

تخمین زمان واخنش کلاس درس با استفاده از شبکه عصبی کانولوشن

معصومه شفیعیان
دکترای تخصصی
استادیار دانشگاه صدا و سیما
shafieian@iribu.ac.ir

سلمان نوری زهاب*
کارشناسی ارشد مهندسی صدا
دانشگاه صدا و سیما
salmanzhab1@gmail.com

تاریخ پذیرش: ۱۴۰۰/۰۶/۱۱

تاریخ دریافت: ۱۴۰۰/۰۵/۲۷

چکیده

کلاس درس به‌عنوان یکی از مهم‌ترین محیط‌های آموزشی نقش عمده‌ای در یادگیری و پیشرفت تحصیلی دانش‌آموزان دارد. زمان واخنش به‌عنوان یکی از پارامترهای آکوستیکی در داخل اتاق‌ها، تأثیر به‌سزایی در کیفیت صدا دارد. عدم کارآیی مناسب فرمول‌های کلاسیک مانند سابین، باعث شد که در این مقاله به بررسی استفاده از روش‌های یادگیری ماشین به‌عنوان یک روش جایگزین برای پیش‌بینی زمان واخنش محیط پرداخته شود. در این پژوهش ابتدا با استفاده از روش‌های مبتنی بر آکوستیک هندسی و با استفاده از نرم‌افزار اودئون به جمع‌آوری مجموعه داده‌گان مورد نیاز در بسامدهای ۵۰۰ و ۲۰۰۰ هرتز پرداخته می‌شود. در این مجموعه داده‌گان از ۴ کلاس درس با فضایی مستطیل شکل، همراه با المان‌هایی مانند میز و صندلی و پنجره و در، استفاده شد. پس از آن به‌منظور ارائه یک سیستم مبتنی بر یادگیری ماشین از شبکه عصبی کانولوشنی استفاده شده است. این مدل ویژگی‌های محیط را در نظر می‌گیرد و در نهایت مقادیر زمان واخنش را به‌عنوان تابعی از بسامد تولید می‌کند. در این پژوهش با استفاده از شبکه عصبی کانولوشنی برای بسامد ۵۰۰ هرتز به ضریب تعیین ۹۳ درصد و برای بسامد ۲۰۰۰ هرتز، ضریب تعیین ۹۵ درصد حاصل شد.

واژگان کلیدی: آکوستیک کلاس درس، شبکه عصبی کانولوشن^۱، ردیابی پرتو، زمان واخنش^۲

۱. مقدمه

تجربی، به رابط‌های برای زمان واخنش رسید که نشان می‌داد زمان واخنش متأثر از حجم محیط و ضریب جذب است که در معادله ۱ بیان می‌شود:

$$T_{\text{sabine}} = \frac{0.16 * V}{S * \bar{\alpha}_{SAB}} \quad (1)$$

$$\bar{\alpha}_{SAB} = \frac{1}{S} \sum_i s_i * \alpha_i \quad (2)$$

در محیط‌های آموزشی بخش عمده‌ای از فرایند آموزش از طریق درست شنیدن اتفاق می‌افتد. تمرکز و یادگیری به محیط بهینه صوتی نیاز دارد، و شرایط آکوستیکی نامناسب باعث اختلال در امر یادگیری دانش‌آموزان می‌شود [۱]. یکی از پارامترهای مهم به‌منظور توصیف ویژگی‌های صوتی محیط، زمان واخنش است. سابین^۳ اولین نفری بود که به پیش‌بینی زمان واخنش پرداخت و با توجه به تحقیقات

محیطی به‌عنوان ورودی‌های این روش عمل می‌کند و مقادیر زمان واخنش به‌عنوان خروجی به‌دست می‌آید.

۲. ادبیات و پیشینه‌ی پژوهش

در این بخش به ارائه مطالب کلی در مورد زمان واخنش و روش ردیابی پرتو اشاره می‌شود و همچنین تحقیقات گذشته در حوزه موضوع پژوهش مورد بررسی قرار می‌گیرد.

۲-۱. زمان واخنش

یکی از پارامترهای مهم در زمینه بررسی و تحلیل محیط‌های آکوستیکی زمان واخنش است. زمان واخنش پارامتری وابسته به بسامد است. مدت زمان لازم برای کم‌رنگ شدن صدا یا تحلیل رفتن صدا در یک فضای بسته را زمان واخنش می‌گویند و در تعریف به معنای مدت زمانی است که بعد از خاموش شدن منبع صدا، طول می‌کشد تا تراز فشار صوت به میزان 60dB کاهش پیدا کند. این پارامتر را با نماد T_{60} نشان می‌دهند.

۲-۲. روش‌های مبتنی بر آکوستیک هندسی

در آکوستیک هندسی فرض ما این است که موج به‌صورت پرتوهایی در راستای مستقیم پخش می‌شود. فرضیات مهم این روش این است که جبهه‌های موج را با پرتوهایی که عمود بر جبهه‌های موج هستند، می‌توان نمایش داد. این فرض در بسامدهای بالا جایی که طول موج صدا در مقایسه با ابعاد سطح و ابعاد کلی فضا کوچک باشد معتبر است اما این فرض در بسامدهای پایین دچار خطاهای در تقریب می‌شود [۳]. اصول آکوستیک هندسی بسیار کلی است، و این روش‌های مدل‌سازی، به‌ویژه ردیابی پرتو^۴ و رادیوزیتی^۵، در زمینه‌های دیگری مانند اپتیک و آکوستیک زیر آب استفاده می‌شود. با این حال، این اصول بیشتر در گرافیک رایانه‌ای که در آن مسأله، شبیه‌سازی رفتار نور در محیط است، استفاده می‌شود. در آکوستیک هندسی چون فرض ما، فرض بسامد بالا بودن صوت نسبت به جزئیات هندسه

که در این روابط، V حجم اتاق، S مساحت کل سطوح و α_i و S_i که به‌ترتیب مساحت و ضریب جذب سطح شماره i ام محیط هستند.

همان‌طور که معلوم است این معادلات کلاسیک، فرمول‌های نسبتاً ساده‌ای هستند و مهم‌ترین فرض این معادلات این است که میدان صوتی در هر محیط محصور به‌صورت کاملاً پخش‌ناشد و اگر میدان صوتی یک محیط پخش‌ناشد زمان واخنش در یک محیط نقطه به نقطه متفاوت خواهد بود و باعث می‌شود که این فرمول‌ها با مقادیر اندازه‌گیری شده اختلاف داشته باشند. از این رو به‌نظر می‌رسد که استفاده از معادلات کلاسیک یا هر رابطه دیگر که چنین فرضی را برای پیش‌بینی زمان واخنش داشته باشد، باید توأم با خطای زیادی باشد [۲]. در این راستا، یکی از مهم‌ترین موضوعات، استفاده از روش‌های جایگزین است که می‌تواند روابط پیچیده‌ای را بین متغیرهای مختلفی که بر زمان واخنش در اتاق‌های محصور تأثیر می‌گذارند، تشخیص دهد. رویکردهای جدیدتر پیش‌بینی شرایط آکوستیکی محیط، روش‌های شبیه‌سازی مبتنی بر پرتو و روش‌های المان محدود هستند که دقت بالاتری را در مقایسه با فرمول‌های کلاسیک نشان می‌دهند. با این حال، استفاده از این روش‌ها نیاز به تخصص رایانه و آکوستیک دارد و مستلزم محاسبات وقت‌گیر است. همچنین آنها فقط در موارد خاص به‌کار می‌روند. زیرا مدل هندسی ایجاد شده برای یک اتاق خاص برای سایر اتاق‌ها کاربرد ندارد و عیب دیگری هم که در این رویکرد وجود دارد این است که هریک از روش‌های شبیه‌سازی در یک بازه بسامدی خاصی دقت بالایی دارند [۳].

استفاده از روش‌های مبتنی بر یادگیری ماشین می‌تواند به‌عنوان یک روش جایگزین برای پیش‌بینی زمان واخنش یک محیط ارائه شود. قابلیت این روش‌ها در تعیین پارامترهای تأثیرگذار در پیش‌بینی زمان واخنش باعث می‌شود تا بتوان نقش عوامل مختلف را سنجید. عوامل

همکاران [۸] با استفاده از شبکه عصبی پرسپترون چند لایه برای تخمین زمان واخس در محیط‌های نوفه‌ای^{۱۳} صنعتی پرداخته شده است. در این کار ابتدا به جمع‌آوری دیتاهای حاصل از محیط‌های نوفه‌ای در کارگاه‌های گلدوزی می‌پردازد و بعد با استفاده از یک شبکه عصبی MLP^{۱۴} به تخمین زمان واخس می‌پردازد.

۳. روش پیشنهادی

سیستم طراحی شده در این پژوهش که به منظور پیش‌بینی زمان واخس مبتنی بر شبکه عصبی کانولوشن یک بعدی است، به چند بخش تقسیم می‌شود. ابتدا مجموعه دادگان با استفاده از رویکرد ردیابی پرتو و به کمک شبیه‌ساز اودتون^{۱۵} جمع‌آوری می‌شود. سپس داده‌های یک بعدی ابتدا به عنوان داده‌های خام جمع‌آوری می‌شود و به منظور آموزش و آزمایش مدل تقسیم‌بندی می‌شود.

از ۲۰ درصد مجموعه دادگان به منظور آزمایش شبکه استفاده می‌شود. سپس مجموعه آموزشی به عنوان ورودی شبکه عصبی استفاده می‌شود و پارامترهای مدل شبکه عصبی تنظیم می‌شود و در نهایت مجموعه آزمایش بر روی مدل آموزش داده شده قرار می‌گیرد و عملکرد مدل مورد ارزیابی قرار می‌گیرد. مجموعه دادگان شامل ۷۸۸ نمونه داده است. ورودی‌های اصلی این شبکه عصبی و اطلاعات آماری ورودی‌ها در جداول ۳ تا ۵ آمده است.

۳-۱. شبکه عصبی کانولوشن

شبکه عصبی کانولوشن از انواع شبکه عصبی و یکی از مدل‌های پر استفاده امروزی، در مباحث یادگیری ماشین است. دلیل نام‌گذاری این نوع شبکه عصبی، وجود لایه کانولوشن یا پیچشی، در مرکز شبکه است. این لایه، عملیاتی به نام کانولوشن را انجام می‌دهد. شبکه عصبی کانولوشن از پرسپترون‌های چند لایه و یک یا چند لایه کانولوشن تشکیل شده است. شکل ۱ یک مدل ساده از شبکه عصبی

محیط است و می‌توان در این شرایط صوت را به صورت انتشار در راستای مستقیم فرض کنیم در نتیجه این شباهت‌ها باعث شده ما همگرایی‌هایی را با مباحث گرافیک رایانه‌ای مشاهده کنیم [۳].

۳-۲. بررسی تحقیقات پیشین

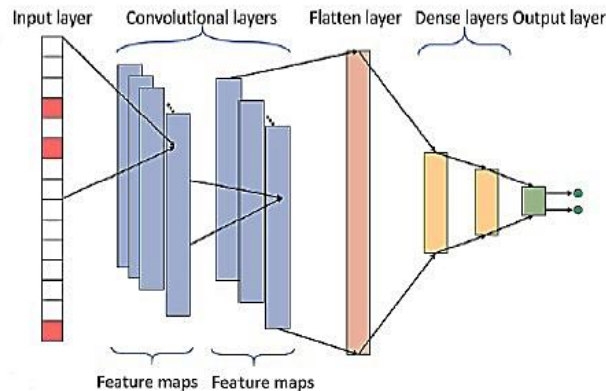
تلاش‌ها در این حوزه در سال‌های اخیر منجر به ارائه رویکردهایی جدید برای پیش‌بینی پارامترهای آکوستیکی شد. هدف این بود که هم عوامل ناشناخته در زمینه پارامترهای آکوستیکی کشف شود و هم اینکه بتوان رویکردی ارائه داد که به واقعیت نزدیک‌تر باشد. نووسویات^{۱۶} [۴،۲] یک الگوریتم برای بهبود فرمول سابین ارائه داد و یک ضریب تصحیح وابسته به بسامد برای اصلاح این رابطه بیان کرد تا بتواند مشکل این فرمول را در فرض یکنواخت بودن میدان صوتی برطرف کند و مقادیر پیش‌بینی شده این فرمول سابین را به مقادیر اندازه‌گیری شده نزدیک‌تر کند. همچنین فالکن^{۱۷} و همکاران [۵] یک روش مبتنی بر یادگیری ماشین و با استفاده از شبکه عصبی پرسپترون چند لایه^{۱۸}، برای تخمین پارامتر آکوستیکی در مسائل واقعیت مجازی طراحی کرد و نتایج حاصل از این کار را با مقدار اندازه‌گیری و نتایج حاصل از فرمول سابین و ایرینگ^{۱۹} مقایسه کردند. این روش به دقتی تقریباً ۹۰ درصد برای بیشتر باندهای بسامدی رسیده است.

وانگ^{۲۰} و همکاران [۶] در مقاله‌ای به تخمین ابعاد اتاق با استفاده از شبکه عصبی کانولوشن پرداختند. در این مقاله از تابع RIR^{۲۱} برای تولید دیتاست استفاده کرده است که این تابع توسط امانوئل هبت^{۲۲} در سال ۲۰۰۶ معرفی شده بود [۷]. پاسخ ضربه‌های به دست آمده توسط تابع RIR به صورت منحنی‌های دامنه-زمان به عنوان ورودی شبکه عصبی استفاده می‌شود، همچنین خروجی شبکه عصبی استفاده شده

در این مقاله ۳ نورون دارد که برای طول و عرض و ارتفاع محیط به کار برده می‌شود. همچنین در مقاله اوحدی و

می‌کنند در صورتی که در شبکه عصبی کانولوشن دو بعدی و سه بعدی که عمدتاً برای کار بر روی تصاویر و ویدئو است، این فیلترها در دو و سه جهت حرکت می‌کنند.

کانولوشن یک بعدی است. به دلیل استفاده از داده‌های یک بعدی از شبکه عصبی کانولوشن یک بعدی استفاده شد. در این گونه از شبکه عصبی، فیلترها فقط در یک جهت حرکت



شکل ۱. معماری یک شبکه عصبی کانولوشن یک بعدی

فیلترها فقط آرایه‌ای از اعداد هستند که در فرایند آموزش به روزرسانی می‌شوند. به حرکت فیلتر استفاده شده روی آرایه ورودی پیچش یا کانوالو می‌گویند. با استفاده از این فیلترها روی داده‌های ورودی پردازش می‌شود. هدف از فیلترهای کانولوشن، استخراج ویژگی‌های سطح بالا از ورودی است. اعداد موجود در فیلترها طی یک فرایند آموزش، از طریق الگوریتم پس انتشار خطا به دست می‌آیند [۹]. همچنین برای پارامتر لایه‌گذاری^{۱۷} در این پژوهش مقدار یکسان^{۱۸} در نظر گرفته می‌شود. از آنجایی که قسمت‌هایی از آرایه که در حاشیه قرار می‌گیرد و نمی‌توان با ابعاد باقیمانده عملیات کانولوشن را انجام داد به همین خاطر، جهت انجام عملیات به نواحی مذکور صفر اضافه می‌کنیم. همچنین مشابه با سایر شبکه‌های عصبی، شبکه عصبی کانولوشن هم از تابع تحریک غیرخطی بعد از لایه کانولوشنی استفاده می‌کند. از مهم‌ترین توابع فعال‌ساز استفاده شده می‌توان به تابع سافت مکس^{۱۹}، تابع رلو^{۲۰}، تابع الو^{۲۱} اشاره کرد [۹]. در این پژوهش از تابع فعال‌ساز رلو، بعد از لایه‌های کانولوشن استفاده شد.

لایه اول در این شبکه عصبی، همیشه یک لایه کانولوشن، و ورودی این لایه یک آرایه از اعداد است. شبکه‌های کانولوشنی معمولاً به یک لایه کانولوشنی محدود نمی‌شوند و لایه‌های بیشتری دارد، در هر یک از این لایه‌ها ویژگی خاصی تشخیص داده می‌شود. معمولاً، لایه‌های ابتدایی برای استخراج ویژگی‌های سطح پایین و لایه‌های بعدی برای استخراج ویژگی‌های سطح بالا به کار برده می‌شود. خروجی این لایه بعد از تبدیل به بردار یک بعدی به لایه شبکه تمام متصل ارسال می‌شود. لایه‌های تمام متصل معمولاً در انتهای معماری‌های شبکه کانولوشن استفاده می‌شوند [۹]. ساختار مدل CNN-1d و تنظیم پارامترها در این مقاله از طریق چندین آزمایش به دست آمده است. ساختار این شبکه عصبی کانولوشن یک بعدی، شامل ۶ لایه کانولوشن، دو لایه پولینگ^{۱۶} و دو لایه تمام متصل است. جدول ۱ پارامترهای معماری شبکه CNN-1d را نشان می‌دهد. در لایه اول از ۸ فیلتر، در لایه دوم از ۱۶ فیلتر، در لایه سوم از ۳۲ فیلتر، در لایه چهارم از ۶۴ فیلتر، در لایه پنجم از ۱۲۸ فیلتر و در لایه ششم از ۲۵۶ فیلتر با سایز ۳ استفاده شد.

جدول ۱. معماری شبکه عصبی کانولوشن

layer	parameter	activation
Conv1D	Filter8/kernel3	ReLU
Conv1D	Filter16/kernel3	ReLU
Conv1D	Filter32/kernel3	ReLU
Maxpooling		
Conv1D	Filter64/kernel3	ReLU
Conv1D	Filter128/kernel3	ReLU
Conv1D	Filter256/kernel3	ReLU
AveragePooling		
Dense	20	ReLU
Dense	1	Linear

محققان، داده‌های خود را با توجه به نیازشان جمع‌آوری می‌کنند. اگرچه به اشتراک گذاشتن این داده‌ها امری معمول است، اما بیشتر اوقات به دلیل کوچک بودن، کمبود تنوع در ویژگی‌ها، قابلیت استفاده در پژوهش را ندارند.

با این حال اکثر مجموعه داده‌های موجود و قابل دسترس، فاقد اطلاعات دقیق در مورد هندسه محیط و ضرایب جذب و اشیا قرار گرفته داخل اتاق هستند. بنابراین ما یک مجموعه داده جدید را بنا به نیاز این پژوهش، تجزیه و تحلیل می‌کنیم. در این مقاله، با توجه به تئوری آکوستیک هندسی مطرح شده در بخش ۲-۲، با استفاده از شبیه‌ساز اودتون به جمع‌آوری مجموعه داده‌ها می‌پردازیم. در نتیجه به‌منظور داشتن یک مجموعه داده استاندارد و کافی برای استفاده در سیستم مبتنی بر یادگیری ماشین در فاز اول ابتدا ۴ کلاس درس با ابعاد مختلف و تعداد اشیاء مختلف، با استفاده از نرم‌افزار اسکچاپ^{۳۳} طراحی شد. اطلاعات کلاس‌های استفاده شده در طراحی در جدول ۲ آمده است. سپس شبیه‌سازی آکوستیکی محیط‌ها به کمک نرم‌افزار اودتون انجام شد. برای هر کلاس ۵۶ ترکیب مختلف از متریاها برای سطوح مختلف در نظر گرفته شد، در انتخاب متریاها سعی شد که تنوع زیادی در ترکیب‌ها لحاظ شود، به‌گونه‌ای که حالات مختلف اتاق با مجموع ضریب جذب بالا تا مجموع ضریب جذب پایین، نسبت به متریاها موجود

همچنین در این معماری از دو لایه پولینگ استفاده شد. لایه پولینگ معمولاً بعد از یک لایه کانولوشن قرار می‌گیرد و از آن برای کاهش اندازه نقشه‌های ویژگی و پارامترهای شبکه می‌توان استفاده کرد. آنچه در این لایه اتفاق می‌افتد، کاهش ابعاد است. لایه پولینگ پارامتر قابلیت آموزش ندارد و صرفاً یک نمونه‌برداری ساده و مؤثر انجام می‌دهد. پولینگ عملکردی شبیه کانولوشن دارد و یک پنجره روی خروجی حرکت می‌کند. رایج‌ترین رویکرد پولینگ، میانگین‌گیری و انتخاب بیشینه مقادیر است.

در رویکرد اول، میانگین مقادیر موجود در هر کدام از پنجره‌ها محاسبه می‌شود و در روش دوم، بیشترین مقدار موجود در این پنجره محاسبه شده و به‌عنوان مقدار در نقشه ویژگی جدید انتخاب می‌شود [۹].

همچنین در این پژوهش به‌منظور حداقل کردن تابع هزینه و به‌روزرسانی پارامترهای شبکه، از الگوریتم بهینه‌ساز آدام^{۳۲} استفاده شد.

۲-۳. جمع‌آوری مجموعه دادگان

الگوریتم‌های یادگیری ماشین توسط داده‌ها تغذیه می‌شوند و عملکرد این روش‌ها برای هر عمل مشخصی با کیفیت و کمیت این داده‌ها محدود می‌شود. برای مسائل مربوط به آکوستیک اتاق، تعداد کمی مجموعه داده استاندارد، بزرگ و قابل دسترسی وجود دارد. در عوض، در بیشتر مطالعات،

شبیه‌سازی در نرم‌افزار اودتون به این صورت است که، ابتدا مدل طراحی شده وارد می‌شود سپس منابع صدا و گیرنده‌ها تعریف می‌شود و بعد از این مرحله باید متریال‌های مورد نظر برای هر سطح انتخاب شود. در مرحله بعد بنا به دقت مد نظر، تعداد پرتو انتخاب و تعیین می‌شود و بعد از آن بنا به نوع خروجی مد نظر، شبیه‌سازی انجام می‌شود. برای تمام شبیه‌سازی‌ها موقعیت استاد یا منبع صوت در محلی که احتمال حضور استاد بیشتر است یعنی در کنار تخته آموزشی در نظر گرفته شد. به همین منظور در ارتفاع ۱۷۰ سانتی‌متری از کف محیط و به فاصله یک متری از دیوار و در وسط اتاق منبع صوتی قرار می‌گیرد. همچنین به تناسب با ابعاد محیط از تعداد گیرنده‌های متفاوتی در این محیط‌ها استفاده می‌شود. این گیرنده‌ها در مکان‌هایی که بیشترین احتمال حضور دانشجویان را دارد انتخاب شده است. در این شبیه‌سازی‌ها از منبع تمام جهته ۸ استفاده شد، به دلیل اینکه معمولاً در بیشتر مواقع، اندازه‌گیری‌های استاندارد پارامترهای آکوستیکی با منابع تمام جهتی انجام می‌شود در نتیجه برای شبیه‌سازی هم از منابع تمام جهته استفاده کردیم، تا با نتایج اندازه‌گیری واقعی بتوانیم مقایسه کنیم.

۳-۳. اطلاعات آماری مجموعه دادگان

در این پژوهش از ۷۸۸ نمونه داده برای آموزش و آزمایش، الگوریتم‌های یادگیری ماشین استفاده شد که اطلاعات آماری مجموعه دادگان در جدول ۳ آمده است. همچنین اطلاعات آماری ضرایب جذب کلی سطوح در بسامد ۵۰۰ هرتز در جدول ۴ آمده است و اطلاعات آماری ضرایب جذب کلی سطوح در بسامد ۲۰۰۰ هرتز هم در جدول ۵ آمده است. جذب کلی سطوح به معنای حاصل ضرب سطح در ضریب جذب است.

ایجاد شود و نتایج به‌دست آمده جامعیت خوبی داشته باشد. تخمین زمان واخس در این پژوهش به این صورت است که، زمان واخس در شبکه عصبی کانولوشن یک بعدی به‌صورت تابعی از ویژگی‌های ورودی تعیین شده در جداول ۳ تا ۵ ارائه می‌شود.

$$y = f(x_1, x_2, x_3, \dots, x_{14}) \quad (3)$$

که در رابطه ۳، y ، زمان واخس پیش‌بینی شده توسط شبکه عصبی است x_1 و تا x_{14} ویژگی‌های ورودی تعیین شده در مجموعه دادگان است که در جداول ۳ تا ۵، مشخصات و اطلاعات آماری این ویژگی‌ها آمده است.



شکل ۲. نمونه کلاس درس طراحی شده

محدوده بسامد صحبت انسان، بین ۳۰۰ هرتز تا ۳ کیلو هرتز یا در بعضی مطالعات بین ۵۰۰ هرتز تا ۲ کیلو هرتز عنوان شده است [۱۰]. در این پژوهش تلاش شد که جمع‌آوری مجموعه دادگان تنها در یک تک بسامد صورت نگیرد در نتیجه با توجه به محدوده بسامد صحبت انسان و همچنین محدودیت در جمع‌آوری مجموعه دادگان، شبیه‌سازی‌ها فقط در دو بسامد ۵۰۰ هرتز و ۲۰۰۰ هرتز صورت گرفت.

جدول ۲. اطلاعات کلاس‌ها و تعداد گیرنده‌ها

تعداد گیرنده	حجم (m^3)	ارتفاع (m)	عرض (m)	طول (m)
۴	۲۲۴	۲/۸	۸	۱۰
۳	۱۷۵	۳/۳	۷	۷/۶
۳	۲۰۷	۳/۳	۷	۹
۴	۲۷۲	۳/۱	۸	۱۱

جدول ۳. اطلاعات آماری مجموعه دادگان

نسبت موقعیت ارتفاع اتاق	نسبت موقعیت عرضی گیرنده به عرض اتاق	نسبت موقعیت طولی گیرنده به طول اتاق	حجم اتاق	ارتفاع (متر)	عرض (متر)	طول (متر)
۰/۴۱۸	۰/۵۱۷	۰/۶۲۵	۲۲۴/۱۱	۳/۰۹۸	۷/۵۷	۹/۵۵
۰/۰۳۴	۰/۲۳۷	۰/۱۵۵	۳۵/۱۲۸	۰/۲۰۷	۰/۴۹	۲۳/۱
۰/۴۱	۰/۵	۰/۳	۲۲۴	۳/۱	۸	۱۰

جدول ۴. اطلاعات آماری جذب کلی سطوح در بسامد ۵۰۰ هرتز

جذب پنجره	جذب تخته	جذب در	جذب دیوارها	جذب سقف	جذب کف	جذب میز و صندلی
۱/۵۵	۰/۱۷۴	۰/۹۶	۱۳/۴۲	۲۱/۸۸	۱۸/۳۴	۱۵/۱۳۹
۱/۲۱	۰/۱۱	۱/۰۶۸	۱۴/۳۰	۲۳/۷۱	۲۲/۷۲	۷/۳۹۱
۱/۱۷	۰/۲۵	۰/۴۹۵	۵/۸۴۷	۸/۸	۶/۴	۱۷/۰۵۶

جدول ۵. اطلاعات آماری جذب کلی سطوح در بسامد ۲۰۰۰ هرتز

جذب پنجره	جذب تخته	جذب در	جذب دیوارها	جذب سقف	جذب کف	جذب میز و صندلی
۲/۱۵۸	۰/۱۱۶	۰/۵۵۵	۱۷/۳۰۲	۱۶/۰۱	۱۳/۸۷۹	۹/۷۳
۱/۵۱	۰/۰۵۸	۰/۴۱۲	۲۲/۸۴۵	۱۹/۱۷۳	۱۹/۰۱۶	۶/۵۴
۲/۱	۰/۱۵	۰/۳۳	۴/۰۴	۶/۱۶	۴/۴	۷/۴۲۷

۳-۴. معیارهای ارزیابی

ضریب تعیین که با R^2 نشان داده می‌شود قدرت توضیح‌دهندگی مدل را نشان می‌دهد [۸،۵]. ضریب تعیین نشان می‌دهد که چند درصد از تغییرات متغیر وابسته توسط متغیرهای مستقل توضیح داده می‌شود. در واقع این ضریب بیانگر میزان احتمال همبستگی میان دو دسته داده در آینده است. این ضریب در واقع نتایج تقریبی پارامتر مورد نظر در آینده را براساس مدل ریاضی تعریف شده که منطبق بر داده‌های موجود است، بیان می‌دارد. ضریب تعیین، معیاری است از این که خط رگرسیون^{۲۴}، چقدر خوب خوانده‌ها را معرفی می‌کند. اگر خط رگرسیون از تمام نقاط بگذرد توانایی معرفی همه متغیرها را دارد و هرچه از نقاط دورتر باشد نشان‌دهنده توانایی کمتر است. اگر $R^2=1$ باشد یعنی زمانی که از متغیرهای مستقل استفاده می‌کنیم هیچ خطایی

وجود ندارد که این بهترین حالت ممکن است. اگر $R^2=0$ باشد یعنی استفاده از متغیرهای مستقل هیچ تأثیری بر برآورد خط رگرسیونی ندارد.

ضریب بین صفر تا یک تغییر می‌کند و مقدار بهینه آن یک است و این حالت مربوط به زمانی است که در آن مقادیر شبیه‌سازی شده دقیقاً مشابه مقادیر مشاهداتی باشد.

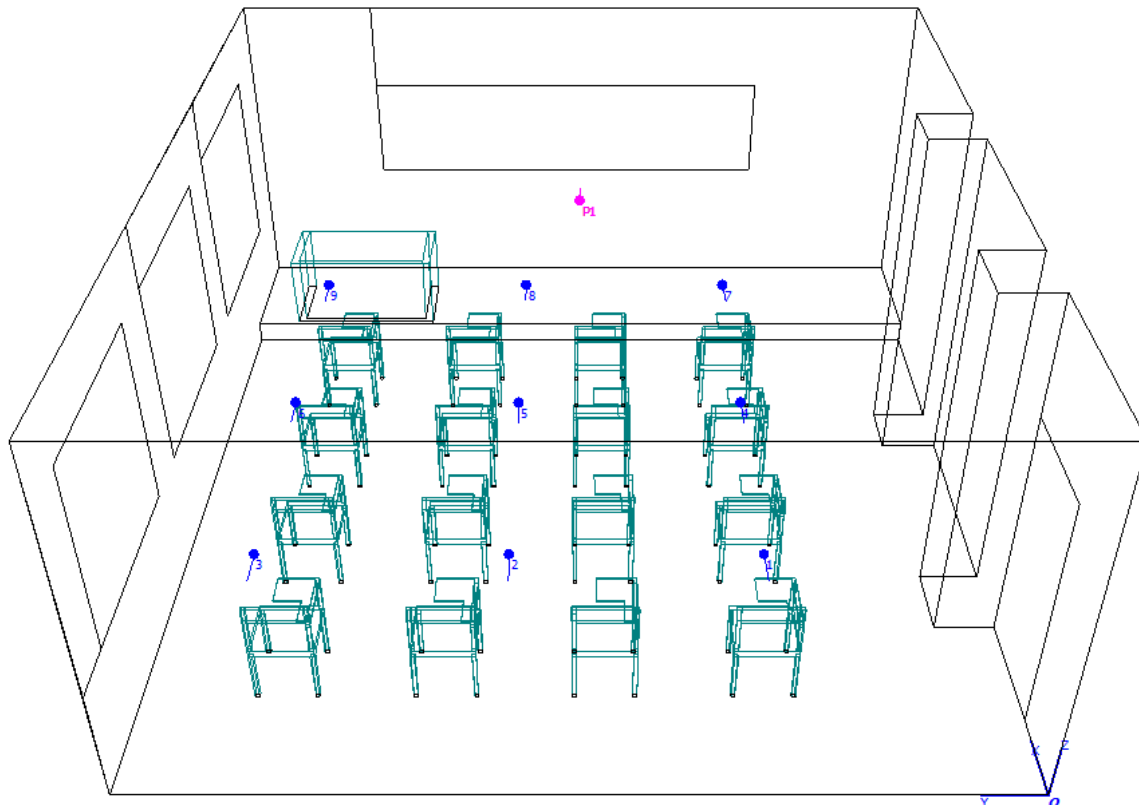
$$R^2 = 1 - \frac{\sum_{i=1}^n (x_{imeas} - x_{ipred})^2}{\sum_{i=1}^n (x_{imeas} - \bar{x})^2} \quad (4)$$

رویکرد دیگری که برای ارزیابی مدل استفاده شد، میانگین مربعات خطا بود. این رویکرد روشی برای برآورد میزان خطاست که در واقع تفاوت بین مقادیر اندازه‌گیری و آنچه تخمین زده شده است. این شاخص که مقداری همواره نامنفی دارد، هرچقدر مقدار آن به صفر نزدیک‌تر باشد،

شده، مقدار متغیر پیشبینی شده و مقدار متغیر اندازه‌گیری شده است.

$$MSE = \frac{1}{n} \sum_{i=1}^n [(x_{imeas} - x_{ipred})^2] \quad (5)$$

نشان‌دهنده میزان کمتر خطاست. مقدار این شاخص به صورت فرمول ۵ بیان می‌شود که در این رابطه n ، x_{imeas} و x_{ipred} به ترتیب برابر با تعداد متغیر اندازه‌گیری



شکل ۳. شبیه‌سازی کلاس ۲۰۲ دانشگاه صدا و سیما در نرم‌افزار اودئون

با توجه به شرایط اندازه‌گیری، کلاس ۲۰۲ اقدام به شبیه‌سازی شد. همچنین با توجه به ثبت ۹ اندازه‌گیری، از ۹ گیرنده با توجه به موقعیت‌های مختلفی که احتمال حضور دانش‌آموزان وجود دارد، استفاده شد. در این اندازه‌گیری از منابع تمام جهت استفاده شده است. نتایج به‌دست آمده از شبیه‌سازی در جدول ۶ نشان‌دهنده دقت قابل قبول شبیه‌سازی است. برای بررسی بیشتر می‌توان گفت که نتایج جدول ۶ بیانگر این نکته است که دقت شبیه‌سازی با اودئون در بسامدهای بالاتر، بیشتر است، زیرا در محدوده بسامد بالا به علت اینکه فرضیات آکوستیک هندسی به خوبی رعایت

۴. اعتبارسنجی نتایج شبیه‌سازی

در این مرحله به منظور بررسی و اطمینان از نتایج نرم‌افزار شبیه‌ساز اودئون، در دو بسامد ۵۰۰ هرتز و ۲۰۰۰ هرتز، از نتایج اندازه‌گیری واقعی استفاده شد، تا مطمئن شویم مجموعه داده‌های شبیه‌سازی شده، جایگزین مناسبی برای اندازه‌گیری واقعی است. به این منظور کلاس شماره ۲۰۲ دانشگاه صدا و سیما مورد بررسی قرار گرفت، این کلاس تقریباً مستطیل شکل است و به طول ۶ متر و عرض ۶/۹ و ارتفاع ۳/۳ متر است. در نتیجه نتایج زمان واخس این کلاس با شبیه‌سازی حاصل از اودئون مقایسه می‌شود. به این منظور

همان‌طور که از شکل ۴ مشخص است، ارتباط خطی بین نتایج حاصل از ساین و نتایج شبیه‌سازی وجود ندارد. همچنین برای بسامد ۲۰۰۰ هرتز، همان‌طور که مشخص است بازهم ارتباط کاملاً خطی بین نتایج حاصل از ساین و نتایج شبیه‌سازی وجود ندارد، ولی ارتباط خطی‌تری نسبت به بسامد ۵۰۰ هرتز وجود دارد. علت این است که در بسامدهای نزدیک‌تر به بسامد شرودر در یک محیط، به علت وجود امواج ایستاده، وابستگی به محل قرارگیری گیرنده بیشتر است و به همین علت در بسامدهای پایین‌تر، فرمول ساین دقت کمتری دارد. همچنین نشان داده شد که فرض یکنواخت بودن میدان صوتی محیط، باعث ایجاد خطا در پیش‌بینی می‌شود و فرمول کلاسیک تخمین زمان واخشن نمی‌تواند به خوبی رفتار صوتی یک محیط رو پیش‌بینی کند. بنابراین، مشهود است که به‌منظور داشتن دقت بیشتر در پیش‌بینی زمان واخشن باید از پارامترهای بیشتری نسبت به فرمول ساین استفاده شود. در نتیجه لزوم استفاده از رویکردهای جایگزین برای پیش‌بینی زمان واخشن ثابت می‌شود.

۴-۱. نتایج شبکه عصبی کانولوشن

در این مرحله نتایج شبکه عصبی کانولوشن یک بعدی استفاده شده به‌منظور پیش‌بینی زمان واخشن مورد بررسی قرار می‌گیرد. در این پژوهش از مجموعه دادگان تولید شده با رویکرد آکوستیک هندسی و به کمک نرم‌افزار اودئون استفاده شد. جزئیات معماری شبکه عصبی کانولوشن یک بعدی، در این پژوهش در بخش ۳-۱ آمده است. همان‌طور که در جدول ۷، نتایج شبکه عصبی آمده است، با توجه به معیارهای ارزیابی مطرح شده در بخش ۳-۴، این سیستم طراحی شده کارایی مناسبی برای پیش‌بینی زمان واخشن در دو بسامد ۵۰۰ و ۲۰۰۰ هرتز ثبت کرده است.

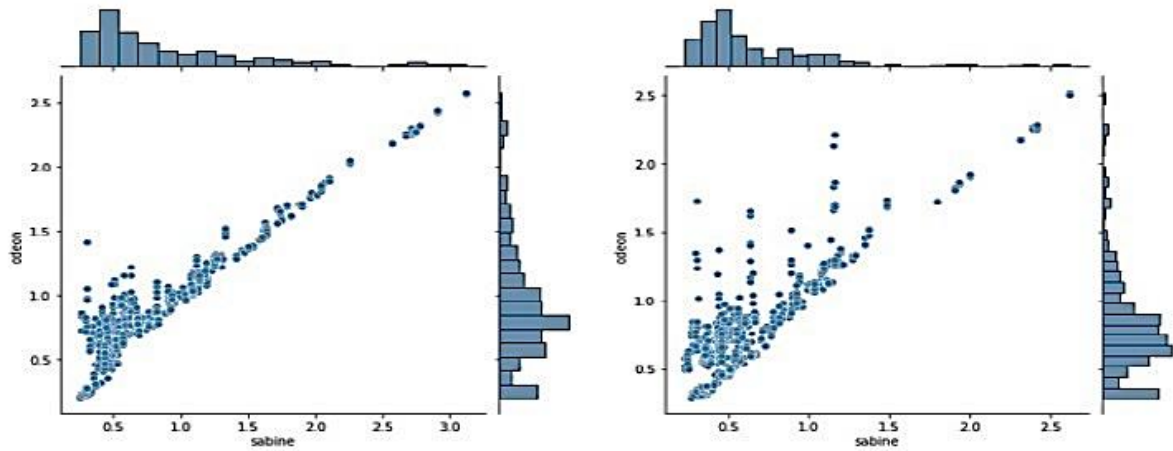
می‌شود در نتیجه شبیه‌سازی‌ها خطای کمتری دارند. برای بررسی نتایج در محدوده بسامد پایین، با استفاده از مفهوم بسامد شرودر^{۲۵} تحلیل انجام می‌شود. طبق گفته شرودر در [۱۱] محدوده بسامدی بالاتر از بسامد شرودر، رفتار صوت به‌صورت پرتوای خواهد بود و در محدوده پایین‌تر از بسامد شرودر، صوت به‌صورت موجی رفتار خواهد کرد. در نتیجه از آنجایی که بسامد شرودر برای کلاس ۲۰۲ عددی تقریباً برابر با ۱۸۰ هرتز است. که با توجه به ضریب ۴ در نظر گرفته شده برای انتقال به ناحیه با حالت پرتوای [۱۱]، شبیه‌سازی‌ها در بسامد ۵۰۰ هرتز با کمی خطای بیشتر همراه است.

جدول ۶. نتایج زمان واخشن اندازه‌گیری و شبیه‌سازی

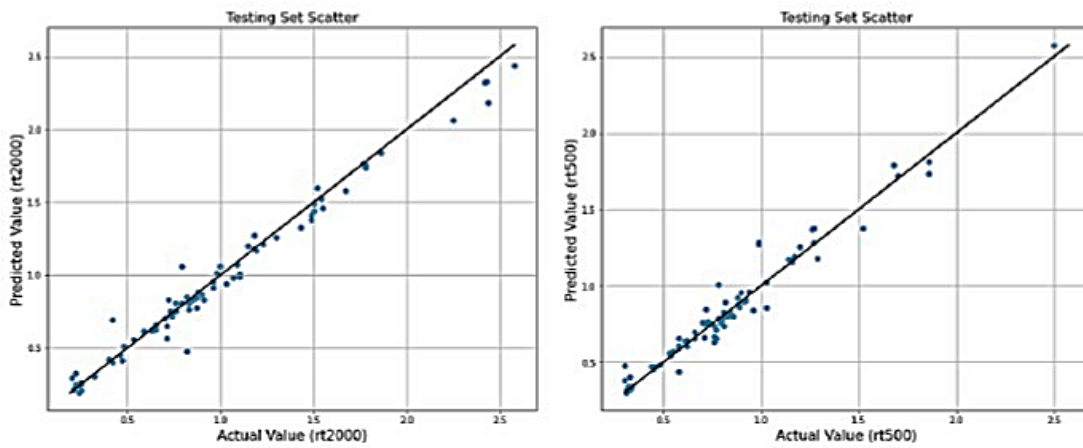
	بسامد ۲۰۰۰ هرتز	بسامد ۵۰۰ هرتز
اندازه‌گیری واقعی	۱/۱۸ (ثانیه)	۱/۴۸ (ثانیه)
نتایج حاصل از شبیه‌سازی اودئون	۱/۲۰ (ثانیه)	۱/۵۴ (ثانیه)
نتایج ساین	۱/۲۳ (ثانیه)	۱/۶۷ (ثانیه)

همان‌طور که از نتایج جدول ۶ مشخص است، شبیه‌سازی اودئون می‌تواند با دقت بالایی زمان واخشن را محاسبه کند. به همین جهت از نتایج شبیه‌سازی اودئون برای آموزش و تست مدل مبتنی بر یادگیری ماشین استفاده می‌کنیم.

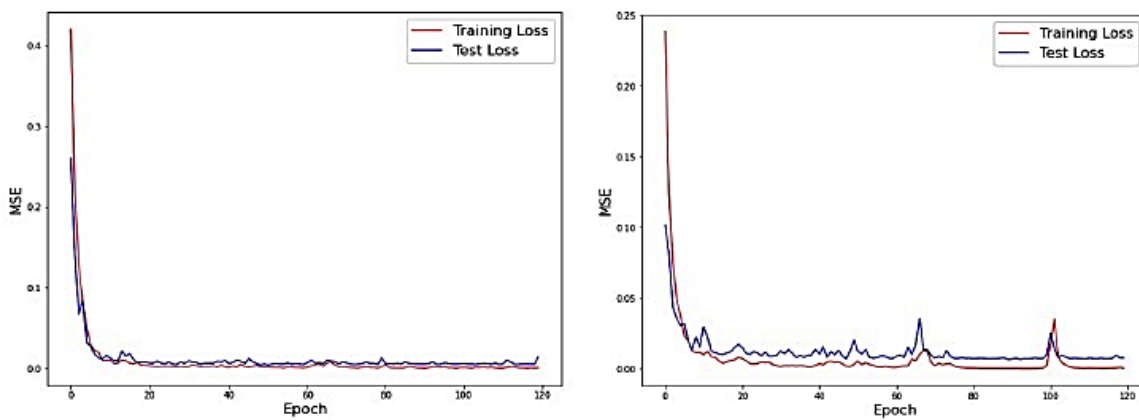
همچنین به‌منظور بررسی بیشتر نتایج شبیه‌سازی، در این مرحله با استفاده از مجموعه دادگان ساخته شده حاصل از شبیه‌سازی، ابتدا نتایج فرمول ساین را استخراج می‌کنیم و این نتایج را با نتایج شبیه‌سازی اودئون مقایسه می‌کنیم. در بررسی‌های صورت گرفته بین نتایج شبیه‌سازی اودئون و نتایج حاصل از فرمول ساین، از نمودار توزیع مشترک بین ساین و اودئون استفاده شد. در بسامد ۵۰۰ هرتز،



شکل ۴. نمودار توزیع مشترک نتایج اودتون و سابین در فرکانس ۵۰۰ هرتز (شکل سمت راست) و ۲۰۰۰ هرتز (شکل سمت چپ)



شکل ۵. نمودار نتایج شبکه عصبی کانولوشن یک بعدی در فرکانس ۵۰۰ هرتز (شکل سمت راست) و ۲۰۰۰ هرتز (شکل سمت چپ)



شکل ۶. نمودار میانگین مربعات خطا براساس ایپوک در فرکانس ۵۰۰ هرتز (شکل سمت راست) و ۲۰۰۰ هرتز (شکل سمت چپ)

۵. نتیجه گیری

با توجه به اهمیت پیش‌بینی زمان واخشی به‌عنوان یکی از مهم‌ترین پارامترهای آکوستیکی در طراحی محیط‌های آموزشی، هدف از این پژوهش، پیش‌بینی زمان واخشی در نظر گرفته شد. در این مقاله یک سیستم برای تخمین زمان واخشی با استفاده از الگوریتم‌های یادگیری ماشین و به‌صورت یادگیری نظارت شده، ارائه شده است. به‌طوری‌که با استفاده از مجموعه دادگان ثبت شده بتوان یک روش جایگزین برای فرمول‌های سابقین برای پیش‌بینی زمان واخشی طراحی کرد. به این منظور با توجه به مطالب گفته شده در بخش ۳-۲، به جمع‌آوری مجموعه دادگان پرداخته شد و سپس مدل، با استفاده از شبکه عصبی کانولوشن یک بعدی بررسی شد، و با تنظیمات مختلف بر روی پارامترهای تنظیمی، نتایج با دقت مناسب ذخیره شد. معماری شبکه عصبی استفاده شده در جدول ۱ آمده است. در این مطالعه، ابعاد و حجم محیط و مقادیر ضرایب جذب تفکیک شده سطوح و موقعیت قرارگیری گیرنده، به‌عنوان ورودی برای این مدل پیشنهادی تعیین شد و زمان واخشی به‌عنوان خروجی این سیستم، در نظر گرفته شد. در این پژوهش نتایج به شرح موجود در جدول ۷ به ثبت رسیده است.

همان‌طور که در جدول ۷ مشخص است در این مدل ضریب تعیین به‌دست آمده در بسامد ۲۰۰۰ هرتز بیشتر از بسامد ۵۰۰ هرتز است، که تا حد زیادی به این علت است که در مجموعه دادگان موجود، در بسامد پایین‌تر، زمان واخشی ثبت شده، تغییرات بیشتری نسبت به زمان واخشی در بسامد ۲۰۰۰ هرتز دارند. همچنین به‌منظور بررسی بیشتر در این پژوهش، نمودار نتایج شبکه عصبی در شکل ۵ و نمودار خطا با توجه به ایپوک^{۲۶} در شکل ۶ برای دو بسامد ۵۰۰ هرتز و ۲۰۰۰ هرتز رسم شد. همان‌طور که از شکل ۵ مشخص است، سیستم طراحی شده در این پژوهش با در نظر گرفتن ویژگی‌های لحاظ شده در مجموعه دادگان، توانست یک مدل برای پیش‌بینی زمان واخشی ارائه دهد، و می‌تواند زمان واخشی را برای دادگان آزمایش به میزان خوبی پیش‌بینی کند. همچنین در شکل ۶ مشخص است که معیار ارزیابی، میانگین مربعات خطا در شبکه عصبی کانولوشن یک بعدی، در حین فرایند آموزش و تست کاهش پیدا می‌کند و سیستم، یادگیری مناسبی را از خود نشان می‌دهد.

جدول ۷. نتایج شبکه عصبی کانولوشن یک بعدی

بسامد	آموزش		آزمایش	
	R^2	MSE	R^2	MSE
۵۰۰ هرتز	۰/۹۷	۰/۰۰۴	۰/۹۳	۰/۰۰۶
۲۰۰۰ هرتز	۰/۹۸	۰/۰۰۱	۰/۹۵	۰/۰۰۷

۷. مآخذ

- [1] Prodi, Nicola, Chiara Visentin, and Andrea Farnetani, "Intelligibility, listening difficulty and listening efficiency in auralized classrooms", *The Journal of the Acoustical Society of America*, 2010, Vol.128, no.1, pp.172-181.
- [2] Nowoświat, Artur, Marcelina Olechowska, and Jan Ślusarek, "Prediction of reverberation time using the residual minimization method", *Applied Acoustics*, 2016, Vol.106, Vol.42-50.
- [3] Hargreaves, J. A., "Synergies between the high-frequency Boundary Element Method and Geometric Acoustics", In *Proceedings of e-Forum Acusticum*, 2020, pp.545-547, HAL, 2020.
- [4] Nowoświat, Artur, and Marcelina Olechowska, "Estimation of reverberation time in classrooms using the residual minimization method", *Archives of Acoustics*, 2017.
- [5] Perez, Ricardo Falcon, Georg Götz, and Ville Pulkki, "Machine-learning-based estimation of reverberation time using room geometry for room effect rendering", In *Proceedings of the 23rd International Congress on Acoustics: integrating 4th EAA Euroregio*, 2019, Vol.9, p.13.

- [6] Yu, Wangyang, and W. Bastiaan Kleijn, "Room Geometry Estimation from Room Impulse Responses using Convolutional Neural Networks", *arXiv preprint arXiv: 1904.00869*, 2019.
- [7] Habets, Emanuel AP., "Room impulse response generator", *Technische Universiteit Eindhoven, Tech. Rep.*, 2006, Vol.2, no.2.4 p.1.
- [8] Aliabadi, Mohsen, Rostam Golmohammadi, Abdolreza Ohadi, Muharram Mansoorizadeh, Hassan Khotanlou, and Mohammad Saber Sarrafzadeh, "Development of an empirical acoustic model for predicting reverberation time in typical industrial workrooms using artificial neural networks", *Acta Acustica united with Acustica*, 2014, Vol.100, no.6, pp.1090-1097.
- [9] Aloysius, Neena, and M. Geetha, "A review on deep convolutional neural networks", In *2017 International Conference on Communication and Signal Processing (ICCSP)*, pp.0588-0592. IEEE, 2017.
- [10] Bhangale, Kishor Barasu, and K. Mohanaprasad, "A review on speech processing using machine learning paradigm", *International Journal of Speech Technology*, 2021, Vol.24, no.2, pp.367-388.
- [11] Khezri, Seyed Mostafa, and Pedram Jafari Shalkouhi, "The Schroeder Frequency of Furnished and Unfurnished Spaces", *Romanian Journal of Acoustics and Vibration*, 2012, Vol.9, no.2, p.113.

پی نوشت:

-
- 1. Convolutional Neural Network
 - 2. Reverberation Time
 - 3. Sabine
 - 4. Ray Tracing
 - 5. Radiosity
 - 6. Nowoswiat
 - 7. Falcon
 - 8. Multilayer Perceptron
 - 9. Eyring
 - 10. Wangyang Yu
 - 11. Room Impulse Response
 - 12. Emanuel A.P.Habets
 - 13. Noise
 - 14. Multilayer Perceptron
 - 15. Odeon
 - 16. Pooling Layer
 - 17. Padding
 - 18. Same
 - 19. Softmax Function
 - 20. Rectified Linear Unit (relu)
 - 21. Exponential Linear Unit (ELU)
 - 22. Adam Optimization
 - 23 Sketchup Software
 - 24. Linear Regression
 - 25. Schroeder Frequency
 - 26. Epoch