایی می‌شود را توضیح می‌دهد. شکل ۱، نمودار جریانی که جزئیات الگوریتم‌های استفاده شده را نشان می‌دهد، ارائه شده است. انجام مدل‌های مختلف، از جمله خود رگرسیون غیرخطی با ورودی‌های بیرونی (NARX)، تابع پایه شعاعی (RBF)، درخت تصمیم‌گیری (DT)، و جنگل تصادفی (RF)، در این نمودار جریان مقایسه شده‌اند.

داده‌های آزمایشی از نتایج آزمایشگاهی در مرحله اول جمع‌آوری شدند. این داده‌ها شامل مشخصات مواد، شرایط آزمون، و نتایج آزمون بودند. سپس داده‌ها پیش‌پردازش شدند، که شامل پالایش، نرمال‌سازی و تبدیل دسته‌بندی‌ها به قالب‌های قابل استفاده برای مدل‌های یادگیری ماشین بود. مجموعه آموزشی، مجموعه اعتبارسنجی، و مجموعه آزمایشی، با نسبت ۷۰٪ – ۱۵٪ – ۱۵٪ [12] تقسیم گردید. در بخش اول براساس تعداد داده های واقعی برای مجموعه آموزشی، ازروش "اعتبارسنجی متقاطع" K-fold که باعث می‌شود دقت مدل‌ها در پیش‌بینی داده‌های بهبود یابد و خطر بیش‌برازش کاهش یابد استفاده شده است. سپس، مجموعه داده‌های واقعی به عنوان یک ورودی برای آموزش مدل‌های برای پیش‌بینی مقاومت فشاری بتن استفاده شد. آموزش مدل‌ها بر اساس ویژگی‌های استخراج شده از مجموعه داده آموزشی انجام گرفت.

در بخش دوم هدف تولید داده های مصنوعی از طریق فرآیندهایی شامل استفاده از شبکه مولد تخاصمی جدول بندی شده (TGAN) و الگوریتم خوشه‌بندی K-means به مجموعه داده‌های مصنوعی تبدیل می‌شود. TGAN به عنوان یک مدل یادگیری عمیق برای تولید داده‌های مصنوعی استفاده می‌شود، در جدول ۱، روش‌های مختلف تولید داده‌های مصنوعی از جمله GAN، CTGAN، WGAN-GP و TGAN بررسی شده‌اند. مطالعات نشان داده‌اند که روش‌های GAN برای داده‌های تصویری مناسب‌ترند، در حالی که TGAN به دلیل بهره‌گیری از LSTM، توزیع داده‌های جدولی مانند بتن را بهتر حفظ می‌کند.

دلیل انتخاب TGAN، توانایی آن در مدل‌سازی روابط پیچیده و حفظ ویژگی‌های کلیدی بتن است. برخلاف روش‌های رایج، TGAN وابستگی‌های زمانی و متغیرهای ترکیبی را بهتر درک می‌کند که در تحلیل مقاومت بتن بسیار مهم است.

برای ارزیابی دقت داده‌های مصنوعی، از مقایسه شاخص‌های آماری، آزمون Kolmogorov-Smirnov، و اعتبارسنجی متقاطع K-Fold استفاده شد که نشان داد داده‌های تولید شده با TGAN شباهت بالایی با داده‌های آزمایشگاهی دارند.

در حالی که K-means برای دسته‌بندی داده‌ها بر اساس ویژگی‌های مشترک آن‌ها به کار می‌رود. سپس، مجموعه داده‌های مصنوعی تولید شده به عنوان یک ورودی جدید برای آموزش مدل‌های مورد نظر مورد استفاده قرار می‌گیرد. این فرآیند به منظور ایجاد تنوع در داده‌ها و ارزیابی تأثیر داده‌های مصنوعی بر عملکرد مدل‌ها انجام می‌شود. تعداده داده های مصنوعی در سه سطح (۱،۰۰۰، ۵،۰۰۰، و۱۰،۰۰۰) براساس روند مطالعات قبلی اشاره شده در جدول ۱ و سایر مطالعات قبلی [13] انتخاب شد، تا تعداد داده‌ها برای هر سن مورد نظر مقاومت فشاری بتن تعیین شود. تعداد بهینه داده‌های مصنوعی با ارزیابی پارامترهای اعتبارسنجی مانند ضریب تعیین (R^2) تعیین شد. پس از تجزیه و تحلیل نتایج، تعداد داده مناسب برای هر مدل بر اساس این ارزیابی‌ها انتخاب گردید. مجموعه آموزشی، مجموعه اعتبارسنجی، و مجموعه آزمایشی، با نسبت ۷۰٪ – ۱۵٪ – ۱۵٪ [12] تقسیم گردید. به منظوراخذ نتایج مطلوب ازالگوریتم خوشه‌بندی K-means در تولید دادهای مصنوعی، برای مجموعه های آموزشی ، ازروش "اعتبارسنجی متقاطع" K-fold نیز استفاده شد.

برای پیش‌بینی‌های دقیق‌تر در مورد تغییرات زمانی، اثرات مواد افزودنی شیمیایی، و کارایی و اعتبارسنجی مدل، مقاومت های ۳ روزه، ۷ روزه، ۲۸ روزه، ۹۰ روزه و یک‌ساله به عنوان ورودی‌ها استفاده شدند، در حالی که مقاومت فشاری سه‌ساله به عنوان خروجی بود. ارزیابی تمام مدل‌ها با استفاده از معیارهای عملکرد معین شده انجام شد. اگر خطاهای قابل توجهی شناسایی شدند، فرایند آموزش تکرار شد تا مدل بهینه‌سازی شود. این روش امکان تجزیه و تحلیل دقیق‌تر تغییرات زمانی، اثرات مواد افزودنی شیمیایی، کارایی مدل‌های پیش‌بینی، مقایسه نتایج مختلف مدل‌ها و عوامل مؤثر بر مقاومت فشاری بتن را فراهم آورد. داده‌های مورد استفاده در این مطالعه از مشخصات مواد منتخب و طرح‌های مخلوط نهایی که در چندین پروژه ساختمانی در ایران به کار رفته‌اند، مشتق شده‌اند. جدول ۲ خلاصه‌ای از ویژگی‌های مواد مرتبط را ارائه می‌دهد.

۳- مشخصات مواد

مواد مورد استفاده در بتن شامل سنگدانه‌ها (شن و ماسه) [14] ، سیمان، آب، و مواد افزودنی شیمیایی است. سنگدانه‌ها، با حداکثر اندازه ذرات ۲۰ میلی‌متر (۴/۳ اینچ)، مطابق با الزامات درجه‌بندی اندازه سنگدانه درشت شماره ۶

هستند که در ASTM C33 مشخص شده است [15]. در طراحی مخلوط نهایی، سنگدانه درشت مورد استفاده شامل شن خرد شده (۵–۲۰ میلی‌متر) با حداکثر اندازه ذره ۲۰ میلی‌متر (۳/۴ اینچ) است که مطابق با درجه‌بندی سنگدانه درشت شماره ۶ ذکر شده در مشخصات ASTM C33 است. مخلوط بتن نهایی از ترکیب ۱:۱ شن طبیعی به عنوان سنگدانه ریز استفاده می‌کند. آب مورد استفاده در تولید و شکل‌گیری بتن قابل شرب بوده و فاقد هرگونه مواد مضر است، که آن را برای آماده‌سازی بتن مناسب می‌سازد. سیمان اصلی مورد استفاده، سیمان نوع II از کارخانه سیمان تهران است. بر اساس استانداردهای ایران، خواص شیمیایی و فیزیکی این سیمان ارائه شده است. به دلیل شرایط خاص پروژه‌های ساختمانی، مواد افزودنی شیمیایی در طراحی مخلوط بتن استفاده شده‌اند [16]. نمونه‌های بتن مطابق با استانداردهای ASTM C31 عمل آوری شده‌اند [17]. در این پژوهش، ۱۲ طرح اختلاط بتن مورد بررسی قرار گرفت، اما به دلیل مسائل اجرایی و علمی، تنها ۶ طرح نهایی برای تحلیل انتخاب شدند. طرح‌های MIX-7 تا MIX-12 به دلیل ناهمگنی در شرایط عمل‌آوری، نوسانات مقاومت فشاری و عدم انطباق با استاندارد ASTM C31 کنار گذاشته شدند. برخی از این طرح‌ها در شرایط رطوبتی و دمایی کنترل‌نشده عمل‌آوری شدند که منجر به کاهش واکنش‌های هیدراسیون و افت مقاومت فشاری شد. همچنین، عدم یکنواختی در نتایج آزمایشگاهی باعث افزایش خطای پیش‌بینی در مدل‌های یادگیری ماشین گردید. انتخاب طرح‌های نهایی بر اساس کیفیت و یکپارچگی داده‌ها بوده و تضمین می‌کند که مدل‌های پیش‌بینی با دقت بالاتری اجرا شوند. لیکن در تحقیقات آتی، پیشنهاد می‌شود تعداد طرح‌های اختلاط افزایش یابد تا طیف گسترده‌تری از ترکیبات مورد بررسی قرار گرفته و دقت مدل‌های یادگیری ماشین بیش از پیش بهبود یابد. جدول ۲، مشخصات طراحی مخلوط نهایی بتن را نشان می‌دهد.

۳-۱- جامعه آماری

جامعه آماری این تحقیق شامل ۸۵۱۲ نمونه آزمایشگاهی از پروژه‌های ساختمانی واقعی در ایران است که برای ارزیابی مقاومت فشاری بتن در سنین مختلف (۳ روزه تا ۳ ساله) مورد استفاده قرار گرفته‌اند. این داده‌ها از پروژه‌هایی که در مناطق مختلف ایران اجرا شده‌اند، جمع‌آوری شده‌اند. تنوع جغرافیایی این پروژه‌ها و استفاده از مصالح مختلف، باعث شده است که این مجموعه داده به عنوان یک جامعه آماری جامع و نماینده در نظر گرفته شود.

۳-۲- روش نمونه‌گیری

روش نمونه‌گیری در این تحقیق به‌صورت هدفمند انجام شده است. در این روش، نمونه‌هایی انتخاب شده‌اند که دارای کیفیت بالا، دقت و قابلیت اطمینان کافی برای تحلیل‌های یادگیری ماشین باشند. برای انتخاب نمونه‌ها، پروژه‌هایی در نظر گرفته شده‌اند که دارای داده‌های جامع و استاندارد از مراحل مختلف عمل‌آوری بتن بوده‌اند. همچنین، برای جلوگیری از بروز خطاهای نمونه‌گیری، داده‌ها به نحوی جمع‌آوری شده‌اند که نشان‌دهنده شرایط مختلف ساخت و ساز و استفاده از مصالح متنوع در پروژه‌های ساختمانی ایران باشند.

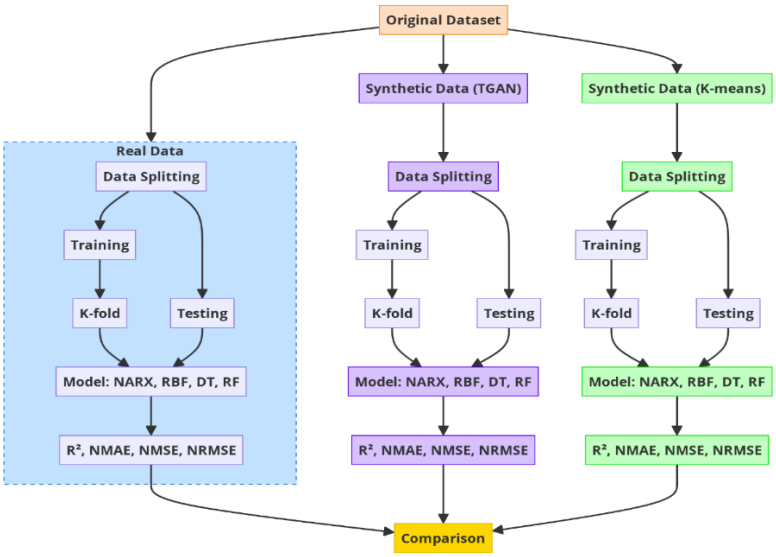
۳-۳- حجم نمونه و سن بتن

به‌طور کلی، این تحقیق شامل ۸۵۱۲ نمونه از پروژه‌های مختلف است. این نمونه‌ها از شش طرح مخلوط مختلف بتن تهیه شده‌اند که برای پروژه‌های اجرایی و آزمایشگاهی در ایران به کار رفته‌اند. داده‌های مربوط به سن بتن شامل نمونه‌هایی با عمر ۳ روزه، ۷ روزه، ۲۸ روزه، ۹۰ روزه، یک‌ساله و سه‌ساله است. برای دو طرح مخلوط خاص، داده‌های سه‌ساله در دسترس نبوده و فقط داده‌های مربوط به سنین پایین‌تر مورد استفاده قرار گرفته‌اند.

تمام نمونه‌های آزمایشگاهی مطابق با استاندارد ASTM C39 آزمایش شده‌اند. این استاندارد یکی از روش‌های معتبر بین‌المللی برای اندازه‌گیری مقاومت فشاری بتن است که دقت بالایی در سنجش رفتار مکانیکی بتن در طول زمان دارد. نمونه‌ها در شرایط کنترل‌شده عمل‌آوری شده و نتایج آزمایش‌های مختلف برای مدل‌سازی استفاده شده است.

جدول ۲: مشخصات طرح‌های اختلاط نهایی بتن

شماره طرح مخلوط	۳روزه	۷روزه	۲۸روزه	۹۰روزه	یک ساله	سه ساله
MIX1	۲۰۱	۱۹۴۹	۱۹۴۹	۱۴۶۵	۱۵۱	۱۵۰
MIX2	۵۰	۸۴	۴۵۰	۷۰	۷۵	۷۵
MIX3	۶۵	۸۵	۲۲۰	۸۰	۸۵	۹۰
MIX4	۵۵	۸۰	۲۰۰	۵۰	۶۰	۱۱۰
MIX5	۴۰	۵۵	۸۰	۸۵	۸۰	۵۰
MIX6	۴۰	۵۵	۸۰	۷۶	۸۰	۶۵



شکل ۱: روش انجام تحقیق

٤- توصیف داده‌ها

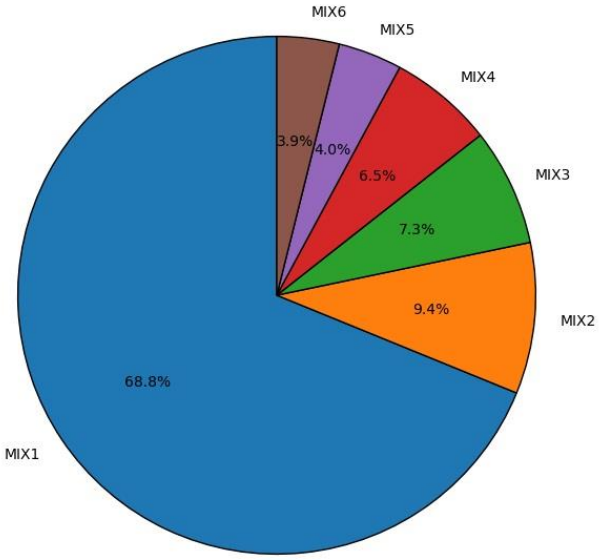
۴-۱- جمع‌آوری و پیش پردازش داده‌ها

ایجاد یک مجموعه داده جامع و قابل اعتماد برای توسعه مدل‌های پیش‌بینی یادگیری ماشین (ML) حیاتی است. نمونه‌های داده استفاده شده شامل متغیرهایی برای ساخت مدل‌ها به منظور پیش‌بینی مقاومت فشاری بتن در سنین مختلف است، بر اساس چهار طرح مخلوط و ۷،۸۴۵ نمونه که قصد استفاده آن‌ها در پروژه‌های عملی و آزمایشگاهی وجود دارد، این نمونه‌ها از پروژه‌های میدانی در ایران به دست آمده‌اند. درصد تعداد نمونه های بتن در طرح اختلاط های مختلف در شکل ۲ نمایش داده شده است.

۴-۲- پیش پردازش داده‌ها

در اولین گام، داده‌ها پالایش شدند تا ناهنجاری‌ها و داده‌های پرت، همچنین موارد دارای مقادیر گمشده حذف شوند [18] [19]. در مرحله بعدی، مشخص شد که تعداد داده‌ها در سنین مختلف بتن (۳ روز، ۷ روز، ۲۸ روز، ۹۰ روز، یک سال و سه سال) نامتعادل است. این نامتعادلی از تفاوت‌ها در تعداد آزمایش‌ها و اندازه‌گیری‌های انجام شده در هر دوره

ناشی می‌شود. به عنوان مثال، ممکن است آزمایش‌های بیشتری در دوره‌های کوتاه‌مدت (۳، ۷ و ۲۸ روز) به دلیل نیاز به نتایج سریع‌تر انجام شده باشد، در حالی که داده‌های کمتری برای دوره‌های بلندمدت (یک سال و سه سال) به دلیل زمان طولانی مورد نیاز برای به دست آوردن نتایج جمع‌آوری شده است. این نامتعادلی نیازمند تنظیم داده‌ها است تا از اختلال در یادگیری مدل جلوگیری شود، زیرا داده‌های آموزشی کافی برای برخی دوره‌ها وجود ندارد [11] .



شکل ۲: درصد تعداد نمونه‌های آزمایشگاهی بتن (واقعی) در طرح اختلاط‌های مختلف

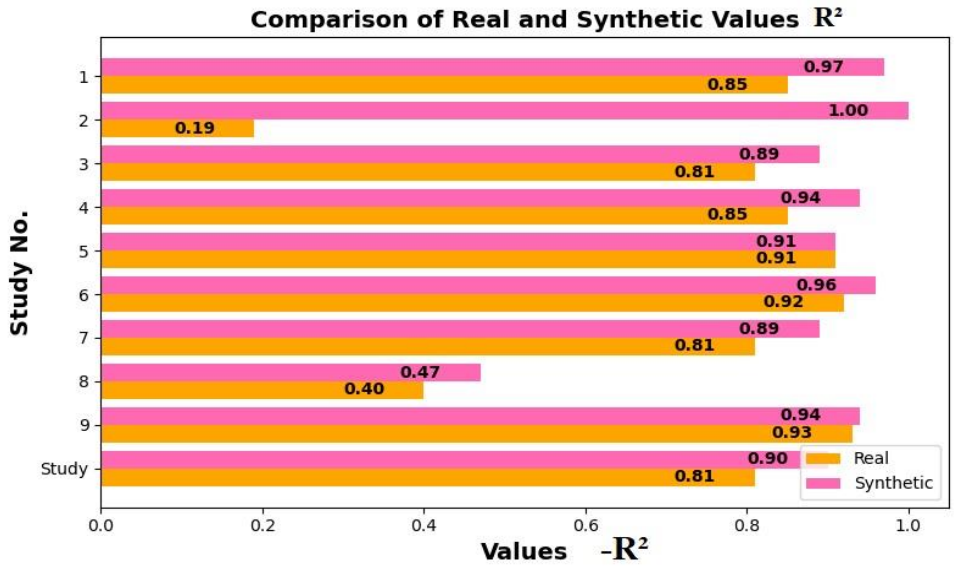
۴-۳- تولید داده‌های مصنوعی

در روش‌های سنتی برای تخمین مقاومت فشاری بتن با استفاده از تکنیک‌های یادگیری ماشین (ML)، مجموعه داده باید به زیرمجموعه‌هایی تقسیم شود که یکی برای آموزش و دیگری برای آزمایش استفاده می‌شود. بنابراین، صرف‌نظر از اندازه مجموعه داده، ۷۰-۸۰٪ از داده‌ها برای آموزش تخصیص می‌یابند، در حالی که تنها ۲۰-۳۰٪ برای اعتبارسنجی و آزمایش عملکرد مدل استفاده می‌شود [12]. با توجه به اندازه کوچک مجموعه داده‌ها برای پیش‌بینی مقاومت فشاری طولانی‌مدت بتن، بخش کوچکی از داده‌ها که برای آزمایش اختصاص می‌یابد، نگرانی‌هایی را در مورد دقت، مقاومت و قابلیت تعمیم مدل‌های توسعه‌یافته برای داده‌های جدید و نادیده ایجاد می‌کند [12][13]. مطالعات نشان دادند که در یک مدل پیش‌بینی مقاومت فشاری با استفاده از شبکه‌های عصبی مصنوعی (ANN)، افزایش تعداد نمونه‌های آزمایشگاهی از ۲۵ به ۱۲۵ درصد خطای پیش‌بینی را از ۲۰٪ به ۵٪ کاهش داده است. در شکل ۳ مقایسه تفاوت‌های

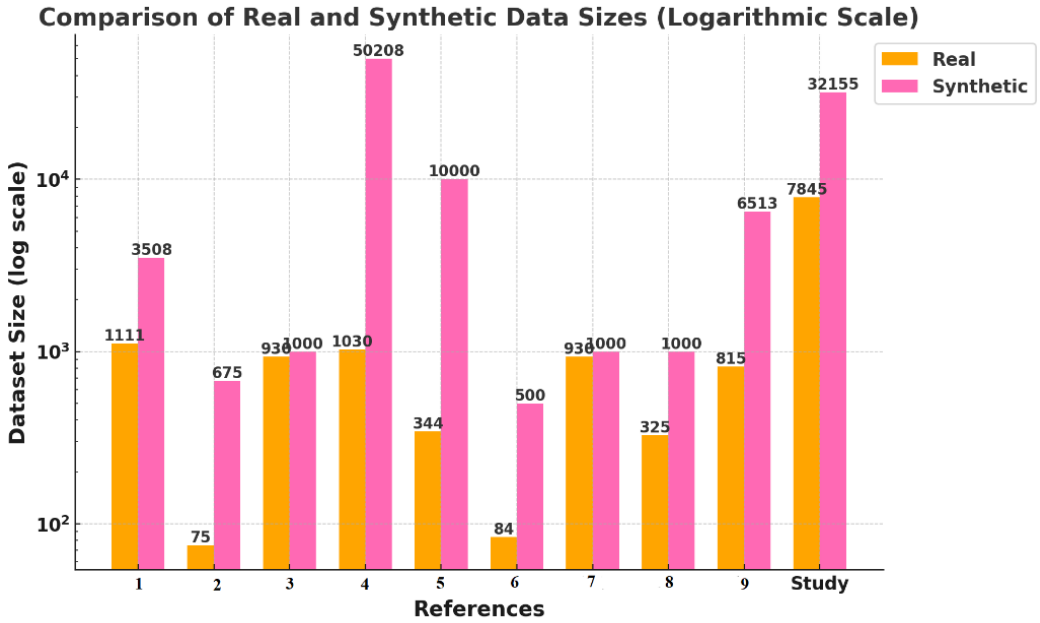
عملکرد صحت‌سنجی (R^2) در بین داده‌های واقعی و مصنوعی برای مجموعه داده‌های مطالعات مختلف براساس مطالعات ارایه شده در جدول ۱ ارایه شده است.تحقیق حاضر، تعداد نمونه‌های داده واقعی به میزان ۷،۸۴۵ می باشد، که این تعداد افزایش قابل توجهی نسبت به مطالعات ارایه شده در جدول ۱ را نشان می دهد. با این حال، جمع‌آوری داده‌های اضافی به دلیل پیچیدگی‌ها، هزینه‌ها و تأثیرات زیست‌محیطی تولید مقادیر زیادی از نمونه‌های بتن طولانی‌مدت، چالش‌برانگیز است و منجر به اتلاف مواد می‌شود. برای غلبه بر این موانع و جبران کمبود داده‌های آزمایشگاهی، استفاده از تکنیک‌های یادگیری ماشین مورد بررسی قرار گرفت. این تحلیل به بررسی استفاده از الگوریتم‌هایی مانند شبکه‌های مولد تخاصمی جدولی [20] (TGAN) و خوشه‌بندی K-means [21] برای تولید حجم زیادی از داده‌های مصنوعی پرداخت، که برای آموزش مدل‌های یادگیری ماشین استفاده شدند. این مدل‌ها برای پیش‌بینی مقاومت فشاری بتن به کار گرفته می‌شوند [20].

٤-۳-۱- تعداد مورد نیاز داده‌های مصنوعی

بر اساس موارد اشاره شده در جدول ۱، رابطه بین تعداد داده های واقعی و مصنوعی در مطالعه قبلی نشان داده شده است. با توجه به مجموعه داده‌های ای تحقیق که شامل ۷،۸۴۵ نمونه واقعی است، داده‌های مصنوعی در سه سطح مختلف کم (۱،۰۰۰)، متوسط (۵،۰۰۰)، بالا (۱۰،۰۰۰) تولید شدند. داده‌های مصنوعی با استفاده از روش‌های مختلف اعتبارسنجی مورد ارزیابی قرار گرفتند تا جامعیت و دقت آن‌ها تضمین شود. پس از آن، یک تحلیل دقیق و مقایسه عملکرد مدل، از جمله ضرایب تعیین (R^2) در سطوح مختلف داده‌ها، انجام شد تا سطحی با بالاترین دقت در تعداد دادها شناسایی شود. ۵،۰۰۰ نمونه داده مصنوعی برای هر سن بتن به عنوان تعداد بهینه برای تمام مدل‌های استفاده شد. این انتخاب نه تنها الگوریتم‌ها را بهینه کرد بلکه کارایی و دقت مدل‌ها را نیز افزایش داد. شکل ۴ نمودار لگاریتمی مقایسه‌ای بین تعداد داده‌های واقعی و مصنوعی را برای مجموعه داده‌های مطالعات مختلف را نشان می دهد.



شکل ۳: مقایسه تفاوت‌های عملکرد صحت‌سنجی (R^2) در بین داده‌های واقعی و مصنوعی برای مجموعه داده‌های مطالعات مختلف



شکل ۴: نمودار لگاریتمی مقایسه‌ای بین تعداد داده‌های واقعی و مصنوعی را برای مجموعه داده‌های مطالعات مختلف

۴-۳-۲- روش‌های تولید داده‌های مصنوعی

روش‌های مختلفی برای تولید داده مصنوعی برای نمونه بتن مطابق جدول ۳ ارایه شده است، لیکن در این مطالعه از دو روش برای تولید داده‌های مصنوعی خوشه‌بندی K-means [21] و روش شبکه‌های مولد تخاصمی جدولی (TGAN) [20] استفاده شد، که هر کدام به صورت جداگانه توضیح داده شده‌اند.

۴-۳-۲-۱- روش خوشه‌بندی K-means

برای مدیریت هم‌ترازی داده‌ها، خوشه‌بندی K-means با ۲۰ خوشه اعمال شد [21]. این تعداد خوشه با استفاده از روش Elbow و تحلیل Silhouette انتخاب شد تا اطمینان حاصل شود که تعداد خوشه‌ها نه تنها داده‌ها را به خوبی نمایندگی می‌کنند بلکه از پیچیدگی بیش از حد مدل نیز جلوگیری می‌کنند. در این تحقیق، استفاده نوآورانه از خوشه‌بندی و داده‌های مصنوعی برای بهبود دقت پیش‌بینی‌ها و جبران کمبود داده‌های واقعی به کار گرفته شد. این روش نه تنها به کاهش خطاهای مدل کمک کرد بلکه با افزودن ویژگی‌های خوشه‌بندی، دقت متعادل‌تری در پیش‌بینی‌های مربوط به مقاومت فشاری بتن ایجاد کرد.

برای مقابله با حساسیت K-means به مقادیر اولیه، تکنیک‌های مختلفی برای انتخاب مراکز اولیه به کار گرفته شد، مانند اجرای چندین بار الگوریتم با تنظیمات اولیه مختلف و انتخاب بهترین نتیجه بر اساس معیار ارزیابی مانند SSE (مجموع خطاهای مربعی) [22]. این رویکرد به یافتن مراکز بهینه برای خوشه‌بندی کمک کرد و احتمال همگرایی به حداقل‌های محلی را کاهش داد. داده‌های پرت موجود در خوشه‌هایی با داده‌های بسیار کم شناسایی و حذف شدند تا تأثیر منفی آن‌ها بر مدل‌سازی به حداقل برسد و تولید داده‌های مصنوعی با کیفیت بالا تضمین شود. استفاده از روش‌های خوشه‌بندی در تخمین مقاومت فشاری بتن توسط Beskopylny و همکاران مورد استفاده قرار گرفت [23]، جایی که خوشه‌بندی به بهبود دقت مدل‌های پیش‌بینی مقاومت فشاری منجر شد. این تکنیک، با تقسیم داده‌ها به گروه‌های همگن (از مقاومت پایین تا بسیار بالا) و افزودن ویژگی‌های خوشه‌بندی، می‌تواند مدل‌های یادگیری ماشین را بهینه کند و نتایج دقیق‌تری ارائه دهد. در مطالعه حاضر، تکنیک خوشه‌بندی برای جبران اثرات کمبود داده‌های آموزشی استفاده شد. در برخی خوشه‌ها، داده‌های مصنوعی تولید شدند تا مدل‌ها در ارزیابی عملکرد نتایج منطقی‌تری کسب کنند. این داد ها این داده‌های مصنوعی بر اساس مراکز خوشه‌های تعیین‌شده توسط خوشه‌بندی K-means در ۲۰ خوشه تولید شدند، با نوسانات کنترل‌شده است، ۱۰٪ در اطراف میانگین، تا اطمینان حاصل شود که ویژگی‌های واقعی داده‌های اصلی به

خوبی نمایندگی شده‌اند. برای جلوگیری از بیش‌برازش و تضمین کیفیت آموزش مدل، داده‌ها به ۱۰ بخش تقسیم شدند، و هر بخش برای آزمون مدل استفاده شد. این فرآیند به ارزیابی عملکرد مدل تحت شرایط داده‌های مختلف کمک می‌کند و تضمین می‌کند که مدل قادر به تعمیم به داده‌های جدید خارج از مجموعه آموزشی است. برای اطمینان از تطابق داده‌های مصنوعی تولید شده با توزیع اصلی، یک تحلیل آماری کامل انجام شد [24]. تکنیک‌های آماری مانند آزمون Kolmogorov-Smirnov برای ارزیابی تطابق توزیع داده‌های مصنوعی با داده‌های واقعی استفاده شد [22]. نتایج نشان داد که توزیع داده‌های مصنوعی با داده‌های واقعی تطابق داشت، به طوری که p-value ها بیشتر از سطح معنی‌داری ۰,۰۵ بودند و بنابراین فرض صفر رد نشد. روش شبیه‌سازی توزیع داده بر اساس خوشه‌بندی به عنوان یک رویکرد جدید برای تولید داده‌های مصنوعی برای مطالعات مواد ساختمانی استفاده شد و با استفاده از TGAN (Tabular GANs) معتبرسازی شد. شکل ۵، نمونه‌هایی از خوشه‌بندی داده‌های ورودی مقاومت فشاری بتن به ترتیب از طرح‌های مخلوط های ۱ تا ۴ در سن ۲۸ روزه را نشان می‌دهد.

۴-۳-۲-۲- روش شبکه‌های مولد تخاصمی جدولی (TGAN)

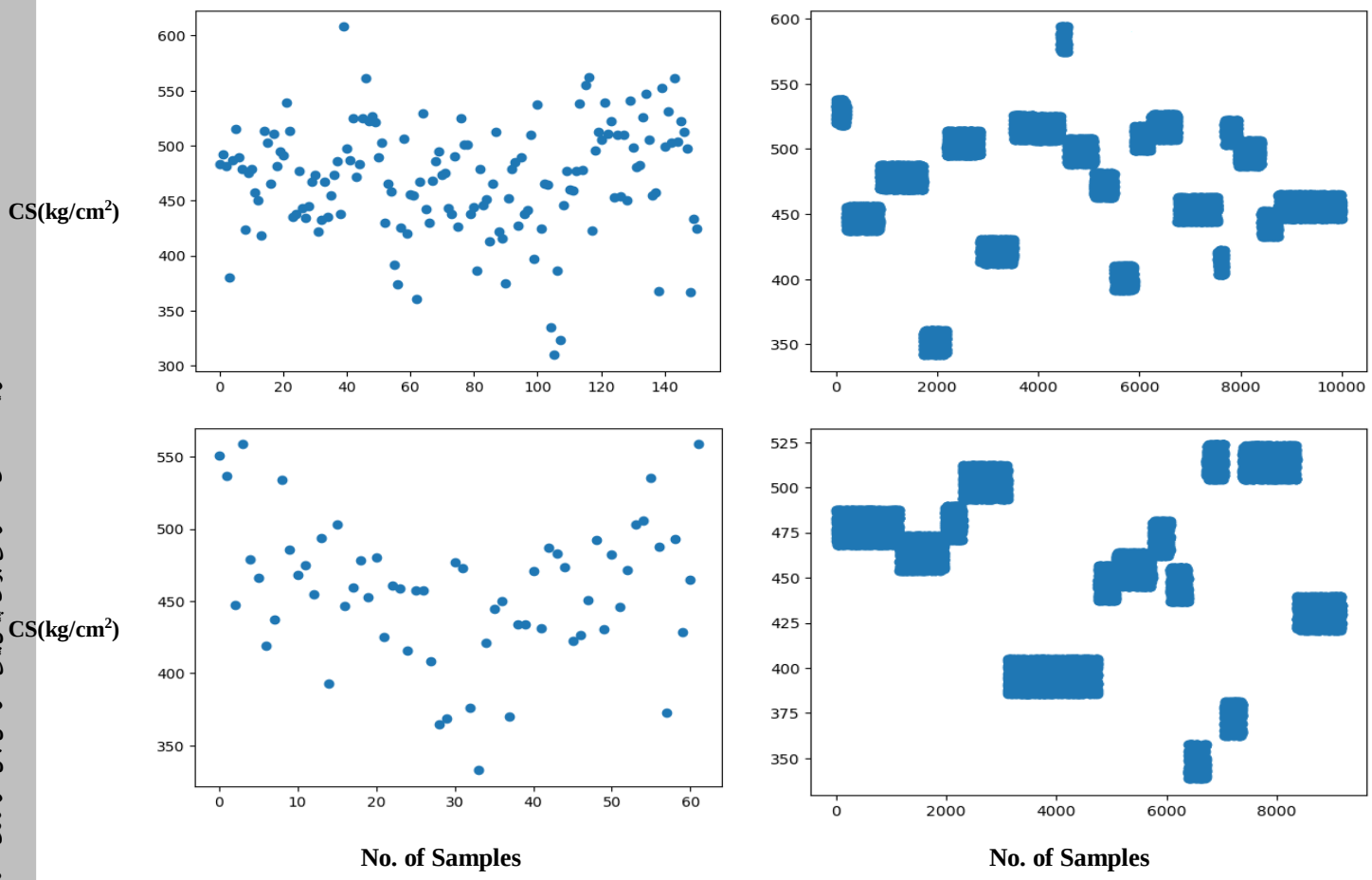
روش شبکه‌های مولد تخاصمی جدولی (TGAN) برای نخستین بار توسط Xu و همکارانش معرفی شد [29] این روش برای تولید داده‌های مصنوعی معتبر از جداولی با متغیرهای پیوسته و گسسته طراحی شده و نسبت به GAN های معمولی عملکرد بهتری دارد.

جدول ۳: اطلاع روش‌های تولید داده‌های مصنوعی برای تعیین مقاومت فشاری بتن

روش	مزایا	معایب	دقت	هزینه‌بر بودن	زمان‌بر بودن
شبکه‌های مولد تخاصمی (GAN)[8]	تولید داده‌های واقعی و دقیق برای ترکیبات پیچیده بتن	نیاز به داده‌های آموزشی زیاد و زمان طولانی برای آموزش	بسیار بالا	زیاد	زیاد
شبکه‌های مولد تخاصمی جدولی (TGAN)[11]	مناسب برای داده‌های عددی بتن (نسبت آب به سیمان، ترکیبات مصالح و غیره)	نیاز به تنظیم دقیق و زمان‌بر برای داده‌های جدولی بتن	بسیار بالا	متوسط	زیاد
کاوش داده (Data Augmentation)[5]	ساده و سریع، افزایش حجم داده‌ها با تغییرات کوچک در نسبت‌ها	ممکن است داده‌های تولیدی تنوع و دقت کافی نداشته باشند	متوسط	کم	کم

زیاد	زیاد	بالا	نیاز به محاسبات سنگین و زمان‌بر	در نظر گرفتن عدم قطعیت و تولید داده‌های متنوع برای سناریوهای مختلف بتن	شبیه‌سازی مونت‌کارلو[25]
زیاد	زیاد	بسیار بالا	زمان‌بر و پیچیده، نیاز به دانش آماری	توانایی مدل‌سازی عدم قطعیت در داده‌های بتن	مدل‌های تصمیم‌گیری بیزین[26]
متوسط	زیاد	بالا	نیاز به داده‌های زیاد برای آموزش	دقت بالا در تولید داده‌های پیچیده و غیرخطی مربوط به بتن	شبکه‌های عصبی مصنوعی (ANNs)[27]
کم	کم	متوسط	دقت پایین‌تر در پیش‌بینی‌های پیچیده بتن	مناسب برای داده‌های ساده و شروع سریع در مدل‌سازی داده‌های بتنی	یادگیری ماشین (Machine Learning)[28]
متوسط	زیاد	بالا	نیاز به منابع محاسباتی بیشتر، مناسب برای داده‌های تصویری و چندبعدی	توانایی تولید تصاویر و داده‌های پیچیده و چندبعدی برای بازرسی بتن	DCGAN (Deep Convolutional GAN)[8][9]
زیاد	زیاد	بسیار بالا	نیاز به منابع بیشتر و محاسبات سنگین‌تر	حل مشکل ناپایداری GANها و بهبود در کیفیت داده‌های تولیدی	WGAN-GP (Wasserstein GAN with Gradient Penalty)[8][9]

TGAN شامل یک مولد (Generator) برای ایجاد داده‌های مصنوعی و یک تمیزدهنده (Discriminator) برای ارزیابی واقعی بودن این داده‌ها است [24]. در این روش، شبکه‌های حافظه طولانی-کوتاه مدت (LSTM) به عنوان مولد و یک پرسپترون چندلایه (MLP) به عنوان تشخیص‌دهنده مورد استفاده قرار می‌گیرند [20]. فرآیند آموزش مدل TGAN با استفاده از بهینه‌ساز Adam و به‌کارگیری واگرایی Kullback–Leibler (KL) در تابع زیان انجام می‌شود تا دقت تولید داده‌های مصنوعی بهبود یابد . این مدل قادر است توزیع داده‌های یک جدول منفرد را با متغیرهای عددی و دسته‌ای تقلید کند و در علم مواد نسبت به روش‌های دیگر مانند autoencoders عملکرد بهتری دارد [31][30]. به دلیل استفاده از LSTM و MLP در ساختار مولد و تشخیص‌دهنده، مدل TGAN دارای چندین پارامتر و ابرپارامتر تنظیمی است که بر کیفیت داده‌های مصنوعی تأثیر می‌گذارند .مقدار بهینه این پارامترها از طریق آزمون و خطا تعیین شده است [31][30]. در این مطالعه، بهینه‌ساز Adam برای تنظیم پارامترها انتخاب شده است و تعداد سلول‌های RNN در مولد ۴۵۰ واحد، تعداد واحدهای کاملاً متصل ۸۵ واحد، نرخ یادگیری ۰,۰۰۱، و اندازه دسته داده‌ها ۲۵۰ واحد در نظر گرفته شده است. تشخیص‌دهنده دارای ۴ لایه با ۲۰۰ واحد در هر لایه است و مدل در ۲۵ دوره با ۷۰۰۰ گام در هر دوره آموزش داده شده است .این تنظیمات برای بهینه‌سازی عملکرد TGAN به کار گرفته شده‌اند.



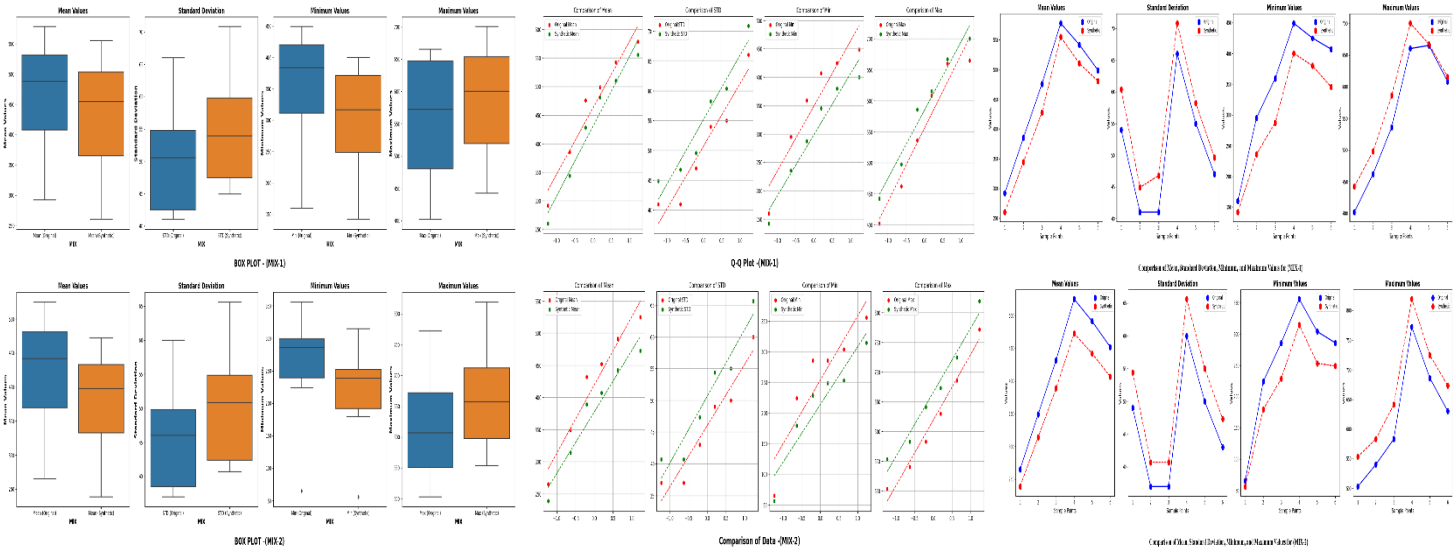
شکل۵. نمونه‌هایی از خوشه‌بندی داده‌های ورودی مقاومت فشاری بتن در طرح‌های مخلوط‌های مختلف (۲۸ روزه)

۴-۳-۳- بررسی وضعیت آماری داده‌های واقعی و مصنوعی

با استفاده از شکل ۶، مقایسه تحلیلی داده‌های واقعی و مصنوعی برای تمام طرح اختلاط‌ها صورت گرفته شده است . در شکل‌های ۷و۸ مقایسه ای دقیقی بین داده‌های واقعی و مصنوعی نشان داده شده است .

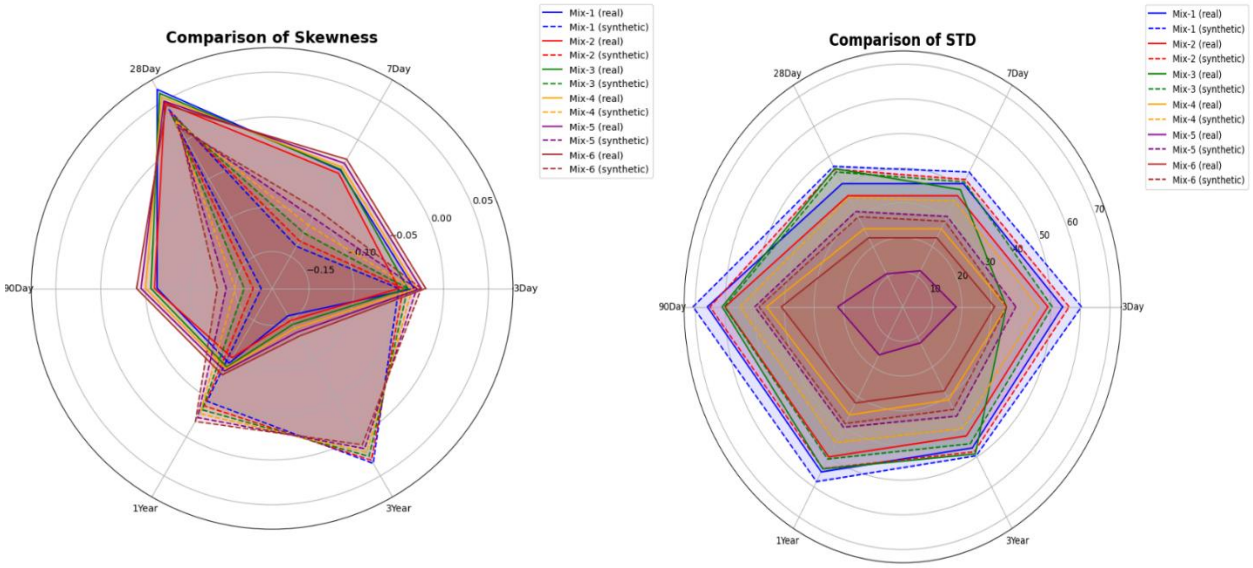
بررسی‌های انجام شده نشان می‌دهند که داده‌های مصنوعی به طور کلی توانسته‌اند مقادیر میانگین داده‌های واقعی را به خوبی شبیه‌سازی کنند، با انطباق مناسبی در بیشتر موارد و اختلاف‌های جزئی در برخی نقاط. از نظر انحراف معیار، داده‌های مصنوعی در بیشتر موارد به داده‌های واقعی نزدیک بوده، با تفاوت‌های کوچک در طرح‌هایی مانند MIX-2 در روز ۹۰ که بر دقت پیش‌بینی‌ها تأثیر می‌گذارد. همچنین، مقادیر حداقل و حداکثر داده‌های مصنوعی دامنه مقادیر واقعی را به خوبی پوشش داده‌اند، که این امر برای ایجاد مدل‌های پیش‌بینی کارآمد حیاتی است. در نهایت، چولگی داده‌های

مصنوعی عمدتاً با داده‌های واقعی هماهنگ است (شکل ۸)، با وجود تفاوت‌های جزئی که به ندرت ساختار توزیع داده‌ها را مختل می‌کند. این یافته‌ها دلالت بر این دارند که داده‌های مصنوعی می‌توانند به عنوان داده‌های پشتیبان موثری در پیش‌بینی‌های مدلی عمل کنند و برای استفاده در تحقیقات بعدی مناسب هستند.



شکل ۶: نمودار مقایسه تحلیلی داده‌های واقعی و مصنوعی برای تعدادی از طرح اختلاط‌ها (Box Plot، Quantile- Plot

(Standard Deviation، Quantile



شکل ۷: وضعیت داده‌های واقعی و مصنوعی برای داده‌های آماری

شکل ۸: تحلیل چولگی در داده‌های واقعی و مصنوعی

به منظور بررسی میزان تطابق داده‌های مصنوعی تولید شده توسط K-Means و TGAN با داده‌های تجربی، از KL Divergence و آزمون Kolmogorov-Smirnov (KS) به‌عنوان معیارهای ارزیابی استفاده شده است. مقدار KL Divergence میانگین برابر با ۰,۱۰۶۸ برای تمامی طرح‌های اختلاط بتن نشان می‌دهد که داده‌های مصنوعی به‌طور کلی رفتار توزیعی مشابهی با داده‌های واقعی دارند. مقدار پایین KL، به‌ویژه در Mix-3 (۰,۰۵۵۳) و Mix-4 (۰,۰۶۴۱)، حاکی از هم‌پوشانی بالای داده‌های مصنوعی و واقعی در این ترکیبات است. همچنین، میانگین مقدار KS D-statistic برابر با ۰,۰۷۳۷ و میانگین p-value برابر با ۰,۰۹۳۱ محاسبه شده است. مقادیر بالاتر p-value نشان‌دهنده عدم تفاوت معنادار بین داده‌های مصنوعی و واقعی در بیشتر ترکیبات بتن است. این نتایج نشان می‌دهد که K-Means و TGAN قادر به تولید داده‌هایی با توزیع آماری نزدیک به داده‌های واقعی بوده و می‌تواند در بهبود عملکرد مدل‌های یادگیری ماشین نقش مؤثری داشته باشد.

۴-۳-۴- تقسیم داده‌ها و اعتبارسنجی متقاطع K- برابر

پس از به‌دست‌آوردن مجموعه داده، آن به سه بخش مجموعه آموزشی، مجموعه اعتبارسنجی، مجموعه آزمایشی، با نسبت ۷۰٪ – ۱۵٪ – ۱۵٪ [12] تقسیم شد. برای کاهش خطاهای نمونه‌گیری تصادفی و اطمینان از نمایندگی مناسب داده‌ها، محققان از اعتبارسنجی متقاطع-K برابر استفاده کردند. در این مطالعه، از اعتبارسنجی متقاطع-K برابر استفاده شد که در آن مجموعه داده به K زیرمجموعه تقسیم شد [32] [12]. در هر تکرار، یک زیرمجموعه به عنوان مجموعه آزمایشی و زیرمجموعه‌های باقی‌مانده برای آموزش استفاده می‌شوند. این فرآیند K بار تکرار می‌شود تا هر زیرمجموعه یک بار به عنوان مجموعه آزمایشی و K-1 بار به عنوان مجموعه آموزشی استفاده شود. این روش دقت و قابلیت تعمیم مدل را ارزیابی می‌کند و از بیش‌برازش جلوگیری می‌کند [12]. مزایای استفاده از اعتبارسنجی متقاطع-K برابر شامل کاهش تعصب نمونه‌گیری تصادفی، بهبود تعمیم‌پذیری مدل، و ارائه ارزیابی قابل‌اعتمادتر از عملکرد مدل است [12] [33]. در این تحقیق، از اعتبارسنجی متقاطع ده‌برابر استفاده شد که در آن مجموعه داده به ده زیرمجموعه تقسیم می‌شود و هر زیرمجموعه یک بار به عنوان داده آزمایشی و نه بار به عنوان داده آموزشی استفاده می‌شود [33] [12] دقت

الگوریتم محاسبه و به‌عنوان میانگین دقت به‌دست‌آمده از ده مدل در ده دوره اعتبارسنجی گزارش می‌شود، که ارزیابی مدل را دقیق و قابل‌اعتماد کرده و عملکرد آن را در شرایط مختلف داده ارزیابی می‌کند [34].

۴-۴- توصیف روش‌های یادگیری ماشین انتخاب‌شده برای مطالعه

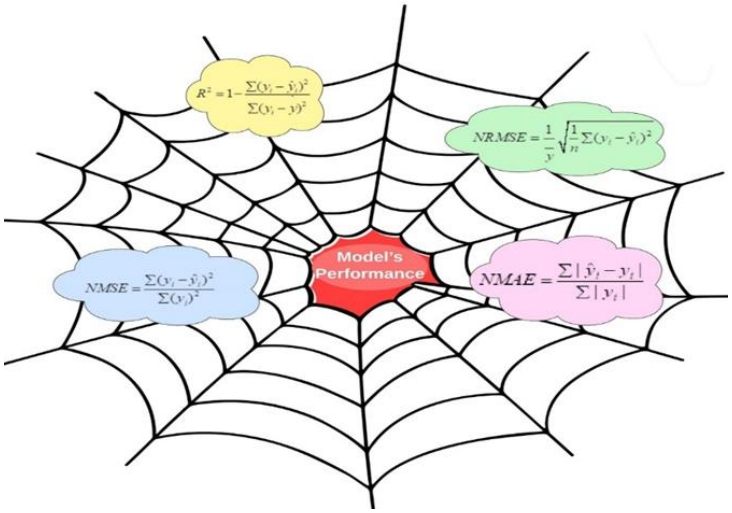
این بخش به توضیح الگوریتم‌های استفاده‌شده در مدل‌سازی برای پیش‌بینی مقاومت فشاری بتن بر اساس ترکیبات افزودنی‌ها می‌پردازد [35] برای پیاده‌سازی مدل‌ها از نرم‌افزارهای Matlab و Anaconda استفاده شد [37] [36]. الگوریتم‌های استفاده‌شده در این مطالعه عبارت‌اند از، خود رگرسیون غیرخطی با ورودی‌های بیرونی (NARX)، تابع پایه شعاعی (RBF)، درخت تصمیم‌گیری (DT) و جنگل تصادفی (RF). مزایا و معایب این الگوریتم‌ها در سناریوهای پیش‌بینی مقاومت فشاری در جدول ۴ خلاصه شده است.

جدول ۴: مقایسه مزایا و معایب مدل‌ها

مدل	مزایا	معایب
NARX	- مناسب برای مدل‌سازی سیستم‌های پویا با داده‌های زمانی - قابلیت مدل‌سازی روابط غیرخطی - مدیریت تعاملات پیچیده بین ورودی‌های خارجی و داده‌های گذشته - قابل پیاده‌سازی با شبکه‌های عصبی	-نیاز به داده‌های زیاد برای آموزش -تنظیم پارامترهای پیچیده -حساسیت به داده‌های گمشده یا نویز -پیچیدگی محاسباتی به ویژه در معماری‌های عمیق
RBF	-مناسب برای مدل‌سازی روابط غیرخطی پیچیده -سرعت بالای همگرایی نسبت به شبکه‌های عصبی سنتی - مناسب برای داده‌های خوشه‌بندی‌شده - قابلیت تعمیم‌پذیری خوب با داده‌های محدود	-حساس به نویز و داده‌های دورافتاده - انتخاب تابع پایه مناسب چالش‌برانگیز - افزایش ابعاد داده منجر به محاسبات پیچیده‌تر -نیاز به تنظیم پارامترهای متعدد
RF	-مقاوم به بیش‌برازش و نویز -عملکرد خوب با داده‌های پیچیده و ابعاد بالا - مدیریت مقادیر گمشده و داده‌های دسته‌بندی‌شده - قابلیت تفسیر با اهمیت ویژگی‌ها	-زمان آموزش طولانی برای داده‌های بزرگ -مصرف حافظه زیاد و محاسبات سنگین -حساسیت به تغییرات کوچک در داده‌ها (نوسانات) - پیچیدگی مدل و کندی در استنتاج زمان واقعی
DT	-ساده و قابل تفسیر -سرعت بالا در آموزش و پیش‌بینی - مناسب برای داده‌های دسته‌بندی‌شده و مداوم - عدم نیاز به مقیاس‌بندی ویژگی‌ها	- مستعد بیش‌برازش (بدون تنظیم مناسب) - حساس به داده‌های نویزی و دورافتاده - عملکرد ضعیف در داده‌های پیچیده و ابعاد بالا -تعمیم‌پذیری ضعیف بدون هرس درخت‌ها

۴-۵- روش‌های ارزیابی عملکرد

برای ارزیابی دقت مدل‌های استفاده شده، معیارهای به کار رفته شامل ضریب تعیین (R^2)، خطای میانگین مربعات نرمال شده (NMSE)، خطای مطلق میانگین نرمال شده (NMAE)، و خطای میانگین ریشه مربعات نرمال شده (NRMSE) هستند. معادلات مربوط به این معیارها در شکل ۹ نشان داده شده‌اند. استفاده از نسخه‌های نرمال شده این معیارها به تنظیم مقادیر خطا نسبت به مقیاس داده کمک می‌کند و مقایسه بین مدل‌ها را آسان‌تر و دقیق‌تر می‌سازد. در شکل ۹ ، n نشان‌دهنده تعداد کل نمونه‌ها، y_i مقدار واقعی، میانگین مقادیر واقعی و $\bar{y}$ مقدار میانگین مقادیر واقعی را نشان می‌دهد و $\hat{y}_i$مقدار پیش‌بینی‌شده می باشد.



شکل ۹: فرمول های نحوه ارزیابی عملکرد مدل‌ها

۵- نتایج و بحث

۵-۱- نتایج

در این بخش، ابتدا تعداد مورد نیاز داده‌های مصنوعی و کیفیت آماری داده‌های مصنوعی و واقعی و همچنین مدل‌های پیش‌بینی مقاومت فشاری براساس انواع داده‌ها بررسی شد. سپس، بر اساس نتایج معیارهای اعتبارسنجی، مدل مناسب انتخاب گردید.در ادامه تغییرات طولانی‌مدت مقاومت فشاری بتن تحت شرایط استفاده از ترکیبات شیمیایی مختلف مورد بررسی و تحلیل قرار گرفت.

۵-۱-۱- تعداد داده‌های مصنوعی

معیارهای ارزیابی R^2 برای تعداد داده‌های مصنوعی تولیدشده در تمام سطوح در شکل ۱۱ ارائه شده است. تمام مدل‌ها بهترین عملکرد خود را در سطح ۵،۰۰۰ داده نشان دادند. تحلیل‌ها نشان می‌دهند که با افزایش تعداد نقاط داده، دقت پیش‌بینی مدل‌ها در مقایسه با داده‌های اولیه آزمایشگاهی بهبود یافته است.

۵-۱-۲- مقایسه روش‌های K-means با TGAN

دراین تحقیق، هدف از بررسی و مقایسه دو روش پیشرفته برای تولید داده‌های مصنوعی خوشه‌بندی K-means و روش شبکه‌های مولد تخاصمی جدولی (TGAN)، ارزیابی قابلیت هر دو روش در تولید داده‌های دقیق و کارآمد برای مدل‌سازی پیش‌بینی مقاومت فشاری بتن بود. جهت بررسی معیارهای ارزیابی، در اولین گام تنوع داده‌های تولیدی از طریق شاخص‌های آماری مانند ضریب تغییرات (CV) مورد ارزیابی قرار گرفت، که برای هر دو روش نزدیک به مقادیر داده‌های واقعی بود، نشان دهنده توانایی آن‌ها در بازتاب خصوصیات داده‌های اصلی است. برای کنترل کیفیت بصری و آماری، کورتوزیس و کشیدگی داده‌های تولیدی به ترتیب نزدیک به ۳ و ۰,۵ محاسبه شد، که این مقادیر بسیار نزدیک به توزیع داده‌های واقعی است، نشان‌دهنده کیفیت بالای داده‌های تولیدی و توانایی هر دو روش در مدل‌سازی دقیق داده‌های بتن است. همچنین دقت پیش‌بینی‌های مدل‌ها براساس داده‌های تولید شده، با استفاده از معیارهای آماری ارزیابی شد. و ضریب تعیین R^2 برای هر دو مدل بیش از ۰,۹۵ بدست آمد که این نتیجه، نشان‌دهنده دقت بالا و توانایی مدل‌ها در پیش‌بینی مقاومت فشاری بتن با استفاده از داده‌های تولیدی است. این نتایج تأیید می‌کنند که هر دو روش، خوشه‌بندی K-means و TGAN، قابلیت‌های مشابهی در تولید داده‌های مصنوعی دارند و هر دو می‌توانند برای افزایش دقت مدل‌های پیش‌بینی در این تحقیق استفاده شوند. لیکن از آنجاییکه روش TGAN نیازمند منابع محاسباتی و پیچیدگی فنی و محاسباتی زیاد و همچنین نیاز به دقت بالا در تنظیم مدل است و نسبت به روش خوشه‌بندی K-means نسبتاً ساده و کم‌هزینه از نظر محاسباتی است. روش تولید داده‌های مصنوعی با استفاده از خوشه‌بندی K-means و تکثیر داده‌ها بر اساس مراکز خوشه‌ها، رویکردی نوآورانه و کار آمد و مقرون به صرفه است که می‌تواند به عنوان یک روش جدید در تولید داده برای تخمین مقاومت فشاری بلند مدت بتن مطرح شود.

۵-۱-۳- تنظیمات اعتبارسنجی متقاطع K برابر

معیارهای عملکرد R^2 برای مدل‌های پیشنهادی که با استفاده از روش اعتبارسنجی متقاطع-K برابر آموزش دیده‌اند، در شکل ۱۱ ارائه شده است.

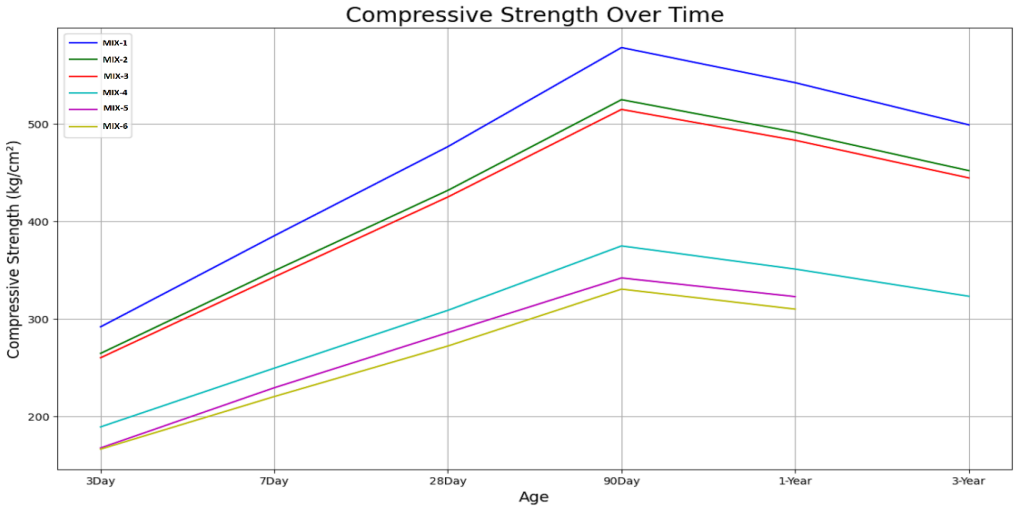
۵-۱-۴- مقاومت فشاری کوتاه‌مدت و بلندمدت بتن

تغییرات متوسط مقاومت فشاری برای شش بتن مختلف MIX-1 تا MIX-6 در دوره‌های زمانی ۳ روز تا ۳ سال مورد بررسی قرار گرفته است. همان‌طور که در شکل ۱۰ نشان داده شده است، تمامی طرح‌های اختلاط افزایش قابل‌توجهی در مقاومت فشاری تا سن ۹۰ روزگی داشته‌اند، که این افزایش ناشی از تکمیل فرآیند هیدراتاسیون سیمان و کاهش تخلخل بتن است.با این حال، پس از ۹۰ روز، برخی از طرح‌های مخلوط دچار افت مقاومت فشاری شده‌اند. در طرح‌های MIX-5 و MIX-6، افت مقاومت بیشتر از سایر طرح‌ها بوده است. نتایج این تحقیق نشان داد که عامل اصلی کاهش مقاومت درازمدت، ترکیبات شیمیایی مواد افزودنی بتن است.

افزودنی‌های شیمیایی که به منظور بهبود خواص مکانیکی بتن در کوتاه‌مدت استفاده می‌شوند، در بلندمدت ممکن است اثرات نامطلوبی بر دوام بتن داشته باشند. این افزودنی‌ها، اگرچه در ابتدا موجب بهبود کارایی، افزایش مقاومت و تسریع یا کنترل هیدراتاسیون می‌شوند، اما در درازمدت ممکن است بر ساختار میکروسکوپی بتن تأثیر گذاشته و منجر به کاهش پیوستگی بین ذرات سیمان و محصولات هیدراتاسیون شوند.بررسی نتایج نشان می‌دهد که در طرح‌هایی که مقدار مشخصی از مواد افزودنی به کار رفته، روند کاهش مقاومت فشاری در بازه‌های یک‌ساله و سه‌ساله محسوس‌تر بوده است. این کاهش می‌تواند ناشی از تغییرات در ساختار کریستالی سیمان هیدراته شده، افزایش تخلخل، تغییرات در توزیع محصولات هیدراتاسیون و واکنش‌های شیمیایی درون ماتریس بتن باشد.این نتایج نشان می‌دهد که ترکیبات شیمیایی افزودنی‌ها، اگرچه در کنترل اولیه مشخصات بتن نقش مهمی دارند، اما در طول زمان می‌توانند به تغییرات نامطلوبی در ساختار بتن منجر شوند که در نهایت باعث افت مقاومت فشاری بتن در سنین بالاتر می‌شود.

در مقایسه بین طرح‌های اختلاط، MIX-1 و MIX-2 به دلیل استفاده بهینه از افزودنی‌ها، عملکرد بهتری در بلندمدت نشان داده‌اند. در این طرح‌ها، مقدار و نوع افزودنی‌های شیمیایی به گونه‌ای انتخاب شده که اثرات منفی بر ساختار بتن در درازمدت کاهش یافته و مقاومت فشاری تا سه سال پایداری بیشتری داشته باشد. در مقابل، طرح‌های MIX-5 و MIX-6 به دلیل ترکیب متفاوت افزودنی‌های شیمیایی، کاهش بیشتری در مقاومت فشاری پس از یک سال

نشان داده‌اند، که نشان‌دهنده تأثیر نامطلوب برخی افزودنی‌ها بر استحکام درازمدت بتن است.با توجه به نتایج به‌دست‌آمده، طرح‌های MIX-1 و MIX-2 گزینه‌های مناسبی برای پروژه‌هایی هستند که نیاز به مقاومت فشاری پایدار در طولانی‌مدت دارند. این طرح‌ها برای سازه‌هایی که در معرض بارهای بلندمدت و شرایط محیطی سخت قرار دارند، مانند پل‌ها، سدها و ساختمان‌های بلندمرتبه، مناسب‌تر هستند. در مقابل، طرح‌هایی مانند MIX-6 که کاهش مقاومت بیشتری دارند، ممکن است برای پروژه‌هایی که اهمیت کمتری به دوام بتن در درازمدت داده می‌شود، استفاده شوند.برای بررسی دقت مدل‌های یادگیری ماشین، پیش‌بینی‌های مدل NARX با داده‌های آزمایشگاهی مقایسه شد. این مقایسه نشان داد که مدل، کاهش مقاومت فشاری در بلندمدت را با دقت بالایی پیش‌بینی کرده و نتایج آن با داده‌های واقعی همخوانی دارد.مدل NARX توانست میزان افت مقاومت فشاری را در سن سه ساله، با میانگین دقت ۹۲٪ پیش‌بینی کند.استفاده از داده‌های مصنوعی تولیدشده با روش TGAN باعث افزایش دقت مدل در تخمین کاهش مقاومت شد.تحلیل مقایسه‌ای نشان داد که طرح‌های دارای افزودنی‌های شیمیایی خاص MIX-5 و MIX-6 دارای خطای بیشتری در پیش‌بینی بودند، که نشان می‌دهد تأثیر افزودنی‌ها در بلندمدت کمتر قابل پیش‌بینی است.



شکل ۱۰: تغییرات متوسط نتایج مقاومت فشاری بتن برای شش مخلوط مختلف Mix-1 تا Mix-6 در طی زمان‌های مختلف

۵-۲- بحث

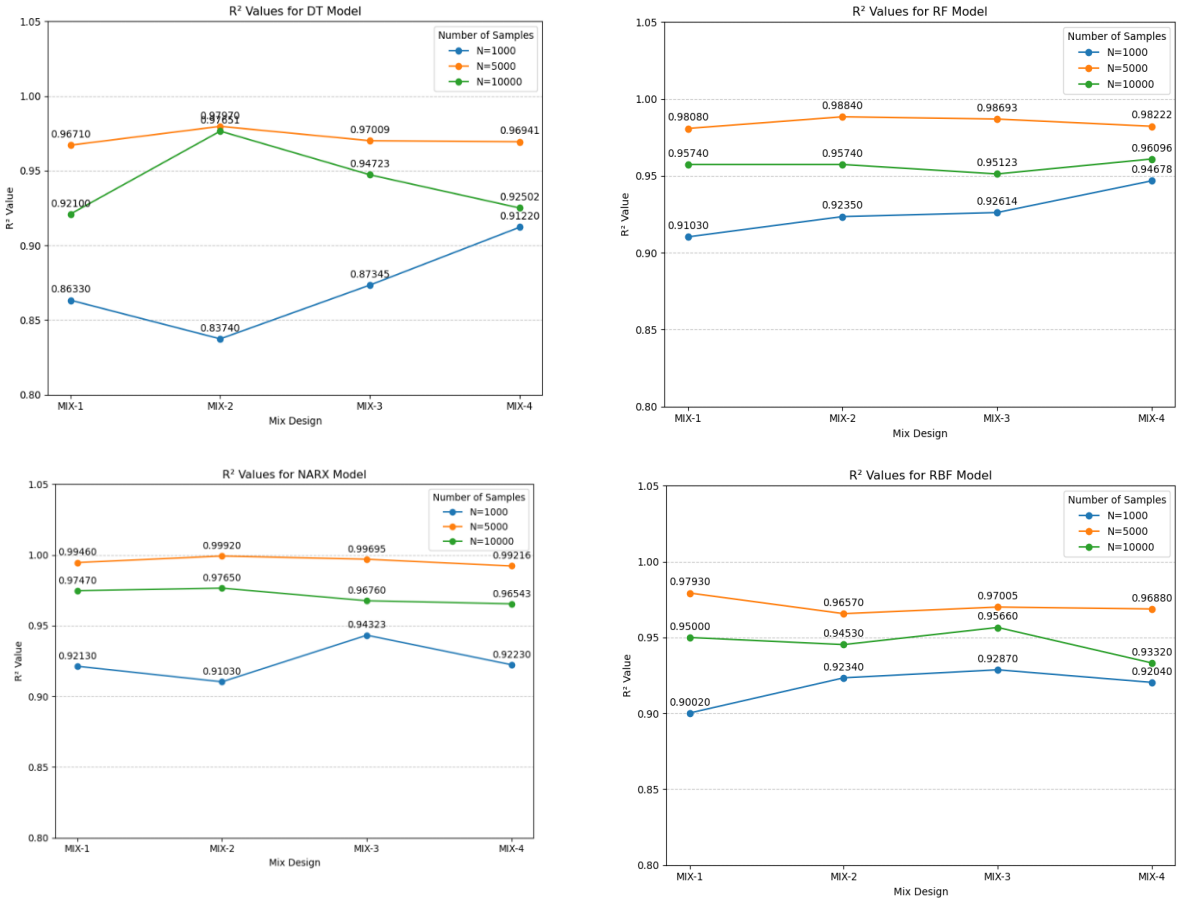
۵-۲-۱- انتخاب وضعیت داده‌های مصنوعی و تعداد داده‌های مصنوعی

نتایج ارزیابی عملکرد مدل‌ها که در شکل ۱۱ نشان داده شده است، نشان می‌دهد که افزایش تعداد نقاط داده‌های مصنوعی به ۵،۰۰۰ در تمام گروه‌های سنی بهترین عملکرد را به همراه دارد. بنابراین، در بخش‌های بعدی این مطالعه، تنها نتایج مدل‌های مبتنی بر این ۵،۰۰۰ داده در نظر گرفته شده است.

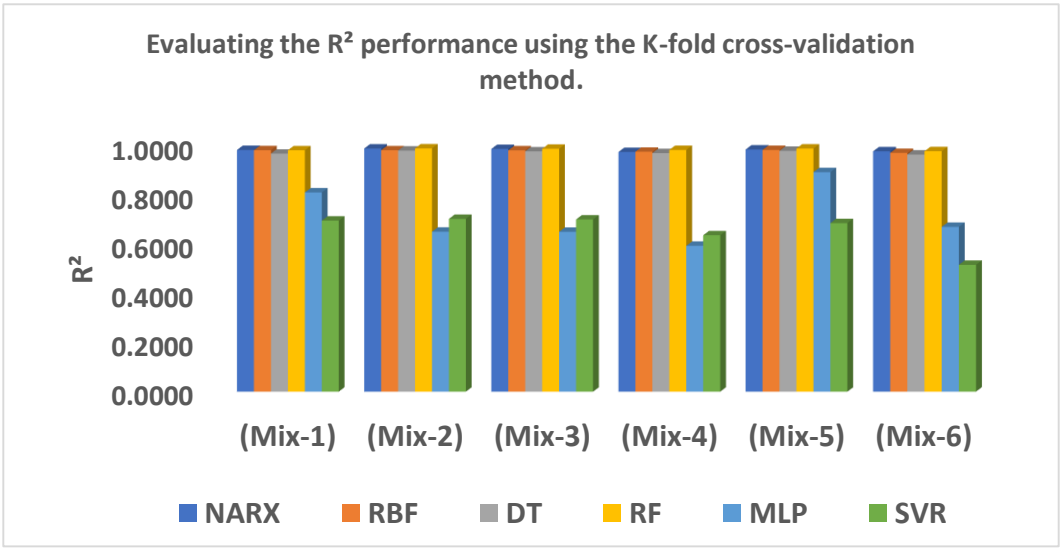
بر اساس مطالعه Shariati و همکاران [13]، در مدل پیش‌بینی مقاومت فشاری بتن (ANN) ، افزایش تعداد نمونه‌های آزمایشگاهی دقت پیش‌بینی مدل را از ۸۰٪ به ۹۵٪ بهبود داد، که نشان‌دهنده بهبود ۱۵٪ در دقت است. با توجه به اینکه افزایش تعداد نمونه‌های آزمایشگاهی دقت تخمین مقاومت فشاری را افزایش می‌دهد، این مطالعه تعداد نمونه‌های آزمایشگاهی را به ۷،۸۴۵ افزایش داده است. این افزایش نه تنها در مقایسه با مطالعات قبلی [38] [13] قابل توجه است، بلکه پتانسیل بهبود دقت پیش‌بینی مدل‌ها را نیز دارد.

به همین دلیل، یکی از ویژگی‌های متمایزکننده این پژوهش در مقایسه با سایر مطالعات، افزایش قابل توجه تعداد نمونه‌های آزمایشگاهی به فراتر از آنچه در تحقیقات قبلی گزارش شده است، می‌باشد. با این حال، برای افزایش بیشتر دقت مدل، نیاز به تولید داده‌های مصنوعی شناسایی شد. با توجه به اینکه تعداد کل داده‌های آزمایشگاهی در این مطالعه ۷،۸۴۵ است، ۳۲،۱۵۵ داده مصنوعی اضافی (در سطح ۵،۰۰۰) تولید شد که دقت و عملکرد مدل‌ها را بهبود بخشید. برای تعیین مقدار داده‌های مصنوعی، تشابه نسبت داده‌های آزمایشگاهی به داده‌های مصنوعی با مطالعه Marani و همکاران [11]، در نظر گرفته شد. این رویکرد اطمینان حاصل می‌کند که نسبت داده‌های تولید شده در این مطالعه با تحقیقات قبلی به خوبی هم‌راستا است.

در تحلیل رتبه‌بندی انطباق داده‌های مصنوعی با داده‌های واقعی برای چهار طرح اختلاط مختلف، MIX-2 بهترین عملکرد را از خود نشان می‌دهد. این طرح با کمترین اختلاف در میانگین، انحراف معیار، و چولگی در مقایسه با سایر طرح‌ها، دارای بالاترین سطح همخوانی است. از طرفی، MIX-1 که در رتبه دوم قرار گرفته، عملکرد قابل قبولی دارد اما نه به اندازه MIX-3 . MIX-2 با اختلاف‌های متوسط نسبت به داده‌های واقعی، در رتبه سوم قرار دارد. در حالی که MIX-6 با بیشترین اختلاف در همه معیارهای بررسی شده، ضعیف‌ترین انطباق را دارد و نشان می‌دهد که داده‌های مصنوعی آن به خوبی با داده‌های واقعی تطابق ندارند.



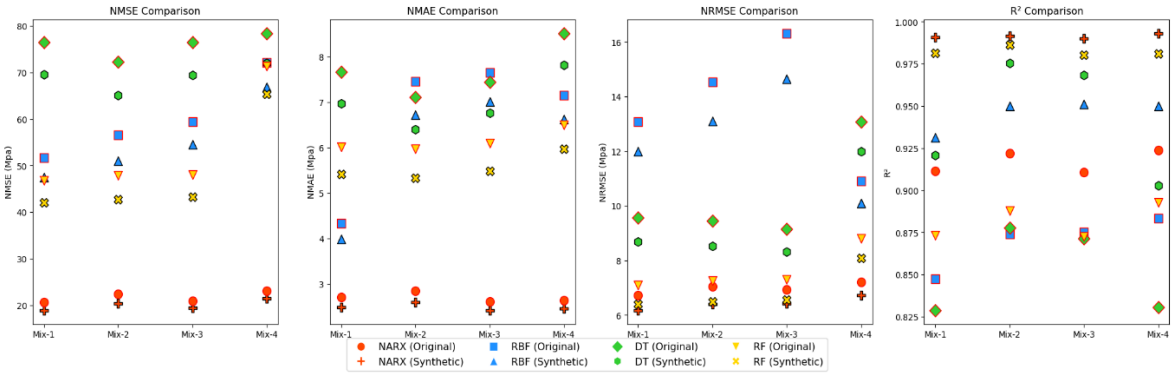
شکل ۱۱: نتایج شاخص برای سطوح داده‌های آزمون (۱۰۰۰، ۵۰۰۰، و ۱۰۰۰۰) برای چهار مخلوط و چهار مدل R^2



شکل ۱۲: شاخص‌های آماری برای اعتبارسنجی متقاطع **K-fold** برابر برای مدل‌ها و تعدادی از طرح اختلاط‌ها

۵-۲-۲- تحلیل آماری داده‌های واقعی و مصنوعی

شکل ۱۳ مقایسه مدل‌های مختلف با استفاده از معیارهای NMSE، NMAE، NRMSE، و R^2 برای داده‌های اصلی و مصنوعی را ارایه می دهد. بررسی شکل نشان می‌دهد که، استفاده از داده‌های مصنوعی به طور کلی منجر به بهبود صحت سنجی و تعمیم پذیری مدل‌ها شده است. داده‌های مصنوعی توانسته‌اند ضریب همبستگی بالاتری با داده‌های واقعی ایجاد کنند و پراکندگی خطاها را کاهش دهند، که نشان می‌دهد این داده‌ها برای بهبود دقت مدل‌ها مؤثر بوده‌اند. با این حال، برخی مدل‌ها نیاز به تنظیمات و داده‌های واقعی بیشتری دارند تا بهترین عملکرد را داشته باشند.به طور کلی، این نمودارها تأیید می‌کند که داده‌های مصنوعی، در بسیاری از مدل‌ها، به عنوان جایگزینی کارآمد برای داده‌های واقعی عمل کرده و می‌توانند در بسیاری از موارد حتی به بهبود دقت پیش‌بینی کمک کنند.



شکل۱۳: مقایسه مدل‌های مختلف براساس معیارهای NMSE, NMAE, NRMSE وR^2 برای داده‌های واقعی و مصنوعی در تعدادی از طرح اختلاط‌ها

نمودار ارائه شده مقایسه‌ای بین داده‌های واقعی و مصنوعی برای مدل‌های مختلط (NARX، RBF، DT، RF)در چند میکس مختلف و بر اساس معیارهای NMSE، NMAE، NRMSE و R^2 را نشان می‌دهد. این معیارها برای ارزیابی عملکرد مدل‌های مختلف در پیش‌بینی مقاومت فشاری بتن استفاده شده‌اند. داده‌های واقعی و مصنوعی در تمامی مدل‌ها و معیارها به دقت قابل قبولی نزدیک به هم هستند، با این حال تفاوت‌هایی نیز مشاهده می‌شود.

در مدل NARX برای Mix-1، داده‌های واقعی و مصنوعی تفاوت اندکی دارند؛ داده‌های واقعی ۲۰,۶۸ و داده‌های مصنوعی ۱۸,۹۷ است. در مدل RBF برای Mix-3، مقدار NMSE برای داده‌های واقعی ۷۶,۴۱ و برای داده‌های مصنوعی ۶۹,۴۶ است که نشان می‌دهد داده‌های مصنوعی با دقت خوبی تولید شده‌اند، اما اختلافی نسبت به داده‌های واقعی وجود دارد. در معیار NMAE، در Mix-2، مدل NARX مقدار NMAE برای داده‌های واقعی ۲,۸۴۹ و برای داده‌های مصنوعی ۲,۵۹ را نشان می‌دهد که تفاوت اندکی دارند. در مدل DT برای Mix-4، داده‌های واقعی و مصنوعی

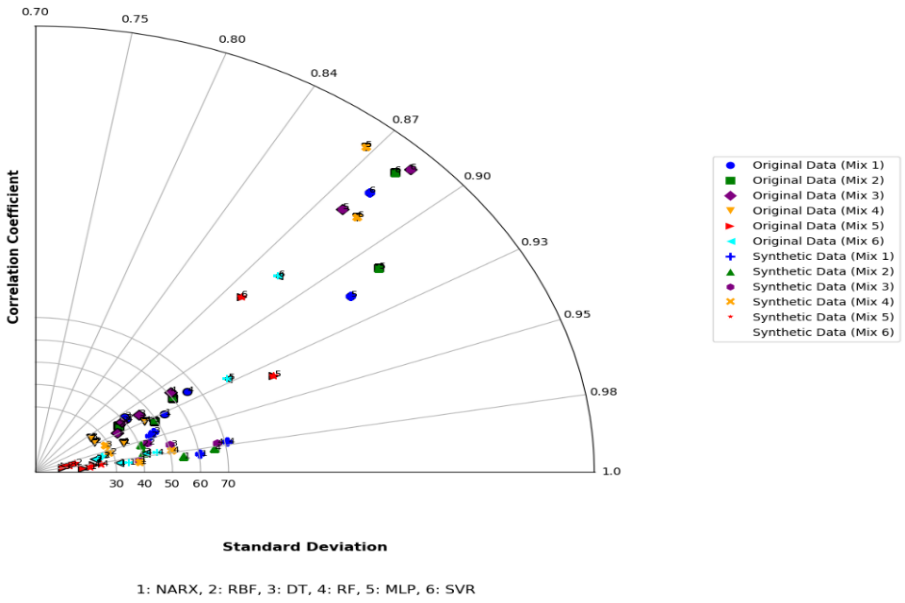
تفاوت بیشتری نشان می‌دهند، اما همچنان به هم نزدیک هستند. در مدل RBF برای Mix-3، مقدار NRMSE برای داده‌های واقعی ۲۰,۳۱ و برای داده‌های مصنوعی ۱۸,۶۴ است که نشان‌دهنده تفاوت قابل توجهی در عملکرد داده‌های مصنوعی و واقعی است. در سایر مدل‌ها، داده‌های مصنوعی و واقعی مقادیر مشابهی دارند و نشان‌دهنده نزدیکی دقیق هستند. داده‌های مصنوعی معمولاً همبستگی بالاتری نسبت به داده‌های واقعی نشان می‌دهند. به عنوان مثال، در Mix-4 برای مدل RF، داده‌های واقعی مقدار ۰,۸۹ و داده‌های مصنوعی مقدار ۰,۹۸ را نشان می‌دهند.

داده‌های مصنوعی تولید شده توسط مدل‌ها توانسته‌اند به خوبی به داده‌های واقعی نزدیک شوند. این داده‌های مصنوعی در بسیاری از موارد نشان داده‌اند که می‌توانند تعمیم‌پذیری بیشتری داشته باشند و حتی در برخی موارد عملکرد بهتری را از داده‌های واقعی ارائه دهند. به‌ویژه در معیار R^2، داده‌های مصنوعی همبستگی بهتری با نتایج پیش‌بینی شده دارند که این امر نشان‌دهنده بهبود در تعمیم‌پذیری و دقت مدل‌ها است.داده‌های مصنوعی تولید شده توسط مدل‌های مختلف در این مطالعه نشان‌دهنده دقت و تعمیم‌پذیری بالایی هستند. در بیشتر موارد، داده‌های مصنوعی به خوبی با داده‌های واقعی همخوانی دارند و در برخی معیارها مانند R^2، داده‌های مصنوعی توانسته‌اند عملکرد بهتری از خود نشان دهند. این امر نشان می‌دهد که داده‌های مصنوعی می‌توانند به عنوان جایگزینی مناسب برای داده‌های واقعی در پیش‌بینی مقاومت فشاری بتن به کار روند.

۵-۲-۳- مقایسه عملکرد مدل‌ها

پیش‌بینی مدل‌های مقاومت فشاری بتن در صنعت به خوبی شناخته شده است. به همین دلیل، محققان مدل‌های مختلفی را برای بهبود دقت پیش‌بینی مقاومت فشاری پیشنهاد داده‌اند [40]. نتایج نشان می دهدکه، برای داده های واقعی را ارائه می‌دهد. مدل NARX با بالاترین مقدار متوسط R^2 برابر با ۰,۹۹۶۸ در تمام طرح‌های اختلاط، بهترین عملکرد را در پیش‌بینی مقاومت فشاری بتن نشان می‌دهد. مدل NARX به دلیل ساختار شبکه عصبی بازگشتی، سازگاری بالا، و توانایی حفظ حافظه بلندمدت، در پیش‌بینی‌های بلندمدت و تحلیل سری‌های زمانی عملکرد برتری نسبت به سایر مدل‌ها نشان داده است. مدل RF نیز عملکرد قوی‌ای از خود نشان داد، به‌خصوص در Mix-1 با R^2 برابر با ۰,۹۸۱۳ که بسیار نزدیک به مدل NARX است. در مقابل، مدل‌های DT و RBF عملکرد قابل قبولی داشتند اما کمتر از مدل‌های NARX و RF موثر بودند. مدل DT با کمترین مقادیر R^2 و بالاترین مقادیر NMSE و NMAE، عملکرد کمتری نسبت به سایر مدلها نشان داد. شکل ۱۴ خلاصه‌ای از نتایج عملکرد مدل‌های مختلف در پیش‌بینی مقاومت فشاری بتن با داده‌های اصلی و مصنوعی را نشان می دهد .

در تحلیل نمودار تفاوت‌ها و شباهت‌های داده‌های واقعی و مصنوعی بین مدل‌های مختلف بررسی شد. در بیشتر مدل‌ها، داده‌های واقعی و مصنوعی به هم نزدیک هستند. به عنوان مثال، در Mix-1 و Mix-2، همبستگی ضریب (Correlation Coefficient) داده‌های واقعی و مصنوعی به‌طور قابل‌توجهی مشابه است. به طوری که مقادیر انحراف معیار (Standard Deviation) نیز نزدیک به یکدیگر هستند. به طور مشخص، در Mix-1، داده‌های مصنوعی با ضریب همبستگی حدود ۰,۹۸ و داده‌های واقعی با ضریب همبستگی حدود ۰,۸۷ مشاهده می‌شود. این شباهت در Mix-2 نیز با همبستگی‌هایی در حدود ۰,۹۹ برای داده‌های مصنوعی و ۰,۸۸ برای داده‌های واقعی حفظ شده است.در برخی موارد مانند Mix-3 و Mix-4، داده‌های مصنوعی و واقعی تفاوت‌های بیشتری دارند. به عنوان مثال، در Mix-4، داده‌های واقعی با همبستگی ضعیف‌تر (در حدود ۰,۸۰) در مقابل داده‌های مصنوعی با همبستگی قوی‌تر (در حدود ۰,۹۸) قرار دارند. در این حالت، انحراف معیار نیز تفاوت قابل توجهی دارد؛ به‌طوری که داده‌های مصنوعی انحراف بیشتری نسبت به داده‌های واقعی نشان می‌دهند.



شکل ۱۴: نمودار ارزیابی عملکرد مدل‌های مختلف در پیش‌بینی مقاومت فشاری بتن با داده‌های اصلی و مصنوعی

در Mix-1، داده‌های واقعی ضریب همبستگی ۰,۸۷ و داده‌های مصنوعی ۰,۹۹ دارند. این تفاوت نشان‌دهنده آن است که داده‌های مصنوعی تعمیم‌پذیری بالاتری را ارائه می‌دهند. همچنین، انحراف معیار برای داده‌های واقعی ۵۴ و

برای داده‌های مصنوعی ۶۰ است. این نشان می‌دهد که داده‌های مصنوعی می‌توانند دامنه‌ی وسیع‌تری از تغییرات را پوشش دهند. درMix-3، داده‌های واقعی ضریب همبستگی ۰,۸۲ و داده‌های مصنوعی ۰,۹۷ دارند. انحراف معیار برای داده‌های واقعی ۴۶ و برای داده‌های مصنوعی ۵۰ است. این موضوع می‌تواند نشان‌دهنده‌ی بهبود تعمیم‌پذیری مدل در داده‌های مصنوعی باشد، زیرا داده‌های مصنوعی قادر به پوشش بازه‌ی بیشتری از مقادیر هستند.

تعامل بین داده‌های مصنوعی و واقعی نشان می‌دهد که داده‌های مصنوعی به‌طور قابل توجهی همبستگی بیشتری با مدل‌های پیش‌بینی دارند. به‌ویژه در Mix-4 که داده‌های واقعی ضعیف‌تر عمل می‌کنند، داده‌های مصنوعی عملکرد بهتری از خود نشان داده‌اند. این نشان می‌دهد که مدل‌های مبتنی بر داده‌های مصنوعی قابلیت تعمیم‌پذیری بالاتری در محیط‌های مختلف دارند. داده‌های مصنوعی با داشتن همبستگی‌های بالاتر ۰,۹۸ در Mix-1 و ۰,۹۹ در (Mix-2 نسبت به داده‌های واقعی، نه تنها دقت بهتری ارائه می‌دهند، بلکه این مدل‌ها در شرایط مختلف نتایج پایدارتری را نیز ارائه می‌دهند. همچنین، انحراف معیار بیشتر در داده‌های مصنوعی) به عنوان مثال در Mix-3 نشان می‌دهد که این داده‌ها قادر به پوشش گسترده‌تری از شرایط پیش‌بینی هستند.

داده‌های مصنوعی به‌طور کلی توانسته‌اند در تمامی مدل‌ها بهبود قابل‌توجهی در دقت و تعمیم‌پذیری ارائه دهند. مدل‌های مصنوعی نه تنها همبستگی قوی‌تری را با متغیرهای پیش‌بینی نشان می‌دهند، بلکه به دلیل انحراف معیار بالاتر، ظرفیت بیشتری برای پوشش دامنه‌های گسترده‌تر دارند. این امر به ویژه در Mix-4 مشخص است که داده‌های مصنوعی توانسته‌اند ضعف‌های داده‌های واقعی را جبران کنند. در نتیجه، داده‌های مصنوعی در این مطالعه به عنوان ابزاری موثر برای بهبود مدل‌های پیش‌بینی در محیط‌های متنوع و پیچیده معرفی می‌شوند. مطالعات کمی به برآورد مقاومت فشاری در بتن حاوی مواد افزودنی شیمیایی متعدد در سنین بلند مدت پرداخته اند [3] [41]. یک مطالعه [42] اثرات طولانی مدت (تا ۳۶۵ روز) ترکیب مواد افزودنی شیمیایی و معدنی در بتن را بررسی کرد. یافته های این مطالعه به خوبی با نتایج تحقیق حاضر همخوانی دارد که باعث کاهش مقاومت بتن تا یک سال می شود. با این حال، تحقیقاتی در مورد سنین بیش از یک سال انجام نشده است.مقایسه روش‌های تولید داده‌های مصنوعی و برتری TGAN این پژوهش عملکرد روش‌های مختلف تولید داده‌های مصنوعی را در پیش‌بینی مقاومت فشاری بتن بررسی کرده است TGAN . نسبت به سایر روش‌ها، دقت بالاتری در حفظ ویژگی‌های اصلی داده‌های آزمایشگاهی نشان داده است. مقایسه مدل‌های یادگیری ماشین با و بدون داده‌های مصنوعی نشان داد که افزودن داده‌های تولید شده توسط TGAN، دقت مدل‌ها را ۱۰ تا ۲۰ درصد بهبود داده است. این تأثیر به‌ویژه در پیش‌بینی‌های بلندمدت مشهودتر است TGAN. در مقایسه با روش‌های دیگر مانند داده‌افزایی و GAN، توزیع داده‌های اصلی را بهتر بازتولید کرده و باعث افزایش دقت مدل‌های

یادگیری ماشین شده است. این امر TGAN را به گزینه‌ای کارآمد برای تولید داده‌های مصنوعی در مهندسی بتن تبدیل می‌کند.

نتایج به‌دست‌آمده به‌خوبی با نتایج پژوهش حاضر همخوانی دارد، که کاهش استحکام بتن را تا یک سال نشان می‌دهد. با این حال، تحقیقاتی برای بررسی سنین بیش از یک سال انجام نشده است. کاهش طولانی‌مدت استحکام بتن به عواملی مانند محیط‌های مخرب، اثرات خستگی، و بارگذاری‌های خاص مرتبط است که مورد بررسی قرار گرفته‌اند [1]. با این حال، کاهش طولانی‌مدت استحکام بتن به دلیل اثرات ترکیب‌های شیمیایی مختلف هنوز مورد تحقیق قرار نگرفته است. بنابراین، با توجه به اهمیت این مسئله در عملکرد نهایی و ایمنی سازه‌های بتنی در دوره عملیاتی، این مطالعه نیاز به بررسی کاهش استحکام فشاری و اثرات آن را به عنوان یک شکاف تحقیقاتی مهم شناسایی می‌کند.مطالعات گذشته [13][39] با چالش‌هایی مانند کمبود داده‌ها برای تخمین استحکام فشاری بتن حاوی ترکیبی از چندین افزودنی شیمیایی، از جمله فوق روان‌کننده‌ها، کندگیرکننده‌ها، و عوامل حباب‌زا مواجه شده‌اند. این تحقیق به‌طور مؤثر نشان داده است که تولید داده‌های مصنوعی و استفاده از فناوری‌های هوش مصنوعی در تخمین استحکام فشاری بتن با استفاده از مدل‌های مختلف، هزینه‌ها، زمان، و مصرف منابع را کاهش می‌دهد. در این مطالعه، استفاده از الگوریتم خاص NARX نسبت به مدل‌های مورد استفاده در تحقیقات، مزیت جدیدی ارائه می‌دهد. برای اطمینان از دقت و قابلیت تعمیم مدل‌های مورد استفاده در این مطالعه در مقایسه با تحقیقات قبلی، علیرغم کمبود تاریخی مطالعات، تلاش‌هایی برای اعتبارسنجی مدل با استفاده از داده‌های مطالعه انجام‌شده توسط [34] صورت گرفت. نتایج نشان می‌دهند که مدل‌های استفاده‌شده در این تحقیق، با بهره‌گیری از داده‌های مطالعه [34]، قابلیت پیش‌بینی قوی‌ای برای استحکام فشاری بتن دارند و ضریب تعیین (R^2) متوسطی برابر با ۰,۹۲ را به‌دست آورده‌اند.

۶- نتیجه‌گیری

در این پژوهش، نقش داده‌های مصنوعی در بهبود دقت و تعمیم‌پذیری مدل‌های پیش‌بینی مقاومت فشاری بتن بررسی شد. نتایج نشان دادند که افزایش تعداد داده‌های مصنوعی تا ۵،۰۰۰ نمونه، بهترین عملکرد را در مدل‌های مختلف به همراه دارد. به‌علاوه، با استفاده از ۷،۸۴۵ داده آزمایشگاهی و تولید ۳۲،۱۵۵ داده مصنوعی اضافی، به دقت بالاتری در پیش‌بینی مقاومت فشاری دست یافتیم. این امر نشان‌دهنده پتانسیل بالای استفاده از داده‌های مصنوعی برای بهبود مدل‌های پیش‌بینی است.

در این مطالعه، دو روش پیشرفته تولید داده‌های مصنوعی، یعنی K-means و TGAN، مورد ارزیابی قرار گرفتند. هر دو روش توانستند داده‌هایی با کیفیت آماری نزدیک به داده‌های واقعی تولید کنند. با این حال، روش K-means به‌عنوان یک رویکرد نوآورانه و مقرون به صرفه در تولید داده‌های مصنوعی انتخاب شد، زیرا علاوه بر دقت و بازتاب خصوصیات داده‌های واقعی، از نظر محاسباتی ساده‌تر و کم‌هزینه‌تر از روش TGAN است. این نتایج تأیید می‌کنند که خوشه‌بندی K-means می‌تواند به‌عنوان یک روش کارآمد برای تخمین مقاومت فشاری بلندمدت بتن به کار رود.با توجه به معیارهای ارزیابی مانند ضریب تعیین (R^2)، مدل‌های مختلف پیش‌بینی مورد بررسی قرار گرفتند. مدل NARX با مقدار متوسط R^2 برابر با ۰,۹۹۶۸ در تمام طرح‌های اختلاط، بهترین عملکرد را در پیش‌بینی مقاومت فشاری بتن از خود نشان داد. مدل RF نیز عملکرد قدرتمندی داشت، به‌ویژه در Mix-1 که R^2 برابر با ۰,۹۸۱۳ به‌دست آمد. سایر مدل‌ها مانند DT و RBF عملکرد قابل قبولی داشتند، اما در مقایسه با NARX و RF از دقت پایین‌تری برخوردار بودند. به‌طور کلی، استفاده از داده‌های مصنوعی توانست دقت پیش‌بینی مدل‌ها را بهبود بخشد، که در بسیاری از موارد به همبستگی بالاتری بین داده‌های واقعی و مصنوعی منجر شد . این یافته‌ها نشان می‌دهند که داده‌های مصنوعی می‌توانند به‌عنوان یک جایگزین کارآمد برای داده‌های واقعی در تخمین مقاومت فشاری بتن استفاده شوند.بررسی مقاومت فشاری بتن نشان داد که پس از افزایش اولیه طی ۹۰ روز، روند کاهش طولانی‌مدت مشاهده می‌شود. این امر بر اهمیت پیش‌بینی مقاومت بلندمدت بتن و نیاز به بازنگری پارامترهای طراحی سازه‌های بتنی تأکید دارد. به‌روزرسانی این پارامترها می‌تواند دوام و عملکرد بتن را بهبود بخشد. همچنین، استفاده از تکنیک‌های یادگیری ماشین موجب افزایش دقت پیش‌بینی و تعمیم‌پذیری مدل‌ها شده است.

این پژوهش به‌طور مؤثر نشان داد که تولید داده‌های مصنوعی و استفاده از مدل‌های هوش مصنوعی می‌تواند به کاهش هزینه‌ها، زمان و مصرف منابع در تخمین مقاومت فشاری بتن کمک کند همچنین با معرفی رویکردهای جدید در تولید داده‌های مصنوعی و مقایسه مدل‌های پیش‌بینی، گام مهمی در افزایش دقت و تعمیم‌پذیری پیش‌بینی مقاومت فشاری بتن برداشته است.. این موضوع به‌ویژه در مواقعی که داده‌های واقعی محدود هستند، اهمیت بیشتری پیدا می‌کند. در نهایت، این مطالعه به‌عنوان یک گام مهم در بهبود روش‌های پیش‌بینی مقاومت فشاری بتن با بهره‌گیری از داده‌های مصنوعی تلقی می‌شود. برای تحقیقات آتی، بررسی اثرات بلندمدت ترکیب‌های مختلف شیمیایی در مقاومت بتن با استفاده از داده‌های مصنوعی پیشنهاد می‌شود.

منابع

[1] A. Harapin, M. Jurišić, N. Bebek, and M. Sunara, "Long-Term Effects in Structures: Background and Recent Developments," *Applied Sciences (Switzerland)*, vol. 14, no. 6. Multidisciplinary Digital Publishing Institute (MDPI), 2024. doi: 10.3390/app14062352.

[2] M. R. Garcez, A. B. Rohden, and L. G. G. de Godoy, "The role of concrete compressive strength on the service life and life cycle of a RC structure: Case study," *J. Clean. Prod.*, vol. 172, pp. 27–38, 2018, doi: 10.1016/j.jclepro.2017.10.153.

[3] K. Bedada, A. Nyabuto, I. Kınotı, and J. Marangu, "Review on advances in bio-based admixtures for concrete," *J. Sustain. Constr. Mater. Technol.*, vol. 8, no. 4, pp. 344–367, 2023.

[4] S. Zeng, X. Wang, L. Hua, M. Altayeb, Z. Wu, and F. Niu, "Prediction of compressive strength of FRP-confined concrete using machine learning: A novel synthetic data driven framework," *J. Build. Eng.*, vol. 94, p. 109918, 2024, doi: https://doi.org/10.1016/j.jobe.2024.109918.

[5] R. Polo-Mendoza, G. Martinez-Arguelles, R. Peñabaena-Niebles, and J. Duque, "Development of a Machine Learning (ML)-Based Computational Model to Estimate the Engineering Properties of Portland Cement Concrete (PCC)," *Arab. J. Sci. Eng.*, no. Ml, 2024, doi: 10.1007/s13369-024-08794-0.

[6] H. A. T. Nguyen, D. H. Pham, and Y. Ahn, "Effect of Data Augmentation Using Deep Learning on Predictive Models for Geopolymer Compressive Strength," *Appl. Sci.*, vol. 14, no. 9, 2024, doi: 10.3390/app14093601.

[7] N. Chen *et al.*, "Virtual mix design: Prediction of compressive strength of concrete with industrial wastes using deep data augmentation," *Constr. Build. Mater.*, vol. 323, no. January, 2022, doi: 10.1016/j.conbuildmat.2022.126580.

[8] T. Aziz, H. Aziz, S. Mahapakulchai, and C. Charoenlarpnopparut, "Optimizing compressive strength prediction using adversarial learning and hybrid regularization," *Sci. Rep.*, vol. 14, no. 1, pp. 1–14, 2024, doi: 10.1038/s41598-024-69434-z.

[9] Q. Tian *et al.*, "Compressive strength of waste-derived cementitious composites using machine learning," *Rev. Adv. Mater. Sci.*, vol. 63, no. 1, 2024, doi: 10.1515/rams-2024-0008.

[10] Y. El Khessaimi *et al.*, "The Effectiveness of Data Augmentation in Compressive Strength Prediction of Calcined Clay Cements Using Linear Regression Learning," *NanoWorld J.*, vol. 9, 2023, doi: 10.17756/nwj.2023-s2-054.

[11] A. Marani, A. Jamali, and M. L. Nehdi, "Predicting ultra-high-performance concrete compressive strength using tabular generative adversarial networks," *Materials (Basel).*, vol. 13,

no. 21, pp. 1–24, 2020, doi: 10.3390/ma13214757.

[12] J. Pachouly, S. Ahirrao, K. Kotecha, G. Selvachandran, and A. Abraham, "A systematic literature review on software defect prediction using artificial intelligence: Datasets, Data Validation Methods, Approaches, and Tools," *Eng. Appl. Artif. Intell.*, vol. 111, no. November 2021, p. 104773, 2022, doi: 10.1016/j.engappai.2022.104773.

[13] M. Shariati, D. J. Armaghani, M. Khandelwal, J. Zhou, and M. Khorami, "Assessment of Longstanding Effects of Fly Ash and Silica Fume on the Compressive Strength of Concrete Using Extreme Learning Machine and Artificial Neural Network," *J. Adv. Eng. Comput.*, vol. 5, no. 1, p. 50, Mar. 2021, doi: 10.25073/jaec.202151.308.

[14] L. Jin, W. Yu, D. Li, and X. Du, "Numerical and theoretical investigation on the size effect of concrete compressive strength considering the maximum aggregate size," *Int. J. Mech. Sci.*, vol. 192, p. 106130, 2021.

[15] A. International, "Standard Specification for Concrete Aggregates- ASTM C33 / C33M-23," 2023. [Online]. Available: 10.1520/C0033_C0033M-23

[16] N. Zeminian, G. Guarino, and Q. Xu, "Chemical admixtures for the optimization of the AAC production," *Ce/papers*, vol. 2, no. 4, pp. 235–240, 2018.

[17] A. International, "Standard Practice for Making and Curing Concrete Test Specimens in the Laboratory: ASTM C192/C192M-24," West Conshohocken, PA, 2024, 2024. doi: 10.1520/C0192_C0192M-24.

[18] V. Çetin and O. Yıldız, "A comprehensive review on data preprocessing techniques in data analysis," *Pamukkale Univ. J. Eng. Sci.*, vol. 28, no. 2, pp. 299–312, 2022, doi: 10.5505/pajes.2021.62687.

[19] I. Nunez, A. Marani, M. Flah, and M. L. Nehdi, "Estimating compressive strength of modern concrete mixtures using computational intelligence: A systematic review," *Constr. Build. Mater.*, vol. 310, p. 125279, 2021, doi: https://doi.org/10.1016/j.conbuildmat.2021.125279.

[20] L. Xu and K. Veeramachaneni, "Synthesizing Tabular Data using Generative Adversarial Networks," 2018, [Online]. Available: http://arxiv.org/abs/1811.11264

[21] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[22] M. Z. Naser and A. H. Alavi, "Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences," *Archit. Struct. Constr.*, vol. 3, no. 4, pp. 499–517, 2023.

[23] A. N. Beskopylny *et al.*, "Prediction of the Compressive Strength of Vibrocentrifuged

Concrete Using Machine Learning Methods," *Buildings*, vol. 14, no. 2, pp. 1–20, 2024, doi: 10.3390/buildings14020377.

[24] F. Barreto, L. Moharkar, M. Shirodkar, V. Sarode, S. Gonsalves, and A. Johns, "Generative Artificial Intelligence: Opportunities and Challenges of Large Language Models BT - Intelligent Computing and Networking," V. E. Balas, V. B. Semwal, and A. Khandare, Eds., Singapore: Springer Nature Singapore, 2023, pp. 545–553.

[25] K. Miok, D. Nguyen-Doan, D. Zaharie, and M. Robnik-Šikonja, "Generating data using Monte Carlo dropout," in *2019 IEEE 15th International Conference on Intelligent Computer Communication and Processing (ICCP)*, IEEE, 2019, pp. 509–515.

[26] D. Kaur *et al.*, "Application of Bayesian networks to generate synthetic health data," *J. Am. Med. Informatics Assoc.*, vol. 28, no. 4, pp. 801–811, 2021.

[27] S. J. Alghamdi, "Prediction of Concrete's Compressive Strength via Artificial Neural Network Trained on Synthetic Data," *Eng. Technol. Appl. Sci. Res.*, vol. 13, no. 6, pp. 12404–12408, 2023.

[28] Y. Lu, M. Shen, H. Wang, X. Wang, C. van Rechem, and W. Wei, "Machine learning for synthetic data generation: a review," *arXiv Prepr. arXiv2302.04062*, 2023.

[29] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling tabular data using conditional gan," *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.

[30] M. Saito, E. Matsumoto, and S. Saito, "Temporal generative adversarial nets with singular value clipping," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2830–2839.

[31] P. Ziolkowski and M. Niedostatkiewicz, "Machine learning techniques in concrete mix design," *Materials (Basel).*, vol. 12, no. 8, 2019, doi: 10.3390/ma12081256.

[32] A. L. Kaminsky, Y. Wang, and K. Pant, "An efficient batch K-fold cross-validation voronoi adaptive sampling technique for global surrogate modeling," *J. Mech. Des.*, vol. 143, no. 1, p. 11706, 2021.

[33] K. Phinzi, D. Abriha, and S. Szabó, "Classification efficacy using k-fold cross-validation and bootstrapping resampling techniques on the example of mapping complex gully systems," *Remote Sens.*, vol. 13, no. 15, p. 2980, 2021.

[34] J. S. Chou and A. D. Pham, "Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength," *Constr. Build. Mater.*, vol. 49, pp. 554–563, 2013, doi: 10.1016/j.conbuildmat.2013.08.078.

[35] H. Song *et al.*, "Predicting the compressive strength of concrete with fly ash admixture

using machine learning algorithms," *Constr. Build. Mater.*, vol. 308, p. 125021, 2021.

[36] "Anaconda, Anaconda," *World's Most Pop. Data Sci. Platf.*, 2019.

[37] T. W. M. P. D. S. Platform, "Matlab R2021a," *World's Most Pop. Data Sci. Platf.*, [Online]. Available: The World's Most Popular Data Science Platform

[38] X. Hu, B. Li, Y. Mo, and O. Alselwi, "Progress in artificial intelligence-based prediction of concrete performance," *J. Adv. Concr. Technol.*, vol. 19, no. 8, pp. 924–936, 2021.

[39] H. Naderpour, A. H. Rafiean, and P. Fakharian, "Compressive strength prediction of environmentally friendly concrete using artificial neural networks," *J. Build. Eng.*, vol. 16, no. January, pp. 213–219, 2018, doi: 10.1016/j.jobe.2018.01.007.

[40] X. Dong, Y. Liu, and J. Dai, "Application of Fully Connected Neural Network-Based PyTorch in Concrete Compressive Strength Prediction," *Adv. Civ. Eng.*, vol. 2024, no. 1, p. 8048645, 2024.

[41] D. H. de Bem, D. P. B. Lima, and R. A. Medeiros-Junior, "Effect of chemical admixtures on concrete's electrical resistivity," *Int. J. Build. Pathol. Adapt.*, vol. 36, no. 2, pp. 174–187, 2018, doi: 10.1108/IJBPA-11-2017-0058.

[42] M. Saridemir, I. B. Topçu, F. Özcan, and M. H. Severcan, "Prediction of long-term effects of GGBFS on compressive strength of concrete by artificial neural networks and fuzzy logic," *Constr. Build. Mater.*, vol. 23, no. 3, pp. 1279–1286, Mar. 2009, doi: 10.1016/j.conbuildmat.2008.07.021.