

طبقه‌بندی آسیب در نمونه‌های ملات‌سیمانی حفره‌دار با استفاده از الگوریتم‌های یادگیری

ماشین

ابوالفضل فقیهی حبیب آبادی، حامد ملاداودی*

دانشکده مهندسی معدن، دانشگاه صنعتی امیرکبیر، تهران، ایران

* davoodi@aut.ac.ir

چکیده

این پژوهش به بررسی کارایی روش غیرمخرب اولتراسونیک در شناسایی و طبقه‌بندی آسیب در نمونه‌های ملات سیمانی با طرح‌های اختلاط مختلف می‌پردازد. بدین منظور، مجموعه‌ای از نمونه‌ها با مقادیر متفاوت محتوای سیمان (CC) تهیه و در دو وضعیت سالم و آسیب‌دیده مورد آزمایش قرار گرفتند. داده‌های حاصل از سیگنال‌های اولتراسونیک، پس از پیش‌پردازش و استخراج ویژگی‌های آماری و زمانی-فرکانسی، به‌عنوان ورودی به سه الگوریتم یادگیری ماشین شامل K نزدیک‌ترین همسایگان (KNN)، بیز ساده (Naïve Bayes) و جنگل تصادفی (Random Forest) اعمال شدند. نتایج تحلیل نشان داد که الگوریتم بیز ساده به‌دلیل توانایی در مدیریت ویژگی‌های مستقل و غیرهمبسته داده‌های اولتراسونیک، در مقایسه با سایر روش‌ها عملکرد دقیق‌تری دارد و با دقت $99/00 \pm 3/16$ درصد و یادآوری $97/5 \pm 7/91$ درصد بالاترین صحت طبقه‌بندی را ارائه کرد. نوآوری اصلی این تحقیق در ترکیب رویکردهای یادگیری ماشین با روش‌های غیرمخرب برای تحلیل تغییرات انرژی سیگنال‌های اولتراسونیک و تشخیص زودهنگام آسیب در ملات‌های سیمانی با نسبت‌های مختلف محتوای سیمان است. این چارچوب تحلیلی داده‌محور و تکرارپذیر، ضمن افزایش دقت و اطمینان در شناسایی آسیب، قابلیت توسعه برای پایش سلامت و ارزیابی مداوم سازه‌های بتنی در مقیاس‌های عملی را دارد و می‌تواند به‌عنوان ابزاری کارآمد در ارتقای دوام و ایمنی زیرساخت‌های عمرانی مورد استفاده قرار گیرد.

کلمات کلیدی

اولتراسونیک، یادگیری ماشین، K نزدیک‌ترین همسایگان، بیز ساده، جنگل تصادفی

بتن یکی از پرکاربردترین مصالح ساختمانی است که به دلیل مقاومت بالا، هزینه تولید پایین و تنوع گسترده، در ساخت سازه‌هایی مانند سدها، پل‌ها و ساختمان‌ها استفاده می‌شود [۱]. با توجه به اهمیت ایمنی و هزینه‌های بالای تعمیر، ارزیابی وضعیت بتن، به‌ویژه در سازه‌های قدیمی، ضروری است [۲].

در این پژوهش از نمونه‌های ملات‌سیمان استفاده شده که شامل ترکیبی از ماسه، سیمان و آب بوده و فاقد سنگدانه‌های درشت است. پس از عمل‌آوری، این مصالح ساختاری مقاوم و مشابه به سنگ طبیعی خواهند داشت. از مزایای این ترکیب می‌توان به وزن کمتر و دانه‌بندی دقیق‌تر نسبت به بتن‌های معمولی اشاره کرد که منجر به شباهت خواص فیزیکی و مکانیکی آن به سنگ طبیعی می‌شود. روش‌های آزمون بتن به دو دسته مخرب و غیرمخرب تقسیم می‌شوند [۳]. روش‌های مخرب، اگرچه دقیق هستند، به دلیل تخریب سازه، هزینه اجرایی بالا و زمان‌بر بودن، محدودیت‌هایی دارند [۴-۶]. در مقابل، روش‌های غیرمخرب^۱ (NDT) به دلیل سرعت بالا، هزینه کمتر و امکان بررسی جامع، به عنوان یک روش کارآمد در ارزیابی وضعیت نمونه‌های ملات‌سیمانی بدون آسیب به ساختار آن شناخته شده‌اند [۴].

این روش‌ها توانایی اندازه‌گیری دقیق پارامترهای کلیدی نمونه‌های ملات‌سیمانی مانند مقاومت فشاری، مدول الاستیسیته، شناسایی ترک‌ها، خوردگی آرماتور و چسبندگی میان نمونه‌های ملات‌سیمانی و آرماتور را دارند و می‌توانند در طول عمر سازه وضعیت آن را تحت نظر داشته باشند [۱،۴]. با وجود محدودیت‌هایی نظیر تأثیر جنس بتن و شرایط محیطی بر دقت نتایج [۱]، روش‌های غیرمخرب با پیشرفت‌های اخیر، به ابزاری مؤثر تبدیل شده‌اند [۷]. انتظار می‌رود که با استانداردسازی و توسعه تجهیزات و نیز همراه کردن این روش‌ها با الگوریتم‌های یادگیری ماشین، این روش‌ها در آینده به طور گسترده‌ای استفاده شوند [۸].

امروزه انواع مختلفی از آزمون‌های غیرمخرب ابداع شده‌اند که امکان بازرسی سازه‌های بتنی را بدون آسیب رساندن به این سازه‌ها فراهم می‌کنند. در این بین آزمایش غیرمخرب اولتراسونیک و آزمایش نشرآویی از اهمیت ویژه‌ای برخوردار می‌باشند. در روش اولتراسونیک می‌توان با تحلیل تغییرات خواص مکانیکی (سرعت یا مدت زمان عبور سیگنال ارسالی، نقص‌های داخلی و مشکلات ساختاری احتمالی را شناسایی کرد. یکی از چالش‌ها در این زمینه، شناسایی آسیب در سازه‌های بتنی ضخیم به دلیل کاهش دامنه سیگنال با افزایش فاصله فرستنده و گیرنده می‌باشد. استخراج ویژگی‌های مختلف سیگنال اولتراسونیک (مانند دامنه، فرکانس، فاز و مدت زمان عبور و بررسی آنان در فضای زمان-فرکانس، می‌تواند به تشخیص دقیق‌تر آسیب‌ها کمک کند [۹].

آزمایش غیرمخرب اولتراسونیک به دلیل نفوذ بالا، حساسیت در تشخیص آسیب‌های داخلی و توانایی ارائه داده‌های لحظه‌ای، یکی از مؤثرترین روش‌ها برای ارزیابی سلامت سازه‌های بتنی محسوب می‌شود [۱۰]. مطالعات متعدد این کارایی را تأیید کرده‌اند و با استفاده از ویژگی‌های چندفرکانسی امواج اولتراسونیک می‌تواند دقت شناسایی آسیب را به شکل قابل توجهی افزایش دهد [۱۱]. همچنین، مرور جامع ارائه‌شده بر پتانسیل بالای این فناوری در تشخیص آسیب داخلی بتن تأکید کرده است [۱۲]. علاوه بر این، ترکیب پردازش سیگنال اولتراسونیک و یادگیری ماشین را با موفقیت در شناسایی آسیب‌های کوچک در بتن نشان داده است [۱۳].

ترکیب تکنیک‌های پردازش سیگنال با الگوریتم‌های یادگیری ماشین، مدل‌های طبقه‌بندی را فراهم می‌کنند تا بتوان با داده‌هایی محدود الگوریتم‌هایی را برای شناسایی آسیب استخراج کرد.

در سال‌های اخیر، استفاده از فناوری اولتراسونیک برای تشخیص آسیب‌های سازه‌های بتنی توجه زیادی را به خود جلب کرده است. این فناوری به دلیل توانایی‌های خود در شناسایی تغییرات در خواص فیزیکی بتن، به‌عنوان یک ابزار مفید در ارزیابی سلامت سازه‌ها مطرح شده است. پژوهشگران متعددی در این زمینه فعالیت کرده‌اند و تکنیک‌های مختلفی برای پردازش و تحلیل سیگنال‌های اولتراسونیک به کار برده‌اند.

برای مثال، با استفاده از الگوریتم‌های یادگیری ماشین، سیگنال حاصل از اولتراسونیک را برای شناسایی آسیب‌های سازه‌های بتنی به کار بردند. این تحقیق به تحلیل سیگنال‌های اولتراسونیک و شناسایی آسیب‌ها از طریق الگوریتم‌های یادگیری ماشین پرداخته است

¹ Non-Destructive Testing

[۱۴]. در همین راستا، با استفاده از تبدیل موجک^۱ (WT)، توانسته‌اند برای شناسایی ترک‌ها در بتن از سیگنال‌های اولتراسونیک بهره ببرند. همچنین آنان با استفاده از پردازش سیگنال اولتراسونیک و ترکیب آن با الگوریتم‌های یادگیری ماشین^۲ (ML)، روشی برای شناسایی آسیب‌ها در بتن ارائه داده‌اند که توانسته است دقت شناسایی را به میزان قابل توجهی افزایش دهد [۱۳].

در ادامه، با شبیه‌سازی و تحلیل سیگنال‌های اولتراسونیک به تشخیص ترک‌ها و آسیب‌های درونی در بتن پرداخته‌اند. آن‌ها به این نتیجه رسیدند که با پردازش صحیح سیگنال‌ها، می‌توان اطلاعات دقیق‌تری از وضعیت سازه‌ها به دست آورد [۱۵]. همچنین، با استفاده از پردازش سیگنال اولتراسونیک و الگوریتم‌های یادگیری عمیق، به شناسایی آسیب‌های داخلی بتن پرداختند و نشان دادند که این روش‌ها می‌توانند دقت شناسایی آسیب‌ها را افزایش دهند [۱۶].

به طور کلی، این تحقیقات نشان‌دهنده پتانسیل بالای فناوری اولتراسونیک در تشخیص آسیب‌های سازه‌های بتنی هستند و استفاده از تکنیک‌های پیشرفته پردازش سیگنال و یادگیری ماشین می‌تواند بهبود قابل توجهی در دقت و سرعت ارزیابی سلامت سازه‌ها ایجاد کنند.

در این پژوهش به طبقه‌بندی آسیب در نمونه‌های شبه‌سنگی با استفاده از آزمایش غیرمخرب اولتراسونیک پرداخته می‌شود. برای شناسایی آسیب در نمونه‌های شبه‌سنگی تبدیل موجک گسسته^۳ (DWT) و تبدیل بسته موجک^۴ (WPT) به کار برده شده است و نتایج تحلیل شده است [۱۷]. بر اساس تبدیلات استخراج‌شده، در این تحقیق یک بانک اطلاعاتی از تبدیلهای موجک تهیه گردید. در نهایت به منظور طبقه‌بندی نمونه‌ها از الگوریتم‌های یادگیری ماشین K نزدیک‌ترین همسایگان^۵ (KNN)، بیز ساده^۶ و جنگل تصادفی^۷ استفاده شد تا مناسب‌ترین مدل برای طبقه‌بندی آسیب بر مبنای پارامترهای آماری شناسایی شود.

مطالعات پیشین در زمینه یادگیری ماشین پیشرفت‌های قابل توجهی را نشان داده‌اند. با این حال، روش‌های یادگیری ماشین به‌کاررفته در طبقه‌بندی نمونه‌های سالم و معیوب، نه تنها به صورت جامع تحلیل نشده‌اند، بلکه دقت پایین این روش‌ها گزارش شده است [۱۸]. در این مقاله، با به‌کارگیری روش‌های یادگیری ماشین، طبقه‌بندی دقیقی برای شناسایی نمونه‌های سالم و معیوب ارائه شده است.

۲- مطالعات تجربی

برای ارزیابی قابلیت‌شناسی عیوب با استفاده از دستگاه Pundit Lab، مجموعه‌ای از نمونه‌ها با محتوای سیمان (CC^۸) و طرح‌های اختلاط متنوع که دارای حفره بودند، مورد آزمایش قرار گرفتند. در ادامه، نمونه‌های مورد استفاده، فرآیند ساخت آن‌ها و روش کار دستگاه Pundit Lab به صورت مفصل تشریح خواهد شد. توضیحات کامل مربوط به روند ساخت، آماده‌سازی، و آزمایش نمونه‌ها پیش‌تر در مطالعه‌ی [۱۷] ارائه شده است و در این مقاله تنها نکات ضروری و مرتبط بیان می‌شود.

۲-۱- مشخصات نمونه‌ها

در گام نخست پژوهش، آزمایش غیرمخرب اولتراسونیک بر روی پنج نمونه ملات سیمانی با مقادیر متفاوت محتوای سیمان (CC) که در آزمایشگاه تهیه شده‌اند انجام شد. شکل ۱ نمای کلی این نمونه‌ها را نشان می‌دهد.

¹ Wavelet Transform

² Machine learning

³ Discrete Wavelet Transform

⁴ Wavelet Packet Transform

⁵ K-Nearest Neighbors

⁶ Naïve Bayes

⁷ Random Forest

⁸ Cement Content



شکل ۱: نمایی از نمونه‌های ملات‌سیمانی با محتوای سیمان (CC) متفاوت

Fig 1: An illustration of cement mortar samples with different cement contents (CC).

همان‌گونه که در جدول ۱ مشاهده می‌شود، متغیر اصلی در این آزمایش نسبت آب به سیمان است که مقدار آن برای هر یک از نمونه‌ها به صورت متفاوت در نظر گرفته شده است.

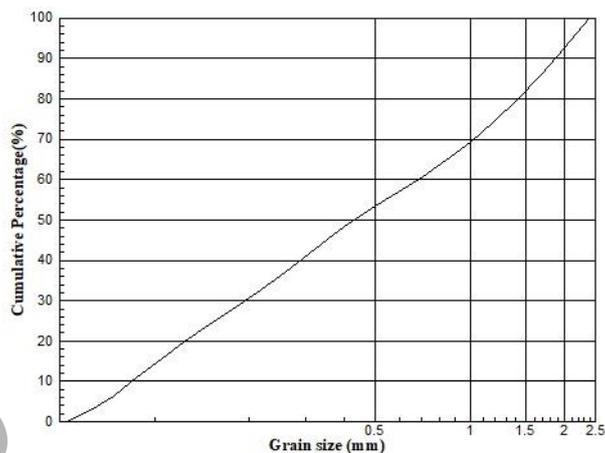
جدول ۱: مشخصات نمونه‌های ملات‌سیمانی با محتوای سیمان (CC) متفاوت

Table 1: Specifications of cement mortar samples with different cement contents (CC).

اسم نمونه	ابعاد نمونه (cm)	آب (%)	سیمان (%)	ماسه (%)
CC10	12 × 12 × 12.5	10	10	80
CC15	12 × 12 × 12.5	10	15	75
CC20	12 × 12 × 12.5	10	20	70
CC25	12 × 12 × 12.5	10	25	65
CC30	12 × 12 × 12.5	10	30	60

این تفاوت در نسبت آب به سیمان منجر به تغییر در چگالی و میزان تخلخل نمونه‌ها شده است. به‌طور طبیعی، با افزایش نسبت آب به سیمان، سرعت پالس اولتراسونیک کاهش می‌یابد. افزایش میزان آب، که موجب تخلخل بیشتر در ساختار نمونه می‌شود، پراکندگی امواج اولتراسونیک را افزایش داده و در نتیجه، سرعت عبور آن‌ها را کاهش می‌دهد. علاوه بر این، به‌منظور بررسی توانایی روش اولتراسونیک در شناسایی آسیب، در مرکز تمامی نمونه‌ها سوراخی استوانه‌ای با قطر ۱ سانتی‌متر ایجاد گردید. این سوراخ‌ها به‌عنوان آسیب مصنوعی در نظر گرفته شدند تا امکان مطالعه‌ی دقیق‌تر ارتباط بین میزان آسیب و سرعت پالس فراهم شود.

مواد اولیه به‌کاررفته در ساخت نمونه‌ها شامل آب شهری، سیمان پرتلند تیپ ۲ و ماسه شسته‌شده فیروزکوه است که مشخصات آن‌ها مطابق با الزامات استانداردهای ملی و بین‌المللی، از جمله استاندارد ملی ایران ISIRI 389 و استاندارد ASTM C150 آمریکا، می‌باشد [۱۹]. منحنی دانه‌بندی ماسه در شکل ۲ ارائه شده است.



شکل ۲: منحنی دانه‌بندی ماسه مورد استفاده در نمونه‌های CC [۱۷].

Fig 2: The grading curve of the sand used in the CC samples [17].

۲-۲- فرآیند آزمایش غیرمخرب اولتراسونیک

به منظور ارزیابی تأثیر ترک بر ویژگی‌های نمونه‌های ملات سیمانی، آزمایش غیرمخرب اولتراسونیک با استفاده از دستگاه Pundit Lab مدل ۲۰۱۷ در آزمایشگاه مهندسی مکانیک سنگ دانشگاه صنعتی امیرکبیر انجام گرفت. نمای کلی دستگاه به همراه تجهیزات جانبی آن در شکل ۳ نمایش داده شده است.

پیش از انجام آزمایش بر روی نمونه‌های ملات سیمانی، کالیبراسیون دقیق دستگاه الزامی بود. بدین منظور، پروب‌های دستگاه به طور کامل با گریس مخصوص پوشانده شده و سپس به دو سر نمونه کالیبراسیون متصل گردیدند. پس از برقراری اتصال مناسب، فرآیند کالیبراسیون دستگاه انجام شد.



شکل ۳: نمای کلی از دستگاه اولتراسونیک

Fig 3: An overview of the ultrasonic device.

در دستگاه Pundit Lab 2017، با اندازه‌گیری ضخامت نمونه و زمان عبور موج، می‌توان سرعت موج عبوری را محاسبه نمود. به منظور تحلیل دقیق‌تر پارامترهای مربوط به سرعت و زمان عبور موج، ثبت و بررسی شکل سیگنال ضروری است. برای این منظور از نرم‌افزار Punditlink استفاده گردید. بر اساس دستورالعمل کاربر دستگاه Pundit Lab، تنظیمات نرم‌افزار مطابق با مقادیر ارائه شده در جدول ۲ تعیین شدند [۱۷].

جدول ۲: تنظیمات اولیه دستگاه اولتراسونیک مطابق با راهنمای کاربر [۱۷].

Table 2: Initial settings of the ultrasonic device according to the user manual [17].

مقدار تنظیم شده	متغیر	واحد	پارامتر
۵۴	۱۵۰, ۸۲, ۵۴, ۳۷, ۲۴	kHz	فرکانس
۹/۳	۵۰۰, ۲۵۰, ۲۲۰, ۲۰۰		
	۱۰۰-۱	μs	عرض پالس

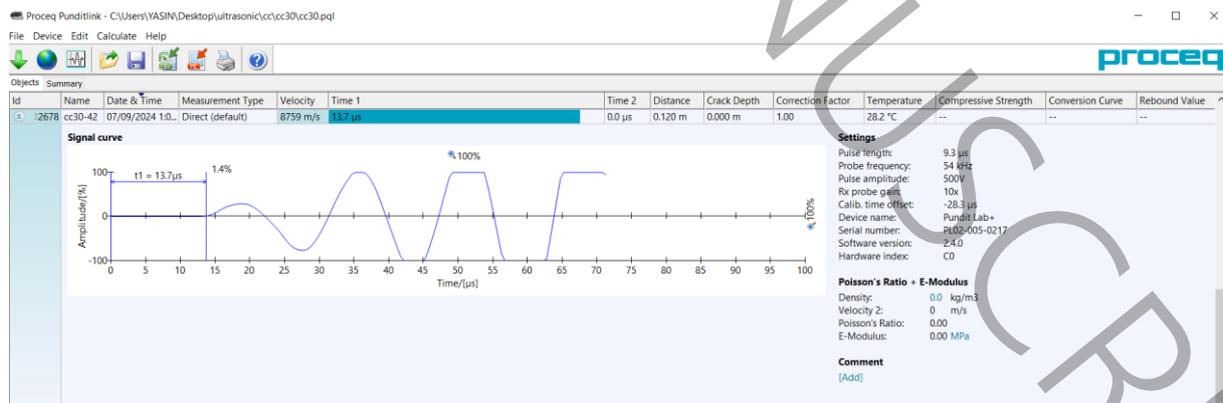
AUTO	۹۹۹۹-۰,۱	μs	دامنه
۰,۱	۱-۰,۱	μs	وضوح
۲/۷	۱۱۰-۱	μs	کالیبراسیون
زمان	زمان، سیگنال	-	میانگین گیری
AUTO	, ۵۰۰, ۳۵۰, ۲۵۰, ۱۲۵	-	ولتاژ تحریک
AUTO	AUTO	-	سودگیرنده
AUTO	x, AUTO ۱۰۰x, ۱۰x, ۱	-	واحدها
M	ft, m	-	عامل تصحیح
۱/۰۰	۱,۳-۰,۰۷	-	

با توجه به محدودیت دستگاه در ارائه‌ی خروجی مستقیم از داده‌های سیگنال، از نرم‌افزار GetDataW برای استخراج داده‌های آماری هر نمودار، از جمله دامنه و مدت‌زمان، استفاده شد. در این نرم‌افزار با تعیین چهار نقطه مرجع (دو نقطه در امتداد محور X و دو نقطه در امتداد محور Y)، مختصات نقاط مربوط به Waveform با دقت استخراج گردید. تمامی آزمایش‌های اولتراسونیک در این پژوهش مطابق با الزامات استانداردهای EN 12504-4 (Europe) و ASTM C597-02 (North America) انجام شده است [۲۰].

طول‌موج‌های بسیار زیاد موجب عدم شناسایی عیوب کوچک می‌شوند و در نتیجه سیگنالی با عمق نفوذ بالا اما حساسیت پایین ایجاد می‌کنند. در مقابل، طول‌موج‌های بسیار کوتاه سبب تضعیف قابل توجه موج ارسالی شده و سیگنالی با حساسیت بالا اما عمق نفوذ محدود به‌دنبال دارند. از این‌رو، انتخاب طول موج مناسب برای ایجاد تعادل میان حساسیت و عمق نفوذ از اهمیت ویژه‌ای برخوردار است. همچنین، واحد کنترل اصلی سیستم باید دارای دقت بالای اندازه‌گیری زمان، حداقل با وضوح ۰,۱ میکروثانیه باشد [۳].

برای هر یک از نمونه‌ها، مجموعه‌ای از آزمایش‌های اولتراسونیک انجام گرفت. نقاط آزمایش بر روی هر نمونه در قالب یک شبکه منظم با فواصل یکسان تعیین شدند و اندازه‌گیری‌ها به‌صورت جفتی، در وجوه مقابل یکدیگر، صورت پذیرفت. تعداد نقاط آزمایش برای هر نمونه بیست نقطه (در قالب آرایه‌ای چهار در پنج) بوده است. شکل مربوط به چیدمان آزمایش در مرجع [۱۷] به‌طور کامل ارائه شده است. برای اطلاع بیشتر به مرجع [۱۷] مراجعه شود. به‌منظور افزایش دقت و اطمینان از تکرارپذیری نتایج، در هر نقطه آزمایش ده مرتبه تکرار شد و میانگین داده‌های حاصل به‌عنوان مقدار نهایی مورد استفاده قرار گرفت.

این آزمایش با استفاده از پروب‌های ۵۴ کیلوهرتزی که به‌صورت مقابل یکدیگر قرار گرفته بودند، انجام شد. بر اساس مطالعات نایک^۱ و مالهوترا^۲، انتخاب این محدوده‌ی فرکانسی به‌منظور کاهش تأثیر تغییرات مقاومت بتن در فرآیند اندازه‌گیری صورت گرفته است [۲۱]. در این آزمایش، پارامترهایی نظیر سرعت موج طولی، زمان عبور موج و شکل موج سیگنال به‌طور مستقیم اندازه‌گیری شده و نتایج حاصل به‌صورت دیجیتالی در نرم‌افزار Punditlink ثبت گردیدند (شکل ۴).



شکل ۴ : نمایی از سیگنال ثبت شده از آزمایش اولتراسونیک

Figure 4: An illustration of the recorded signal from the ultrasonic test.

¹ Naik

² Malhotra

شایان ذکر است که آزمایش سرعت پالس اولتراسونیک در این پژوهش مطابق با استاندارد ASTM C-597-0 انجام شده است [۲۲].

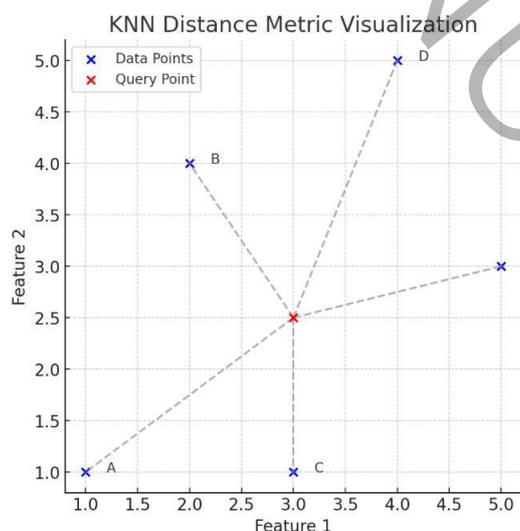
۳- یادگیری ماشین

اگرچه پارامتر انرژي استخراج شده از تحلیل موجک میان مناطق سالم و آسیب دیده تفاوت قابل توجهی دارد، اما تعیین آستانه‌ای مشخص برای جداکردن این دو طبقه دشوار و تا حدی کیفی است. استفاده از مدل‌های طبقه‌بندی یادگیری ماشین با آموزش روی مجموعه‌ای از ویژگی‌های موجکی مربوط به سیگنال‌های سالم و آسیب‌دیده، امکان تفکیک کمی و خودکار نمونه‌ها را فراهم می‌کند. پس از تهیه یک بانک اطلاعاتی کامل از سیگنال‌های برچسب‌خورده، می‌توان علاوه بر صرفه‌جویی در اجرای مکرر تحلیل موجک، نمونه‌های بعدی را بدون پردازش مجدد مستقیماً طبقه‌بندی نمود.

در این پژوهش، از سه الگوریتم یادگیری ماشین شامل K نزدیک‌ترین همسایگان (KNN)، بیز ساده و جنگل تصادفی برای طبقه‌بندی نمونه‌های ملات‌سیمانی به دو گروه سالم و آسیب‌دیده، بر اساس ویژگی‌های استخراج شده از داده‌های آزمایش اولتراسونیک، استفاده شده است. دلیل انتخاب این سه الگوریتم برای مقایسه، کارایی گسترده آن‌ها در طبقه‌بندی داده‌های دودویی است [۲۳]. در ادامه به شرح مختصر هر یک از این سه الگوریتم یادگیری ماشین پرداخته شده است.

۳-۱- K نزدیک‌ترین همسایگان (KNN)

الگوریتم K نزدیک‌ترین همسایگان نقاط داده‌های جدید را بر اساس معیارهای شباهت با استفاده از مجموعه داده‌های موجود طبقه‌بندی می‌کند. به عنوان مثال، از فاصله اقلیدسی برای محاسبه شباهت بین نقاط استفاده می‌کند و دسته مناسب را بر اساس نزدیک‌ترین همسایه‌ها اختصاص می‌دهد. یکی از چالش‌های اصلی الگوریتم K نزدیک‌ترین همسایگان انتخاب تعداد خوشه طبقه‌بندی (K) است، زیرا این امر مستقیماً بر دقت و عملکرد الگوریتم تأثیر می‌گذارد و باید به دقت تنظیم شود. مقدار بسیار کمی از K می‌تواند الگوریتم را به نوبت حساس کند و باعث برآزش بیش از حد شود. در چنین مواردی، مدل کاملاً با داده‌های آموزشی مطابقت دارد، اما ممکن است هنگام پیش‌بینی داده‌های جدید ضعیف عمل کند که منجر به دقت و عملکرد غیرواقعی بالا می‌شود. همچنین، مقادیر بسیار بزرگ ممکن است به دلیل در نظر گرفتن داده‌های نامربوط و مواجهه با زیر برآزش، دقت طبقه‌بندی را کاهش دهد. این بدان معنی است که مدل قادر به یادگیری روابط پیچیده در داده‌ها نیست و در تعمیم به داده‌های جدید ضعیف عمل می‌کند. همچنین عملکرد این روش نه تنها به مقدار بهینه K، بلکه به معیار فاصله مورد استفاده (مانند فاصله اقلیدسی، فاصله منهتن، یا فاصله کسینوسی) نیز وابسته است. ۳-۱- شکل ۵، نحوه تعیین مرکز دسته‌ها به صورت شماتیک و بر اساس فاصله پنج داده مختلف از نقطه مورد نظر نشان داده شده است. این تصویر گویای چگونگی تأثیر فاصله بر فرآیند تخصیص دسته به داده‌های جدید است [۲۴].



شکل ۵ : نحوه انتخاب مرکز دسته‌ها بر اساس فاصله پنج داده مختلف [۲۴]

Fig 5: The method of selecting cluster centers based on the distance of five different data points [24].

الگوریتم K نزدیک‌ترین همسایگان به دلیل سادگی در پیاده‌سازی و کارایی بالا برای مجموعه داده‌های کوچک تا متوسط، بسیار مناسب است. با این حال، برای مجموعه داده‌های بزرگ، به دلیل هزینه بالای محاسباتی ناشی از محاسبه فاصله برای تمام نمونه‌ها، استفاده از این الگوریتم کمتر مقرون به صرفه است. معادله (۱) نحوه محاسبه فاصله اقلیدسی را برای هر نقطه داده X_i نشان می‌دهد:

$$d(X_q, X_i) = \sqrt{\sum_{j=1}^n (X_{qj} - Y_{ij})^2} \quad (1)$$

در معادله (۱)، d نمایانگر فاصله داده مورد بررسی (X_q) از داده‌های دلخواه (X_i) است و n نشان‌دهنده تعداد ویژگی‌های بررسی شده در مجموعه داده است [۲۵].

در بررسی مقادیر K چهار معیار اصلی برای مدل‌سازی و ارزیابی عملکرد مدل وجود دارد که عبارتند از:

- ۱- دقت: میزان پیش‌بینی‌های درست مدل را نشان می‌دهد و به عنوان یک معیار کلی برای سنجش عملکرد مدل طبقه‌بندی مورد استفاده قرار می‌گیرد. این پارامتر از تقسیم تعداد پیش‌بینی‌های صحیح به کل پیش‌بینی‌ها به دست می‌آید.
- ۲- خطا: معیاری برای سنجش میزان خطای مدل است که نشان می‌دهد که چقدر مدل در پیش‌بینی‌ها دچار اشتباه می‌شود.
- ۳- یادآوری: به این سوال پاسخ می‌دهد که از بین تمام موارد درست (مثلاً موارد معیوب)، چند مورد توسط مدل به درستی شناسایی شده است. این پارامتر توانایی مدل در شناسایی تمامی موارد صحیح را اندازه‌گیری می‌کند.
- ۴- صحت: نشان می‌دهد از بین پیش‌بینی‌های درست مدل، چند مورد واقعاً صحیح بوده‌اند و به نوعی دقت پیش‌بینی‌های مثبت مدل را بررسی می‌کند. برای تحلیل دقیق‌تر عملکرد مدل‌های طبقه‌بندی، این چهار پارامتر به‌طور گسترده استفاده می‌شوند [۲۶].

۳-۲- بیز ساده

مدل بیز ساده یک الگوریتم طبقه‌بندی احتمالی است که بر اساس قضیه بیز عمل می‌کند. این الگوریتم فرض می‌کند که ویژگی‌ها کاملاً مستقل از یکدیگر هستند، که به همین دلیل به آن *نا نا آگاهانه* گفته می‌شود. به عبارت دیگر، در این مدل فرض بر این است که هر ویژگی در داده‌ها بدون در نظر گرفتن سایر ویژگی‌ها می‌تواند در یک کلاس خاص ظاهر شود. این مدل احتمال حضور هر ویژگی در یک کلاس را با فرض توزیع نرمال تخمین می‌زند و سپس نمونه‌ها را به کلاسی با بالاترین احتمال تخصیص می‌دهد. مدل بیز ساده مانند K نزدیک‌ترین همسایگان از نظر محاسباتی بسیار کارآمد است و برای مجموعه داده‌های کوچک عملکرد مناسبی دارد.

با این حال، یکی از محدودیت‌های اصلی این الگوریتم، فرض استقلال ویژگی‌ها است. در داده‌های واقعی، این فرض ممکن است برقرار نباشد، زیرا بسیاری از ویژگی‌ها ممکن است به طور بالقوه با یکدیگر وابستگی داشته باشند. در چنین شرایطی، این فرض می‌تواند دقت نتایج را تحت تأثیر قرار دهد و باعث کاهش عملکرد مدل در طبقه‌بندی داده‌ها شود. بیز ساده در این الگوریتم به شکل معادله (۲) است:

$$P(C_k | x) = \frac{P(C_k) \cdot P(x | C_k)}{P(x)} \quad (2)$$

در این معادله

$P(C_k | x)$: احتمال حضور داده در دسته بعدی C_k با توجه به داده x

$P(C_k)$: احتمال حضور داده در دسته قبلی C_k

$P(x | C_k)$: احتمال مشاهده داده x در دسته C_k

$P(x)$: احتمال مشاهده داده x (که برای همه دسته ثابت است) می‌باشد [۲۷].

جنگل تصادفی یک روش یادگیری گروهی^۱ است که با ساخت چندین درخت تصمیم^۲ و ترکیب پیش‌بینی‌های آن‌ها، نتیجه نهایی را به دست می‌آورد. این الگوریتم به‌طور خودکار با ایجاد تعدادی درخت تصمیم و انجام پیش‌بینی براساس هر یک از آن‌ها، نتایج را ترکیب می‌کند تا به یک پیش‌بینی نهایی برسد. برای بهینه‌سازی عملکرد مدل، پارامترهای مختلف مانند تعداد درخت‌ها و حداکثر عمق درخت‌ها از طریق جستجوی شبکه تنظیم می‌شوند.

همچنین، یکی از مزایای مهم این الگوریتم، ارزیابی اهمیت ویژگی‌ها است که به شناسایی مهم‌ترین ویژگی‌ها برای طبقه‌بندی کمک می‌کند. این ویژگی برای تحلیل و شناسایی ویژگی‌های اثرگذار در فرآیند طبقه‌بندی آسیب مفید است [۲۶].

در جنگل تصادفی، چهار معیار اصلی برای مدل‌سازی و ارزیابی عملکرد مدل وجود دارد که عبارتند از:

۱. دقت^۳: درصد پیش‌بینی‌های درست نسبت به کل پیش‌بینی‌ها، که نشان‌دهنده توانایی مدل در دسته‌بندی صحیح داده‌ها است.
۲. شاخص جینی^۴: معیاری برای اندازه‌گیری ناخالصی یا تنوع داده‌ها در هر گره درخت تصمیم است. مقادیر نزدیک به صفر نشان‌دهنده خلوص بالا و مقادیر نزدیک به یک نشان‌دهنده تنوع زیاد در گره‌ها هستند.
۳. افزایش اطلاعات^۵: معیاری برای سنجش کاهش بی‌نظمی^۶ در نتیجه تقسیم داده‌ها بر اساس یک ویژگی خاص است. این معیار معمولاً در درختان تصمیم برای انتخاب ویژگی تقسیم‌کننده استفاده می‌شود.
۴. نسبت بهره^۷: نسبت اینفرمیشن گین به طول مدت تقسیم برای کنترل تمایل مدل به انتخاب ویژگی‌هایی با تعداد مقادیر زیاد است. این معیار کمک می‌کند تا انتخاب‌های مناسب‌تری برای ویژگی‌های تقسیم‌کننده در درخت‌های تصمیم انجام شود [۲۶].

۴- تحلیل نتایج

ابتدا ماتریس همبستگی بین ویژگی‌های استخراج‌شده محاسبه گردید تا ساختار درونی داده‌ها و روابط خطی میان مؤلفه‌ها روشن شود. این گام به شناسایی همبستگی‌های قوی و مؤلفه‌های کمک می‌کند. پس از بررسی مجموعه ویژگی‌ها با تکیه بر تحلیل همبستگی، داده‌ها برای آموزش مدل‌های طبقه‌بندی یادگیری ماشین مورد استفاده قرار گرفتند تا جداسازی کمی کلاس‌های سالم و آسیب‌دیده انجام شود.

۴-۱- تحلیل نمونه‌ها با استفاده از ماتریس همبستگی

سیگنال‌های حاصل از آزمایش غیرمخرب اولتراسونیک بر روی نمونه‌ها به طور کلی به دو دسته سیگنال‌های مربوط به ناحیه‌های آسیب دیده و سالم قابل تقسیم هستند. بنابراین می‌توان از دو برجسب صفر نماینده وضعیت آسیب و یک نشانگر حالت سالم را به داده‌های حاصل از آزمایش نسبت داد. بر این اساس می‌توان از الگوریتم‌های یادگیری ماشین برای طبقه‌بندی داده‌ها استفاده نمود تا مناسب‌ترین مدل برای طبقه‌بندی داده‌ها با توجه به ویژگی‌های مختلف موجود در سیگنال‌های مربوط به نمونه‌ها شناسایی شود. فرایند آموزش و ارزیابی مدل‌ها با استفاده از نرم‌افزار Rapid miner انجام شد که یک نرم‌افزار قدرتمند و شناخته‌شده برای پیاده‌سازی و تحلیل مدل‌های داده‌کاوی و یادگیری ماشین است [۲۸]. براساس نتایج حاصل از آزمایش غیرمخرب اولتراسونیک و تحلیل‌های تبدیل موجک گسسته و تبدیل بسته موجک، داده‌هایی با ۲۲ ویژگی مستقل از یکدیگر به دست آمده است که در ۴-۱-جدول ۳ عنوان این مشخصه‌ها ارائه شده‌اند.

¹ Ensemble Learning

² Decision Trees

³ Accuracy

⁴ Gini Index

⁵ Information Gain

⁶ Entropy

⁷ Gain Ratio

جدول ۳: ویژگی و اختصارات استفاده شده در ماتریس همبستگی

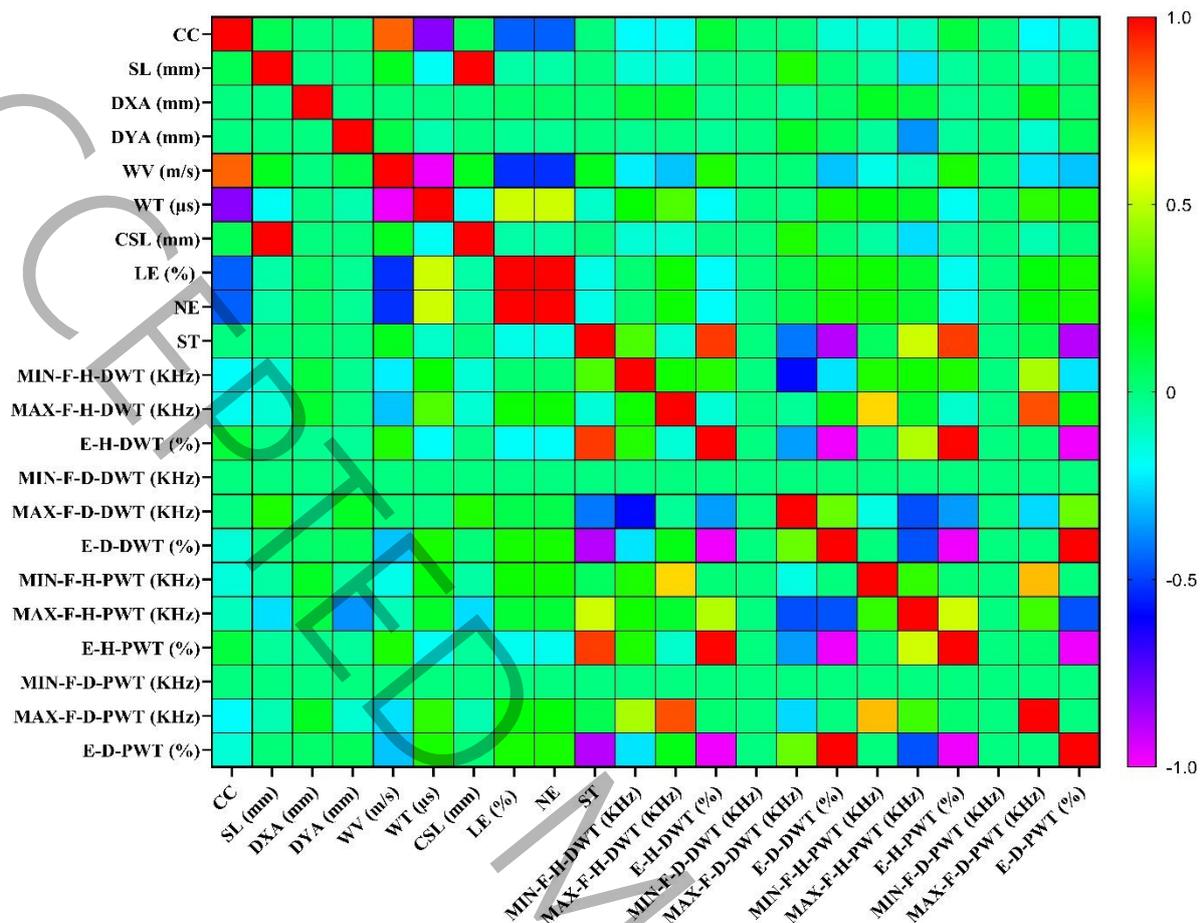
Table 3: Features and abbreviations used in the correlation matrix.

Symbol	Full Text	Persian Equivalent
CC	Cement Content (%)	محتوای سیمان
SL	Sample Length	طول نمونه
DXA	Distance from X-Axis	فاصله از محور X
DYA	Distance from Y-Axis	فاصله از محور Y
WV	Wave Velocity	سرعت موج
WT	Wave Time	زمان موج
CSL	Calculated Sample Length	طول نمونه محاسبه شده
LE	Length Error	خطای طول
NE	Normal Error	خطای نرمال
ST	Sample Type	نوع نمونه
MNFH-DWT	Minimum Frequency in the Healthy Section with Discrete Wavelet Transform	حداقل فرکانس در بخش سالم با تبدیل موجک گسسته
MXFH-DWT	Maximum Frequency in the Healthy Section with Discrete Wavelet Transform	حداکثر فرکانس در بخش سالم با تبدیل موجک گسسته
EHS-DWT	Energy in the Healthy Section of the Sample with Discrete Wavelet Transform	انرژی در بخش سالم نمونه با تبدیل موجک گسسته
MNFD-DWT	Minimum Frequency in the Damaged Section with Discrete Wavelet Transform	حداقل فرکانس در بخش آسیب دیده با تبدیل موجک گسسته
MXFD-DWT	Maximum Frequency in the Damaged Section with Discrete Wavelet Transform	حداکثر فرکانس در بخش آسیب دیده با تبدیل موجک گسسته
EDS-DWT	Energy in the Damaged Section of the Sample with Discrete Wavelet Transform	انرژی در بخش آسیب دیده نمونه با تبدیل موجک گسسته
MNFH-WPT	Minimum Frequency in the Healthy Section with Wavelet Packet Transform	حداقل فرکانس در بخش سالم با تبدیل بسته موجک
MXFH-WPT	Maximum Frequency in the Healthy Section with Wavelet Packet Transform	حداکثر فرکانس در بخش سالم با تبدیل بسته موجک

EHS-WPT	Energy in the Healthy Section of the Sample with Wavelet Packet Transform	انرژی در بخش سالم نمونه با تبدیل بسته موجک
MNFD-WPT	Minimum Frequency in the Damaged Section with Wavelet Packet Transform	حداقل فرکانس در بخش آسیب‌دیده با تبدیل بسته موجک
MXFD-WPT	Maximum Frequency in the Damaged Section with Wavelet Packet Transform	حداکثر فرکانس در بخش آسیب‌دیده با تبدیل بسته موجک
EDS-WPT	Energy in the Damaged Section of the Sample with Wavelet Packet Transform	انرژی در بخش آسیب‌دیده نمونه با تبدیل بسته موجک

در این پژوهش، تعداد ۲۲ ویژگی استخراج شده از سیگنال‌های اولتراسونیک برای مدل‌سازی در نظر گرفته شد. شکل ۶ ماتریس همبستگی حاصل را برای هر ۲۲ ویژگی نشان می‌دهد. با توجه به محدود بودن تعداد ویژگی‌ها و اینکه همبستگی پیرسن تنها روابط خطی را تشخیص می‌دهد، تمامی ویژگی‌ها وارد مدل‌های یادگیری ماشین شدند. علاوه بر این، اهمیت نسبی ویژگی‌ها پس از آموزش مدل‌ها و از طریق معیارهای اهمیت ویژگی^۱ در الگوریتم جنگل تصادفی مورد ارزیابی قرار گرفت تا نقش ویژگی مهم‌تر در تشخیص آسیب مشخص شود.

¹ Feature Importance



شکل ۶: ماتریس همبستگی برای نمونه‌های CC

Fig 6: Correlation matrix for CC samples.

بر اساس ماتریس همبستگی محاسبه‌شده برای ۲۲ ویژگی نمونه‌ها، مقادیری که دارای همبستگی بالا (۰/۸ تا ۱) بودند، مورد بررسی قرار گرفتند. نتایج نشان داد که میزان سیمان مصرفی در نمونه‌های CC همبستگی قابل توجهی (۰/۸۵) با سرعت موج عبوری دارد. این رابطه بیانگر آن است که با افزایش درصد سیمان مصرفی، سرعت موج اولتراسونیک عبوری از نمونه‌ها افزایش می‌یابد. این پدیده را می‌توان با کاهش میزان حفره‌ها و تخلخل نمونه‌ها در اثر افزایش سیمان مصرفی توجیه کرد.

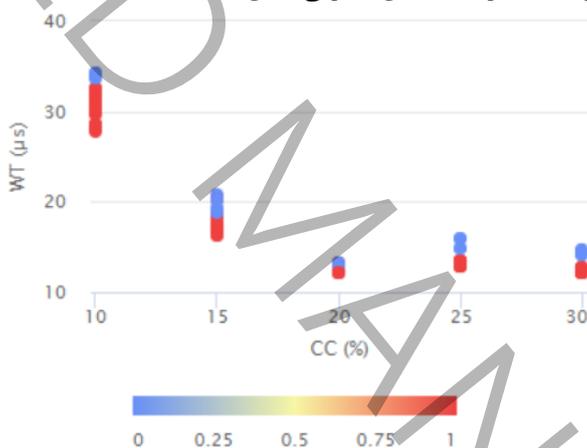
شکل ۷ منحنی تغییرات سرعت موج عبوری از نمونه‌ها را بر اساس درصد سیمان مصرفی نشان می‌دهد. در این منحنی، دو رنگ قرمز (سالم) و آبی (آسیب دیده) به ترتیب نشان‌دهنده آزمایش غیرمخرب اولتراسونیک برای نواحی سالم و آسیب دیده هستند.



شکل ۷: منحنی سرعت عبوری موج از نمونه‌های سالم و آسیب‌دیده بر حسب درصد سیمان مصرفی

Fig 7: The wave propagation velocity curves for healthy and damaged samples versus the percentage of cement content used.

چنانچه در شکل فوق مشاهده می‌شود سرعت بخش سالم (قرمز رنگ) بیش از سرعت بخش آسیب‌دیده (آبی رنگ) است. در شکل ۸ روند تغییرات سرعت زمان عبوری بر حسب درصد سیمان مصرفی نشان داده شده است.



شکل ۸: منحنی مدت زمان عبور موج برای نمونه‌های سالم و آسیب‌دیده بر حسب درصد سیمان مصرفی

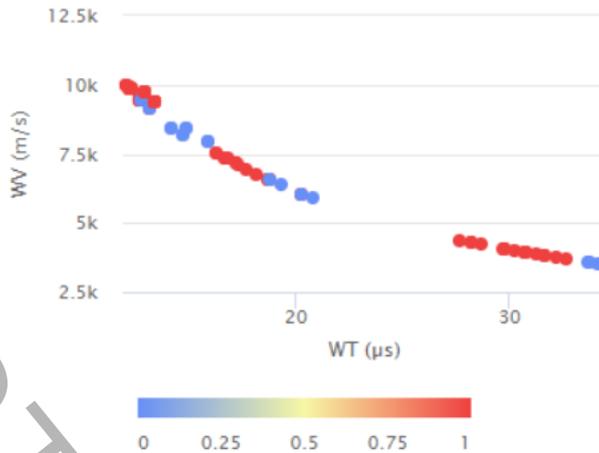
Fig 8: The wave travel time curves for healthy and damaged samples versus the percentage of cement content used.

چنانچه در شکل فوق مشاهده می‌شود با افزایش درصد سیمان مصرفی، مدت زمان عبور موج در بخش سالم و آسیب‌دیده نمونه‌ها کاهش می‌یابد. مدت زمان عبور موج در بخش آسیب‌دیده (آبی) بیش از بخش سالم (قرمز) است. این نتایج نشان‌دهنده آن است که افزایش درصد سیمان مصرفی به بهبود همگنی ساختار نمونه‌های CC منجر شده است، که خود عامل مهمی در ارتقای یکنواختی و کیفیت ساختار داخلی بتن به شمار می‌رود.

قابل توجه است که در هر دو گروه نمونه‌ها، با افزایش درصد سیمان مصرفی، شیب نمودار به‌طور چشم‌گیری کاهش یافته است. با افزایش درصد سیمان مصرفی به دلیل کاهش میزان تخلخل‌های موجود در نمونه‌ها، تأثیر درصد سیمان بر رفتار موج اولتراسونیک عبوری از نمونه‌های سیمانی کمتر می‌شود. به عبارت دیگر، در سطوح بالاتر سیمان مصرفی، ساختار نمونه‌ها به‌مراتب همگن‌تر شده و اثرگذاری تغییرات سیمان بر ویژگی‌های موج عبوری کاهش می‌یابد.

بر این اساس با درصد سیمان مصرفی بالاتر و همگن‌تر شدن نمونه‌ها، موج با سرعت بیشتری و مدت زمان کمتری از فرستنده به گیرنده می‌رسد. این روند افزایشی سرعت موج تا درصد سیمان مصرفی حدود ۲۰ درصد ادامه دارد و پس از آن، با کاهش تخلخل‌ها به دلیل مصرف بیشتر سیمان، سرعت موج به یک مقدار ثابت نزدیک شده و شیب تغییرات کاهش می‌یابد.

در شکل ۹ نیز، تغییرات منحنی سرعت موج عبوری بر حسب مدت زمان عبور موج نشان داده شده است.

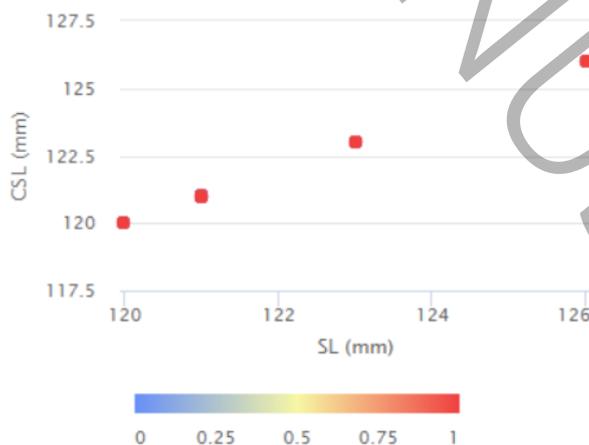


شکل ۹: منحنی سرعت موج عبوری بر حسب مدت زمان عبور موج برای نمونه‌های سالم و آسیب‌دیده

Fig 9: The wave velocity curve versus wave travel time for healthy and damaged samples.

چنانچه در شکل فوق مشاهده می‌شود، در سرعت موج‌های زیاد که در نمونه‌هایی با درصد مصرف سیمان زیاد اتفاق می‌افتد، سرعت استخراجی برای یک نوع نمونه‌ها تقریباً برابر بوده‌اند که این امر با توجه به آستانه اندازه‌گیری دستگاه و نیز کاهش تاثیر درصد سیمان در سرعت عبوری برای نمونه‌های بالای ۲۰ درصد امری منطقی می‌باشد. همچنین با افزایش مدت زمان عبور موج، سرعت موج عبوری روندی کاهشی با شیب منفی خواهد داشت. شدت همبستگی بین دو پارامتر سرعت موج عبوری و مدت زمان عبور موج، با مقدار -0.97 در ماتریس همبستگی ارزیابی شده است.

علاوه بر سه پارامتر درصد سیمان مصرفی، سرعت موج عبوری و مدت زمان عبور موج، تحلیل ماتریس همبستگی نشان داد که رابطه مستقیم و معنادار بسیار قوی‌ای بین طول واقعی نمونه و طول محاسبه‌شده با دستگاه اولتراسونیک برای نمونه‌ها وجود دارد. به گونه‌ای که ضریب همبستگی آن برابر با 0.999 است. این نتیجه نشانگر دقت بالای دستگاه در اندازه‌گیری فاصله بین پروب ارسال‌کننده و دریافت‌کننده است. شکل ۱۰ طول محاسبه‌شده با دستگاه اولتراسونیک را در مقایسه با طول واقعی برای نمونه‌های مختلف در شرایط سالم و آسیب‌دیده نشان داده است.

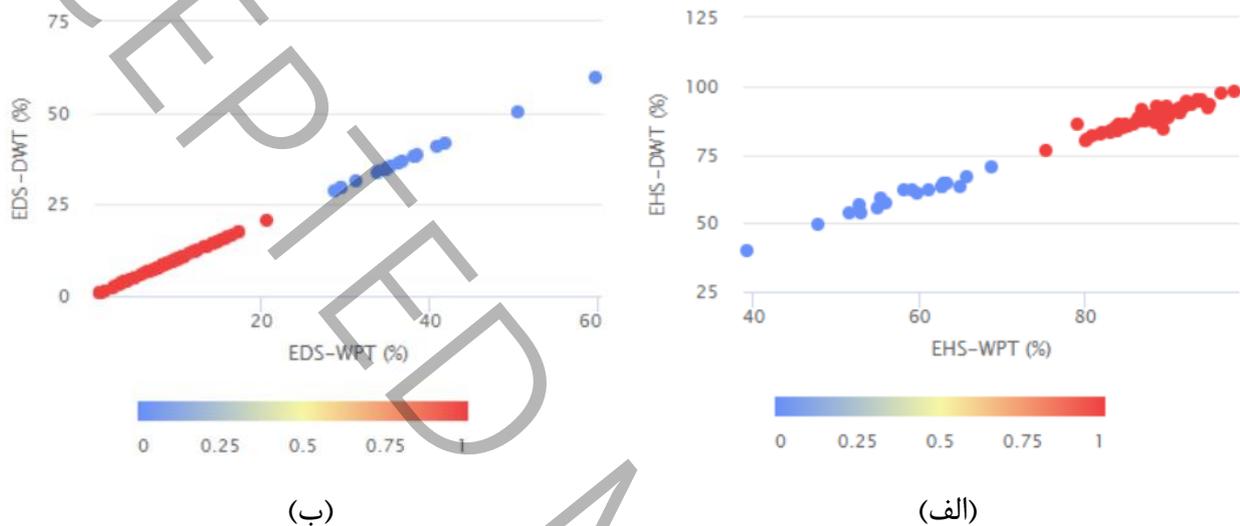


شکل ۱۰: مقایسه طول محاسبه شده با دستگاه اولتراسونیک نسبت به طول واقعی نمونه‌ها

Fig 10: Comparison of the length calculated by the ultrasonic device with the actual length of the samples.

مقادیر خطای اندازه‌گیری نیز برای نمونه‌های مختلف محاسبه شد و به ترتیب برابر بودند با: ۰/۲۴۳ درصد برای نمونه‌های ۱۲۰ میلی متری، ۰/۵۲۸ درصد برای نمونه‌های ۱۲۱ میلی متری، ۰/۴۶۱ درصد برای نمونه‌های ۱۲۳ میلی متری، و ۰/۳۱۲ درصد برای نمونه‌های ۱۲۶ میلی متری. نتایج فوق نشانگر دقت بالای دستگاه اولتراسونیک است.

با توجه به اینکه یکی از اهداف اصلی این پژوهش ارزیابی کارایی انرژی نسبی سیگنال‌های اولتراسونیک به‌عنوان روشی مناسب برای تشخیص نواحی سالم و آسیب دیده در نمونه‌های CC است. بدین منظور، در شکل ۱۱ (الف) انرژی‌های نسبی در فرکانس‌های آسیب دیده در نواحی سالم و آسیب دیده با تبدیلهای موجک بسته و گسسته مقایسه شدند. به علاوه در شکل ۱۱ (ب) نیز انرژی‌های نسبی در فرکانس‌های سالم در نواحی سالم و آسیب دیده با تبدیلهای موجک بسته و گسسته مقایسه شدند.

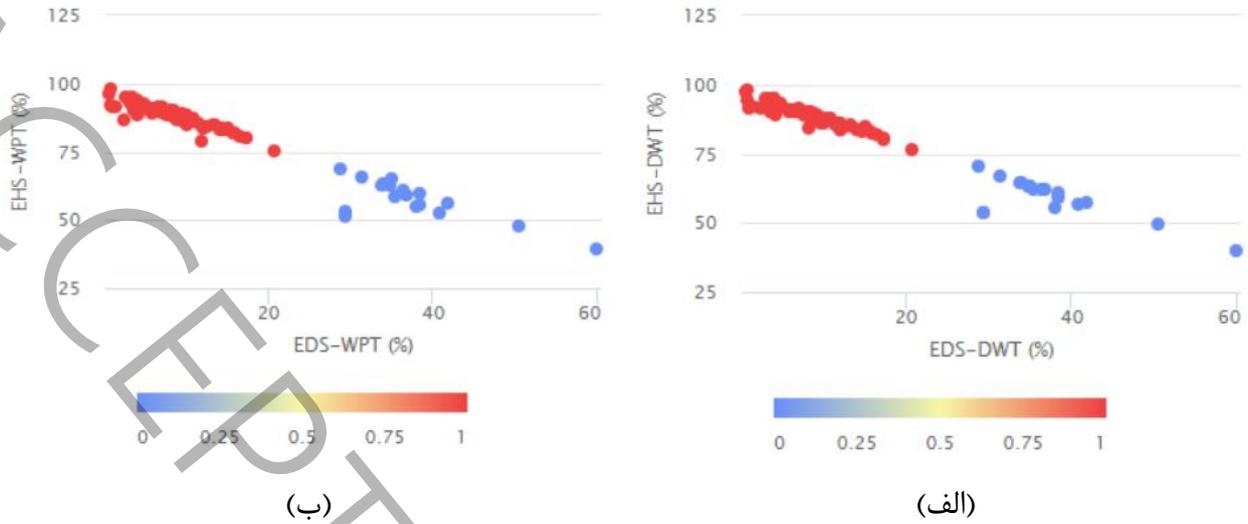


شکل ۱۱: مقایسه تبدیلهای موجک گسسته و بسته در تحلیل میزان انرژی بازه فرکانس مربوط به نمونه‌های بتنی (الف) سالم و (ب) آسیب دیده

Fig 11: Comparison of discrete and continuous wavelet transforms in analyzing the energy level of the frequency band related to concrete samples (a) healthy and (b) damaged.

نتایج نشان می‌دهند که هر دو روش (تبدیل موجک گسسته و تبدیل بسته موجک) تطابق مناسبی در تمایز بین نواحی سالم و آسیب دیده ارائه می‌دهند. به‌طور خاص، انرژی در بازه فرکانسی آسیب (بازه فرکانسی کمتر از بازه فرکانسی سالم) در نواحی آسیب دیده به‌طور چشمگیری بیشتر از نواحی سالم است (شکل ۱۱ (الف)). در مقابل، انرژی سیگنال در بازه فرکانسی سالم به‌طور قابل توجهی در نواحی سالم بیشتر از نمونه‌های آسیب دیده بیشتر است (شکل ۱۱ (ب)). تمایز در انرژی فرکانسی در نواحی سالم و آسیب دیده، قابلیت این دو روش تبدیل موجک را در تشخیص و تفکیک نواحی آسیب دیده و سالم را مشخص می‌کند.

شکل ۱۲ نسبت انرژی مربوط به بازه فرکانسی سالم به انرژی مربوط به بازه فرکانسی آسیب را در هر یک از روش‌های تحلیل فرکانسی تبدیل موجک گسسته و بسته نمایش می‌دهد. نتایج تحلیل اهمیت ویژگی‌ها نشان داد که ویژگی‌های مرتبط با انرژی سیگنال، فرکانس غالب و ویژگی‌های استخراج شده از تبدیل موجک بیشترین تأثیر را در تشخیص آسیب داشته‌اند.



شکل ۱۲: میزان انرژی مربوط به بازه فرکانسی سالم نسبت به میزان انرژی مربوط به بازه فرکانسی آسیب: (الف) تبدیل موجک گسسته و (ب) تبدیل بسته موجک

Fig 12: The energy level of the healthy frequency band compared to the damaged frequency band: (a) discrete wavelet transform and (b) continuous wavelet transform.

۲-۴- تحلیل نمونه‌ها با استفاده از یادگیری ماشین بر اساس طبقه‌بندی

در این پژوهش، سه الگوریتم یادگیری ماشین شامل K نزدیک‌ترین همسایگان، بیز ساده و جنگل تصادفی برای طبقه‌بندی نمونه‌های ملات‌سیمانی به دو گروه سالم و آسیب‌دیده، بر اساس ویژگی‌های استخراج‌شده از داده‌های آزمایش اولتراسونیک، مورد بررسی قرار گرفتند.

۲-۴-۱- K نزدیک‌ترین همسایگان

در این تحلیل، از فاصله اقلیدسی به‌عنوان معیار محاسبه نزدیکی بین داده‌ها استفاده شده است. مقدار بهینه K از طریق روش اعتبارسنجی متقاطع^۱ تعیین گردید تا بهترین عملکرد ممکن حاصل شود. به منظور تعیین مقدار بهینه K ، از تحلیل پارامترهای آماری نظیر دقت، خطا، صحت و یادآوری استفاده شد.

انتخاب مقدار بسیار کوچک برای K می‌تواند حساسیت مدل را به نویز افزایش دهد و منجر به پدیده $Overfitting$ (حالتی که مدل به‌طور بیش از حد به داده‌های آموزشی منطبق می‌شود و در تعمیم‌دهی ضعیف عمل می‌کند) شود. در این شرایط، مدل با داده‌های آموزشی کاملاً منطبق است اما ممکن است در پیش‌بینی داده‌های جدید عملکرد مناسبی نداشته باشد، و دقت و صحت به‌طور غیرواقعی افزایش یابد. در مقابل، انتخاب مقدار بسیار بزرگ برای K می‌تواند باعث $Underfitting$ (حالتی که مدل قادر به یادگیری الگوهای اصلی داده نیست و عملکرد پایینی دارد) شود، به این معنا که مدل قادر به یادگیری روابط پیچیده در داده‌ها نیست و عملکرد ضعیفی در تعمیم‌دهی به داده‌های جدید دارد.

برای جلوگیری از بروز $Underfitting$ و $Overfitting$ در این پژوهش، مقادیر K از ۱ تا ۱۰ بررسی شدند. نتایج حاصل از این تحلیل در جدول ۴ ارائه شده است. این رویکرد به شناسایی مقدار بهینه K کمک کرد تا مدل از تعادل مناسبی بین دقت و تعمیم‌پذیری برخوردار باشد.

¹ Cross Validation

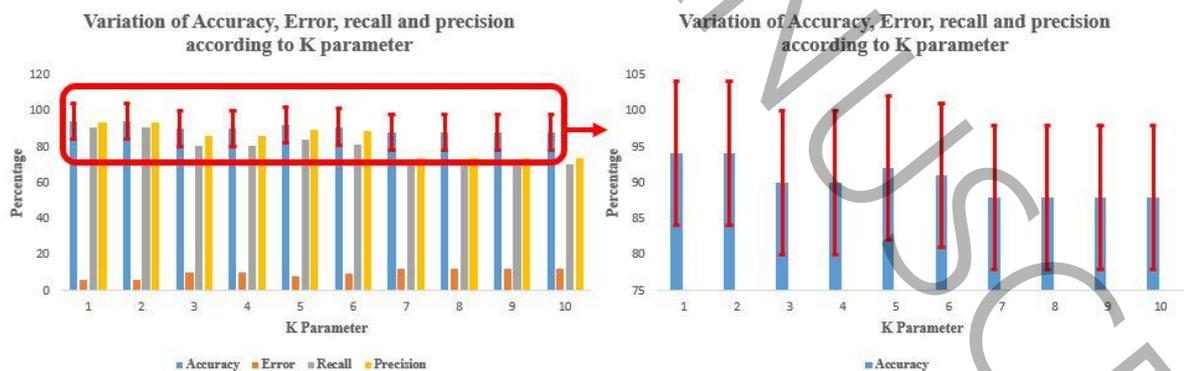
جدول ۴: بررسی مقادیر K در الگوریتم طبقه‌بندی KNN

Table 4: Evaluation of K values in the KNN classification algorithm.

K	Accuracy	Error	Recall	Precision
1	94.00 ± 8.43	6.00 ± 8.43	90.62 ± 12.93	93.26 ± 11.59
2	94.00 ± 8.43	6.00 ± 8.43	90.62 ± 12.93	93.26 ± 11.59
3	90.00 ± 8.16	10.00 ± 8.16	80.62 ± 16.26	86.15 ± 19.46
4	90.00 ± 8.16	10.00 ± 8.16	80.62 ± 16.26	86.15 ± 19.46
5	92.00 ± 7.89	8.00 ± 7.89	83.75 ± 16.72	89.28 ± 18.85
6	91.00 ± 7.38	9.00 ± 7.38	81.25 ± 15.87	88.72 ± 18.58
7	88.00 ± 7.89	12.00 ± 7.89	70.00 ± 19.72	73.78 ± 29.15
8	88.00 ± 7.89	12.00 ± 7.89	70.00 ± 19.72	73.78 ± 29.15
9	88.00 ± 7.89	12.00 ± 7.89	70.00 ± 19.72	73.78 ± 29.15
10	88.00 ± 7.89	12.00 ± 7.89	70.00 ± 19.72	73.78 ± 29.15

مطابق جدول ۴، مقدار بهینه K برابر با ۵ تعیین شده است. این مقدار به دلیل دستیابی به تعادل مناسب میان دقت و تعمیم‌پذیری، بهترین گزینه است؛ چرا که در این حالت الگوریتم نه دچار Overfitting شده و نه به دلیل Underfitting کاهش قابل توجهی در دقت داشته است.

شکل ۱۳، منحنی تغییرات چهار پارامتر آماری (دقت، صحت، یادآوری، و خطا) را بر حسب مقادیر مختلف K نشان می‌دهد. همان‌طور که مشاهده می‌شود، برای مقادیر $K = 1$ و $K = 2$ ، الگوریتم به دلیل حساسیت بالا به نویز دچار Overfitting شده است. در مقابل، در مقادیر بالاتر از ۲، رفتار پارامترهای آماری به صورت سهمی بوده و قله بیشینه آن در $K = 5$ مشاهده می‌شود. این یافته‌ها نشان می‌دهند که انتخاب مقدار $K = 5$ علاوه بر جلوگیری از پدیده Overfitting و Underfitting، به الگوریتم اجازه داده است که عملکرد بهینه‌ای در طبقه‌بندی داده‌ها داشته باشد.



شکل ۱۳: تغییرات پارامترهای آماری در اثر تغییر مقدار K در الگوریتم KNN

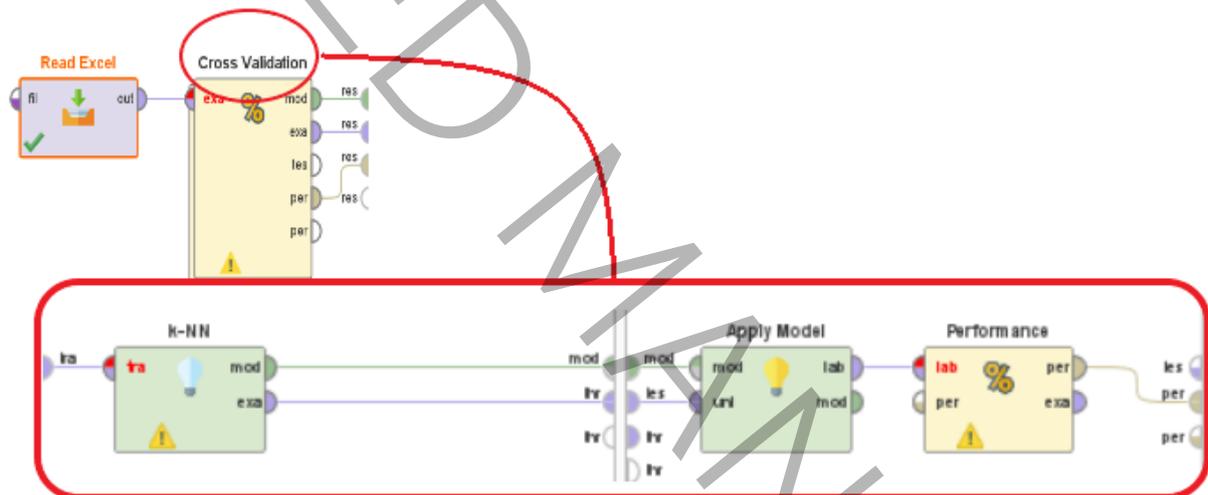
Fig 13: Changes in statistical parameters due to variation of K value in the KNN algorithm.

در این پژوهش، با توجه به ضرایب پایین ماتریس همبستگی، ارتباط قوی میان ویژگی‌ها مشاهده نشده است. بنابراین، استفاده از مدل بیز ساده برای دسته‌بندی داده‌ها منطقی به نظر می‌رسد، چرا که این مدل فرض می‌کند ویژگی‌ها مستقل از یکدیگر هستند و این فرض با ویژگی‌های داده‌های این مطالعه هم‌خوانی دارد.

۴-۲-۳- جنگل تصادفی

روش جنگل تصادفی توانایی مدیریت تعاملات غیرخطی و پیچیدگی ویژگی‌ها را به خوبی دارد، به همین دلیل برای مجموعه داده‌های پیچیده و با تعداد زیادی ویژگی مناسب است. نکته مهم دیگری که باید به آن اشاره کرد، استفاده از رأی اکثریت یا میانگین پیش‌بینی‌ها از تمامی درخت‌ها برای به دست آوردن نتیجه نهایی است. این ویژگی باعث می‌شود که مدل Overfitting (تناسب بیش از حد با داده‌های آموزشی) به حداقل برسد و دقت و کنترل پیش‌بینی افزایش یابد. بنابراین، مدل جنگل تصادفی با استفاده از درخت‌های تصمیم متعدد، معمولاً دقیق‌تر از یک مدل تک‌درختی^۱ عمل می‌کند و توانایی آن در پیش‌بینی داده‌های پیچیده را بهبود می‌بخشد.

در این پژوهش، مدل طبقه‌بندی نمونه‌های سالم و آسیب دیده با استفاده از الگوریتم‌های مختلف در نرم‌افزار Rapid miner پیاده‌سازی شد و نتایج به‌دست‌آمده در شکل ۱۴ نشان داده شده است.



شکل ۱۴: مدل طبقه‌بندی اجرا شده در نرم‌افزار Rapid miner با الگوریتم جنگل تصادفی

Fig 14: The classification model implemented in RapidMiner software using the Random Forest algorithm.

لازم به ذکر است که برای دو الگوریتم دیگر طبقه‌بندی نیز مدل مشابهی با آنچه در شکل ۱۴ نشان داده شده است، اجرا شده است و تنها الگوریتم مورد استفاده برای آموزش داده‌ها تغییر کرده است. در نهایت، نتایج حاصل از طبقه‌بندی هر یک از این الگوریتم‌ها با استفاده از چهار پارامتر آماری مورد بررسی قرار گرفتند که نتایج در جدول ۵ آمده است.

¹ Single Decision Tree

جدول ۵: نتایج حاصل از طبقه‌بندی هر یک از این الگوریتم‌های یادگیری ماشین

Table 5: The classification results obtained from each of these machine learning algorithms.

Algorithm	Accuracy	Error	Recall	Precision	True0	True1	
KNN	92.00 ± 7.89	8.00 ± 7.89	83.75 ± 16.72	89.28 ± 18.85	6	78	Pred1
					14	2	Pred0
Naïve Bayse	99.00 ± 3.16	1.00 ± 7.89	97.50 ± 7.91	99.44 ± 1.76	1	80	Pred1
					19	0	Pred0
RF (Accuracy)	94.00 ± 8.43	5.00 ± 7.07	87.50 ± 17.68	92.33 ± 18.57	6	80	Pred1
					14	0	Pred0
RF (Gain ratio)	93.00 ± 8.23	6.00 ± 9.66	85.00 ± 24.15	82.00 ± 28.98	7	80	Pred1
					13	0	Pred0
RF (Inf gain)	90.00 ± 9.43	8.00 ± 9.19	83.75 ± 21.08	82.65 ± 24.67	7	77	Pred1
					13	3	Pred0
RF (gini index)	96.00 ± 5.16	5.00 ± 7.07	87.50 ± 17.68	92.33 ± 18.57	4	80	Pred1
					16	0	Pred0

در جدول ۵، ردیف‌ها نشان‌دهنده برچسب واقعی و ستون‌ها بیانگر برچسب پیش‌بینی شده توسط مدل هستند. در این پژوهش، کلاس‌بندی به صورت ۱ به‌عنوان نمونه‌های سالم و ۰ به‌عنوان نمونه‌های آسیب‌دیده تعریف شده است. بر این اساس، مقدار true1-predicted1 تعداد نشان‌دهنده تعداد نمونه‌های سالمی است که مدل به‌درستی سالم تشخیص داده است، و مقدار true1-predicted0 تعداد نمونه‌های سالمی را نشان می‌دهد که مدل به اشتباه آسیب‌دیده برچسب‌گذاری کرده است. همچنین مقدار true0-predicted1 بیانگر تعداد نمونه‌های آسیب‌دیده‌ای است که مدل به اشتباه سالم تشخیص داده است، در حالی که true0-predicted0 تعداد نمونه‌های آسیب‌دیده تشخیص‌داده‌شده به‌صورت صحیح توسط مدل را نشان می‌دهد. این ساختار امکان محاسبه شاخص‌هایی همچون دقت، حساسیت، ویژگی‌پذیری و صحت کلی مدل را فراهم می‌سازد و در آن true1 معرف کلاس واقعی سالم و true0 معرف کلاس واقعی آسیب‌دیده است.

چنانچه در جدول ۵ مشخص است، مدل طبقه‌بندی بیز ساده در تمامی چهار پارامتر آماری عملکرد بهتری از خود نشان داده است. این نتیجه‌گیری باتوجه به کم بودن ضرایب ماتریس همبستگی برای تعداد زیادی از ویژگی‌های ورودی، امری طبیعی به نظر می‌آید. می‌توان این‌گونه برداشت کرد که بیشتر ویژگی‌های ماتریس همبستگی به‌طور مستقل از یکدیگر عمل می‌کنند. با توجه به فرض استقلال ویژگی‌ها در الگوریتم بیز ساده، منطقی است که این مدل طبقه‌بندی عملکرد بهتری نسبت به سایر الگوریتم‌ها داشته باشد.

نکته قابل توجه دیگری که در جدول ۵ مشاهده می‌شود، بالا بودن میزان عدم قطعیت پارامترهای آماری یادآوری و صحت در هر چهار الگوریتم جنگل تصادفی است. از این موضوع می‌توان این‌گونه برداشت کرد که با توجه به کم بودن میزان همبستگی میان ویژگی‌های ورودی، درختان تصمیم ایجاد شده در الگوریتم جنگل تصادفی نتوانسته‌اند به لایه‌های عمیق پیش‌روی کنند. عدم وابستگی داده‌ها باعث افزایش تعداد شاخه‌های هر درخت تصمیم و کاهش طول شاخه‌ها شده است. همین امر موجب افزایش عدم قطعیت در پارامترهای یادآوری و صحت در مدل‌های طبقه‌بندی مبتنی بر جنگل تصادفی شده است. شکل ۱۵ درخت تصمیم مورد استفاده در مدل طبقه‌بندی بر مبنای جنگل تصادفی را نشان می‌دهد.



شکل ۱۵ : درخت تصمیم ایجاد شده در الگوریتم جنگل تصادفی مدل طبقه‌بندی

Fig 15: The decision tree generated in the Random Forest classification model.

چنانچه مشاهده می‌شود، جنگل تصادفی تنها توانسته است درخت تصمیم را تا یک سطح گسترش دهد. این امر به دلیل عدم همبستگی بیشتر نمونه‌ها می‌باشد که موجب محدودیت در عمق درختان تصمیم و در نتیجه کاهش دقت و یادآوری شده است.

۵- نتیجه‌گیری

این پژوهش نشان داد که استفاده از روش غیرمخرب اولتراسونیک، در ترکیب با الگوریتم‌های یادگیری ماشین، رویکردی کارآمد برای طبقه‌بندی آسیب در نمونه‌های ملات سیمانی است. با بهره‌گیری از الگوریتم‌های یادگیری ماشین KNN، بیز ساده و جنگل تصادفی، داده‌ها طبقه‌بندی شدند که در این میان الگوریتم بیز ساده با دقت $99/00 \pm 3/16$ درصد و یادآوری $97/5 \pm 7/91$ درصد، بهترین عملکرد را داشت. نوآوری این تحقیق در ارائه یک چارچوب یکپارچه است که از الگوریتم‌های یادگیری ماشین برای طبقه‌بندی دقیق آسیب به کار می‌گیرد؛ رویکردی که تاکنون به صورت جامع برای ملات سیمانی با محتوای سیمان متغیر مورد استفاده قرار نگرفته بود. این روش به دلیل غیرمخرب بودن و دقت بالا برای پایش سلامت سازه‌ها و کنترل‌های دوره‌ای مناسب است. با گردآوری یک بانک داده کامل از ویژگی‌های موجکی استخراج‌شده و آموزش مدل‌های یادگیری ماشین، می‌توان سیگنال‌های جدید را مستقیماً توسط مدل آموزش‌دیده طبقه‌بندی کرد؛ بنابراین نیازی به تکرار تحلیل موجک روی نمونه‌های بعدی نیست و شناسایی آسیب تنها با یک آزمایش اولتراسونیک امکان‌پذیر می‌شود. پیشنهاد می‌شود در مطالعات آینده، این چارچوب در مقیاس سازه‌های واقعی و با در نظر گرفتن اثر شرایط محیطی، رطوبت و دما ارزیابی شود. همچنین ترکیب آن با روش‌های غیرمخرب دیگر مانند تصویربرداری حرارتی یا رادار نفوذی زمین بررسی گردد تا دقت و جامعیت پایش آسیب‌ها افزایش یابد.

مراجع

- [1] J. Chakraborty and M. Stolinski, "Signal-level fusion approach for embedded ultrasonic sensors in damage detection of real RC structures," *Mathematics*, vol. 10, no. 5, p. 724, 2022.
- [2] F. A. J. Mohamad et al., "NDT-Defect Detection on Concrete using Ultrasonic: A Review," *J. Tomogr. Syst. Sensors Appl.*, 2021.
- [3] G. Karaiskos, A. Deraemaeker, D. G. Aggelis, and D. Van Hemelrijck, "Monitoring of concrete structures using the ultrasonic pulse velocity method," *Smart Mater. Struct.*, vol. 24, no. 11, p. 113001, 2015, doi: 10.1088/0964-1726/24/11/113001.
- [4] L. Ge et al., "High-resolution ultrasonic imaging technology for the damage of concrete structures based on total focusing method," *Comput. Electr. Eng.*, vol. 105, p. 108526, 2023.
- [5] R. Jones, "The ultrasonic testing of concrete," *Ultrasonics*, vol. 1, no. 2, pp. 78–82, 1963, doi: 10.1016/0041-624X(63)90058-1.
- [6] I. Ivanchev and V. Slavchev, "About the possible limitations in the usage of the non-destructive ultrasonic pulse velocity method for assessment of cracks in reinforced concrete structures, subjected to direct environmental exposure," *Buildings*, vol. 9, no. 9, p. 202, 2019, doi: 10.3390/buildings9090202.
- [7] A. Lorenzi, J. Campagnolo, and L. C. P. Silva Filho, "Application of artificial neural network for interpreting

- ultrasonic readings of concrete,” *Int. J. Mater. Prod. Technol.*, vol. 26, no. 1–2, pp. 57–70, 2006, doi: 10.1504/IJMPT.2006.008980.
- [8] R. Jones, “Non-destructive testing of concrete,” (No Title), 1962.
- [9] J. Xu and H. Wei, “Ultrasonic testing analysis of concrete structure based on S transform,” *Shock Vib.*, vol. 2019, pp. 1–9, 2019.
- [10] K. Komlos, S. Popovics, T. Nürnbergerová, B. Babal, and J. S. Popovics, “Ultrasonic pulse velocity test of concrete properties as specified in various standards,” *Cem. Concr. Compos.*, vol. 18, no. 5, pp. 357–364, 1996.
- [11] J. Wang, M. Chen, N. Zhu, and H. Liu, “A damage zone detection method in concrete hydraulic structures based on multi-frequency ultrasonic characteristics,” *Sci. Rep.*, vol. 15, no. 1, p. 11173, 2025.
- [12] L. Wang, S. Yi, Y. Yu, C. Gao, and B. Samali, “Automated ultrasonic-based diagnosis of concrete compressive damage amidst temperature variations utilizing deep learning,” *Mech. Syst. Signal Process.*, vol. 221, p. 111719, 2024.
- [13] T. Hu, J. Zhao, R. Zheng, P. Wang, X. Li, and Q. Zhang, “Ultrasonic based concrete defects identification via wavelet packet transform and GA-BP neural network,” *PeerJ Comput. Sci.*, vol. 7, p. e635, 2021.
- [14] M. D. E. Candelaria, N. M. M. Chua, and S.-H. Kee, “Evaluation of heat-induced damage in concrete using machine learning of ultrasonic pulse waves,” *Materials (Basel)*, vol. 15, no. 22, p. 7914, 2022.
- [15] S. Grabke, F. Clauß, K.-U. Bletzinger, M. A. Ahrens, P. Mark, and R. Wüchner, “Damage detection at a reinforced concrete specimen with coda wave interferometry,” *Materials (Basel)*, vol. 14, no. 17, p. 5013, 2021.
- [16] J. A. Mukhti, K. P. V Robles, K.-H. Lee, and S.-H. Kee, “Evaluation of early concrete damage caused by chloride-induced steel corrosion using a deep learning approach based on RNN for ultrasonic pulse waves,” *Materials (Basel)*, vol. 16, no. 9, p. 3502, 2023.
- [17] A. Faghihi and H. Molladavoodi, “Damage detection in cement mortar samples using wavelet transform analysis of non-destructive ultrasonic test results,” 2025.
- [18] M. N. A. B. G. Taj, N. Alruwais, H. M. Alshahrani, J. Vijayalakshmi, N. Shanmugapriya, and S. Jayaprakash, “Precision crack analysis in concrete structures using CNN, SVM, and KNN: a machine learning approach,” *Matéria (Rio Janeiro)*, vol. 29, p. e20240551, 2024.
- [19] D. W. Dunker and J. Farny, “Your Pipe Lining Has a New Cement—ASTM C150 Compared to ASTM C595,” in *Pipelines 2024*, pp. 312–320.
- [20] S. A. Proceq, “Operating Instructions Pundit Lab,” *Pundit Lab+ Ultrason. Instrum.*, vol. 32, 2017.
- [21] N. S. M. Ravi Kumar, T. Barkavi, and C. Natarajan, “Structural health monitoring: detection of concrete flaws using ultrasonic pulse velocity,” *J. Build. Pathol. Rehabil.*, vol. 3, pp. 1–9, 2018.
- [22] S. Hannachi and M. N. Guetteche, “Review of the ultrasonic pulse velocity Evaluating concrete compressive strength on site,” in *Proceedings of Scientific Cooperation International Workshops on Engineering Branches, Istanbul, Turkey, 2014*, no. August, pp. 103–112.
- [23] I. H. Sarker, “Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci* 2: 160.” ed, 2021.
- [24] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [25] K. Fernandes, P. Vinagre, and P. Cortez, “A proactive intelligent decision support system for predicting the popularity of online news,” in *Progress in Artificial Intelligence: 17th Portuguese Conference on Artificial Intelligence, EPIA 2015, Coimbra, Portugal, September 8-11, 2015. Proceedings 17, 2015*, pp. 535–546.
- [26] I. H. Sarker, A. S. M. Kayes, and P. Watters, “Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage,” *J. Big Data*, vol. 6, no. 1, pp. 1–28, 2019.
- [27] B. Mahesh, “Machine learning algorithms-a review,” *Int. J. Sci. Res. (IJSR)*. [Internet], vol. 9, no. 1, pp. 381–386, 2020.
- [28] M. Hofmann and R. Klinkenberg, *RapidMiner: Data mining use cases and business analytics applications*. CRC Press, 2016.