مطالعه تطبیقی سه مدل ماتریس در ایمنی هوانوردی با استفاده از یادگیری ماشین بـا داده های واقعی

تاریخ دریافت: ۱۴۰۵/۰۷/۰۶
تاریخ پذیرش: ۱۴۰۵/۰۸/۱۸

رضا صفرزاده[۱]، فرشاد پازوکی[۲]، شهرام اعتمادی حقیقی[۳]، احمد مامندی[٤]

۱- دانشجوی دکتری، گروه مهندسی هوا فضا، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران،
۲- استادیار، گروه مهندسی هوا فضا، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران، Pazooki_Fa@iau.ac.ir
۳- دانشیار، گروه مهندسی مکانیک، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران،
٤- استاد، گروه مهندسی هوا فضا، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران،

چکیده

ایمنی هوانوردی به شناسایی دقیق و اولویت‌بندی مؤثر مخاطرات عملیاتی وابسته است. ماتریس ریسک ایکائو با وجود سادگی و مقبولیت جهانی، در شناسایی سناریوهای پیچیده و مرزی محدودیت دارد. این مقاله سه مدل ماتریس ریسک شامل ماتریس استاندارد ایکائو، ماتریس فیلتر دوگانه و ماتریس تاب‌آور اصلاح‌شده را از نظر قابلیت یادگیری توسط شبکه عصبی و اعتبارسنجی با داده‌های واقعی مقایسه کرده است. برخلاف مطالعات پیشین، از ۴۷،۷۲۳ گزارش رویداد واقعی پایگاه داده ناسا و پنج ویژگی ورودی با شبکه پرسپترون چندلایه استفاده شده است. عملکرد مدل‌ها با معیارهای MSE، RMSE، MAE و R^2 ارزیابی گردیده است. مدل تاب‌آوراصلاح‌شده با $R^2=0.9612$ و MAE=1.108 بهترین عملکرد را نشان می دهد؛ سپس مدل فیلتر دوگانه و ماتریس استاندارد ایکائو قرار گرفتنده است. آزمون فریدمن برتری معنادار شبکه عصبی را تأیید کر. ($p < 0.001$) افزودن سطوح میانی به ماتریس ایکائو، بدون تغییر منطق ضرب احتمال و شدت، قابلیت یادگیری را بهبود می‌بخشد. مدل فیلتر دوگانه با وجود عملکرد خوب، به دلیل پیچیدگی محاسباتی برای کاربردهای برخط توصیه نمی‌گردد. اعتبارسنجی با داده‌های واقعی نشان داده است شکاف معناداری بین عملکرد مدل‌ها روی داده‌های شبیه‌سازی و واقعی وجود دارد و استفاده از داده‌های واقعی در پژوهش‌های آتی ضروری است.

واژه‌های کلیدی: *ماتریس تاب آور، ماتریس ریسک ایکائو، ماتریس فیلتر دوگانه، یادگیری ماشین، سیستم مدیریت ایمنی*

A Comparative study of three matrix models in aviation safety using machine learning with real data

Reza Safarzadeh[1], Farshad Pazooki[2], Shahram Etemadi Haghighi[3], Ahmad Mamandi[4]

1- Department of Aerospace Engineering, SR.C., Islamic Azad University, Tehran, Iran
2- Department of Aerospace Engineering, SR.C., Islamic Azad University, Tehran, Iran
3- Department of Mechanical Engineering, SR.C., Islamic Azad University, Tehran, Iran
4- Department of Aerospace Engineering, SR.C., Islamic Azad University, Tehran, Iran

Abstract

Aviation safety depends on accurate identification and effective prioritization of operational hazards. The ICAO risk matrix, despite its simplicity and global acceptance, has limitations in identifying complex and borderline scenarios. This article compares three risk matrix models the standard ICAO matrix, the dual-filter matrix, and the modified resilient matrix in terms of learnability by a neural network and validation with real data. Unlike previous studies, 47,723 real event reports from the NASA database and five input features were used with a multilayer perceptron network. Model performance was evaluated using MSE, RMSE, MAE, and R^2 criteria. The modified resilient model showed the best performance with $R^2=0.9612$ and MAE=1.108; the dual-filter model and the standard ICAO matrix ranked next. The Friedman test confirmed the significant superiority of the neural network ($p < 0.001$). Adding intermediate levels to the ICAO matrix, without changing the logic of multiplying probability and severity, improves learnability. The dual-filter model, despite its good performance, is not recommended for online applications due to computational complexity. Validation with real data has shown that there is a significant gap between model performance on simulated and real data, and the use of real data in future research is essential.

Keywords: *Resilient Matrix, ICAO Risk Matrix, Dual Filter Matrix, Machine Learning; Safety Management System*

۱. مقدمه

ایمنی هوانوردی همواره به عنوان یکی از اولویت‌های اصلی صنعت هوایی شناخته شده است. رشد ترافیک هوایی، افزایش پیچیدگی ناوگان و تغییر الگوهای عملیاتی، ارزیابی ریسک را به چالشی اساسی تبدیل کرده است. در میان ابزارهای مختلف، ماتریس ریسک ایکائو که در راهنمای مدیریت ایمنی معرفی شده، به دلیل سادگی و کاربرد گسترده در سطح جهان شناخته می‌شود [۱]. این ماتریس با ترکیب دو بعد احتمال و شدت، مخاطرات را در سطوح مختلف طبقه‌بندی می‌کند.

پژوهش‌های اخیر نشان داده‌اند که ماتریس استاندارد در شناسایی ریسک‌های مرزی و شرایط عملیاتی پیچیده دچار ضعف است [۲, ۳]. مدل‌های اصلاح شده‌ای مانند ماتریس فیلتر دوگانه [۴] و ماتریس با سطوح تاب‌آوری بیشتر [۵] پیشنهاد شده‌اند، اما تحلیل مقایسه‌ای نظام‌مند میان این مدل‌ها با رویکرد کمی و مبتنی بر یادگیری ماشین بر روی داده‌های واقعی کمتر مورد توجه قرار گرفته است. همچنین ارزیابی همزمان این مدل‌ها از نظر دقت در بازتولید سطوح ریسک، شکافی است که پژوهش حاضر به دنبال پر کردن آن است.

این مطالعه برای نخستین بار یک چارچوب مقایسه‌ای مبتنی بر شبکه عصبی برای ارزیابی عملکرد سه مدل ماتریس ریسک در بازتولید خروجی خود مدل‌ها با استفاده از داده‌های واقعی ارائه می‌دهد. پرسش اصلی پژوهش این است: آیا تغییر در ساختار ماتریس ریسک (از طریق افزودن سطوح میانی تاب‌آوری یا اعمال فیلترهای دوگانه) بدون تغییر در منطق ضرب احتمال و شدت، قابلیت یادگیری و بازتولید سطوح ریسک را توسط یک شبکه عصبی پرسپترون چندلایه بهبود می‌بخشد؟

۲. مرور ادبیات

ارزیابی ریسک نقش اساسی در مدیریت ایمنی هوانوردی ایفا می‌کند. یکی از قدیمی‌ترین چارچوب‌ها، ماتریس ریسک ایکائو است که در نشریه راهنمای مدیریت ایمنی تشریح شده است [۱]. مطالعات متعددی به محدودیت‌های ماتریس کیفی اشاره کرده‌اند. کوکس (۲۰۰۸) نشان داد که ماتریس ریسک استاندارد ممکن است تناقض‌های ریاضی داشته باشد [۳]. همچنین پژوهش‌های تجربی در صنعت هوانوردی نشان داده‌اند که ماتریس ایکائو در شناسایی ریسک‌های میانی (ناحیه زرد) حساسیت کافی ندارد [۶].

در پاسخ به این محدودیت‌ها، مدل‌های پیشرفته‌تری مانند ماتریس فیلتر دوگانه ارائه شده است [۴]. مدل دیگر، ماتریس با سطوح تاب‌آوری اصلاح شده است که همان ساختار ایکائو را حفظ می‌کند اما با تقسیم ناحیه زرد به دو سطح (نارنجی و زرد) سعی در افزایش حساسیت به سناریوهای مرزی دارد [۵].

پژوهش‌های اخیر از رویکردهای یادگیری ماشین برای ارزیابی ریسک استفاده کرده‌اند. چن و همکاران (۲۰۲۴) از شبکه عصبی برای پیش‌بینی ریسک در نگهداری هواپیما بهره بردند [۸]. کومار و سینگ (۲۰۲۵) مدل ترکیبی یادگیری ماشین را برای پیش‌بینی ریسک عملیات زمینی ارائه کردند [۹]. وانگ و همکاران (۲۰۲۴) چارچوبی مبتنی بر شبکه ترنسفورمر برای تحلیل متون رویدادهای ناایمن هوانوردی توسعه دادند [۱۰].

با وجود این پیشرفت‌ها، شکاف تحقیقاتی در مقایسه کمی سه مدل فوق با یک معماری یکسان

شبکه عصبی و با استفاده از داده‌های واقعی باقی مانده بود. پژوهش حاضر به پر کردن این شکاف می‌پردازد.

جدول ۱. پیشینه پژوهش

ردیف	پژوهش	مدل	مزیت اصلی	محدودیت/ شکاف
۱	ایکائو (۲۰۱۸ , ۲۰۲۲) [۱] [۱۲]	ماتریس ریسک ۵×۵	سادگی، پذیرش جهانی، سهولت پیاده‌سازی	عدم حساسیت به ریسک‌های مرزی
۲	کوکس (۲۰۰۸) [۳]	تحلیل نظری محدودیت‌های ماتریس ریسک	شناسایی رسمی مشکلات ریاضی	فاقد مدل جایگزین عملیاتی
۳	داجیم (۲۰۱۵)[۶]	افزایش سطوح، تعریف دقیق سلول‌ها	بهبود روش‌شناختی در طراحی ماتریس	عدم ارائه مدل عددی مشخص و عدم ارزیابی با داده واقعی
۴	لی (۲۰۱۹) [۴]	ماتریس فیلتر دوگانه	افزایش دقت در اولویت‌بندی با در نظر گرفتن زمینه عملیاتی	پیچیدگی محاسباتی بالا
۵	هلنیج (۲۰۱۴) [۵]	رویکرد تاب‌آوری	تغییر نگاه از پیشگیری به پاسخ‌دهی و بازیابی	فاقد ابزار عملیاتی کمی
۶	ال و همکاران (۲۰۱۹)[۲]	ارزیابی تجربی محدودیت‌های ماتریس ایکائو در عملیات واقعی	ارائه شواهد تجربی از ضعف ماتریس استاندارد	پیشنهاد مدل جایگزین مشخص ارائه نشده است
۷	چن و همکاران (۲۰۲۴)[۸]	شبکه عصبی برای پیش‌بینی ریسک در نگهداری هواپیما	اولین تلاش برای ترکیب یادگیری ماشین و ارزیابی ریسک	عدم مقایسه مدل عصبی با ماتریس‌های استاندارد
۸	کومار (۲۰۲۵) [۱۰]	مدل ترکیبی یادگیری ماشین	دقت بالا در پیش‌بینی ریسک عملیات	قابلیت تفسیر پایین
۹	پاسمن و راجر (۲۰۲۰) [۷]	مدل ارزیابی ریسک پویا برای ایمنی هوانوردی	معرفی بعد زمان در ارزیابی ریسک	پیچیدگی زیاد
۱۰	زنگ وهمکاران (۲۰۲۴)[۱۲]	یادگیری برای مدیریت ایمنی پیش‌بینی‌کننده	حفظ حریم خصوصی داده در ناوگان چند شرکتی	عدم مقایسه با ماتریس استاندارد به عنوان خط پایه

بر اساس جدول فوق، سه شکاف اصلی در ادبیات موجود شناسایی می‌شود: (۱) عدم وجود مقایسه کمی نظام‌مند میان مدل‌های مختلف ماتریس ریسک با استفاده از یک معماری یکسان یادگیری ماشین و داده‌های واقعی، (۲) تمرکز بیشتر پژوهش‌ها بر پیش‌بینی ریسک واقعی به جای ارزیابی قابلیت بازتولید ساختار داخلی مدل‌ها، و (۳) تأیید نشدن تجربی این موضوع که افزودن سطوح میانی به ماتریس ایکائو واقعاً قابلیت یادگیری توسط شبکه عصبی را بهبود می‌بخشد.

جدول ۲. نوآوری نسبت به پژوهش‌های پیشین

معیار مقایسه	چن و همکاران (۲۰۲۴) [۸]	کومار و سینگ (۲۰۲۵) [۱۰]	پژوهش حاضر
نوع داده	داده واقعی	داده واقعی عملیات زمینی	شبیه‌سازی شده اثبات مفهوم
تعداد مدل‌های ریسک مقایسه‌شده	۱ مدل خروجی شبکه	۱ مدل ترکیبی	۳ مدل (استاندارد، فیلتر دوگانه، تاب‌آور)
معماری یکسان برای همه مدل‌ها	ندارد (فقط یک مدل)	ندارد	دارد (یکسان)
تفکیک بازتولید ساختار مدل از پیش‌بینی ریسک	خیر (پیش‌بینی مستقیم)	خیر	بله (صراحتاً تفکیک شده)

معیار مقایسه	چن و همکاران [۸](۲۰۲۴)	کومار و سینگ [۱۰](۲۰۲۵)	پژوهش حاضر
بررسی اثر تقسیم ناحیه زرد	خیر	خیر	بله

۳. روش شناسی

۱،۳ رویکرد کلی

روش‌شناسی این تحقیق از چهار مرحله اصلی تشکیل شده است: (۱) گردآوری و پیش‌پردازش داده‌های واقعی، (۲) تعریف مدل‌های ریسک و محاسبه شاخص هدف، (۳) پیاده‌سازی شبکه عصبی پرسپترون چندلایه و (۴) تحلیل مقایسه‌ای عملکرد.

۲،۳ داده‌های واقعی

در این پژوهش از پایگاه داده ایمنی ناسا استفاده شده است. این پایگاه داده یک سیستم گزارش‌دهی داوطلبانه، محرمانه و غیرتنبیهی است که از سال ۱۹۷۶ تاکنون بیش از ۱،۷ میلیون گزارش رویداد ایمنی هوانوردی را جمع‌آوری کرده است [۱۱] [۱۲]. هر گزارش شامل اطلاعات دقیقی درباره زمان، مکان، محیط، نوع هواپیما، و شرح رویداد می‌باشد.

برای این مطالعه، ۴۷،۷۲۳ گزارش از سال‌های ۲۰۱۰ تا ۲۰۲۳ انتخاب گردید. معیارهای انتخاب شامل: (۱) گزارش‌های مربوط به عملیات پروازی خطوط هوایی تجاری، (۲) گزارش‌های دارای داده‌های کامل در خصوص پنج ویژگی ورودی اصلی و (۳) گزارش‌هایی که دارای برچسب شدت رویداد توسط کارشناسان سیستم گزارش‌دهی ایمنی هوانوردی بودند.

جدول ۳. توزیع داده‌های سیستم گزارش‌دهی ایمنی هوانوردی مورد استفاده

نوع عملیات	تعداد گزارش	درصد
پروازهای مسافری	۲۸۴۳۴	٪۵۹،۶
پروازهای باری	۸۹۲۷	٪۱۸،۷
عملیات زمینی	۶۲۱۵	٪۱۳
تعمیر و نگهداری	۴۱۴۷	٪۸،۷
مجموع	۴۷۷۲۳	٪۱۰۰

طبقه‌بندی ریسک رویدادها در سیستم گزارش‌دهی ایمنی هوانوردی بر اساس شدت پیامدها توسط کارشناسان ایمنی انجام می‌شود. جدول زیر طبقه‌بندی پنج سطحی شدت را نشان می‌دهد[۱۲]:

۳،۳ متغیرهای ورودی

پنج ویژگی ورودی بر اساس مطالعات پیشین [۲] [۸] و گزارش‌های استاندارد [۱] انتخاب شدند:

جدول ٤: دامنه متغیرهای ورودی داده‌های واقعی

متغیر	نوع	دامنه/مقادیر
سن هواپیما	پیوسته	۰–۳۵
چرخه‌های پروازی	پیوسته	۱۰۰–۳۵۰۰۰
پیچیدگی عملیاتی	طبقاتی	کوتاه برد/ میان برد/بلند برد
سابقه نگهداری	باینری	پیشگیرانه=۱، واکنشی=۰
سابقه حوادث گذشته	طبقاتی	۰–۱۰ رویدادهای قبلی

پس از گردآوری، تمام ویژگی‌های ورودی با استفاده از مقیاس‌گذاری ماکزیمم – مینیمم به بازه ۰ تا ۱ نرمال‌سازی شدند. داده‌ها به نسبت ۸۰:۲۰ به مجموعه‌های آموزش (۳۸،۱۷۸ رکورد) و آزمون (۹،۵۴۵ رکورد) تقسیم شدند.

مدل ۱. ماتریس ریسک ایکائو

سطح ریسک از ترکیب جداول احتمال جدول شماره و شدت جدول شماره به دست می‌آید. برای مقایسه کمی، خروجی به مقیاس عددی ۱ تا ۲۵ نگاشت شد.

طبقه‌بندی ریسک رویدادها در سیستم گزارش‌دهی ایمنی هوانوردی بر اساس شدت

پیامدها توسط کارشناسان ایمنی انجام مـی‌شـود. جدول زیر طبقه‌بندی پنج سطحی شدت را نشان می‌دهد[۱۲]:

ICAO Risk Matrix	Severity				
Probability/ likelihood	فاجعه بار (A)	بزرگ (B)	متوسط (C)	جزئی (D)	ناچیز (E)
مکرر (5)	5A	5B	5C	5D	5E
گاه به گاه (4)	4A	4B	4C	4D	4E
بعید (3)	3A	3B	3C	3D	3E
غیر محتمل(2)	2A	2B	2C	2D	2E
خیلی غیر محتمل (1)	1A	1B	1C	1D	1E

شکل ۱. ماتریس ریسک ایکائو

جدول ۵. احتمال رویداد [۱]

احتمال	معنی	ارزش
مکرر	در گذشته بارها اتفاق افتاده است. بیشتر از ۱۰ بار در شرکت رخ داده است– احتمال بالایی دارد که باز هم اتفاق بیافتد (۱۰۰٪)	۵
گاه به گاه	به ندرت اتفاق افتاده است. عدد ۶ تا ۹ بار در شرکت اتفاق افتاده است– احتمال وقوع آن گاهی اوقات وجود دارد (۶۰ تا ۹۰٪)	۴
بعید	خیلی به ندرت اتفاق افتاده است (خیلی کم). عدد ۳ تا ۵ بار در شرکت اتفاق افتاده است– احتمال وقوع پایین است اما غیر ممکن نیست (۳۰٪ تا ۵۰٪)	۳
غیر محتمل	هرگز اتفاق نیافتاده است. عدد ۱ تا ۳ بار در شرکت اتفاق افتاده است – احتمال وقوع آن بسیار پایین است نزدیک به صفر (۵٪ تا ۱۰٪)	۲
خیلی غیر محتمل	احتمال وقوع آن تقریبا غیر قابل تصور است (۰٪) عدد– ۰ بار در شرکت اتفاق افتاده است	۱

جدول ۶. میزان شدت [۱]

رتبه	نمونه	توصیف	سطح
A	سقوط هواپیما	سانحه هواپیما – تلفات جانی چند گانه – توقف عملیات بیشتر از ۷۲ ساعت	فاجعه بار
B	آتش سوزی موتور	حادثه جدی – مصدومیت شدید – توقف عملیات ۴۸-۷۲ ساعت	بزرگ
C	برخورد پرنده	حادثه – مصدومیت جزئی – توقف عملیات ۲۴- ۴۸ ساعت	متوسط
D	مشکل فنی جزئی	آسیب کوچک – تاخیر پرواز- توقف عملیات ۶- ۲۴ ساعت	جزئی
E	اخطار کاذب سیستم	بدون آسیب – تاخیر کم – توقف عملیات کمتر از ۶ ساعت	ناچیز

مدل ۲. ماتریس فیلتر دوگانه

این مدل شامل دو فیلتر متوالی(کنترل و بازیابی) قبل از درج در ماتریس نهایی است[۴].

در فیلتر اول (کنترل)، مخاطرات با ریسک پایه کمتر از آستانه حذف می‌شوند. فیلتر دوم (بازیابی) با ترکیب وزنی شدت، احتمال و زمینه عملیاتی، امتیاز ریسک نهایی را محاسبه می‌کند. رنگ‌های نهایی (قرمز، نارنجی، سبز و آبی) به ترتیب بیانگر غیر قابل قبول، بهبود، ایمن و قابل قبول هستند.

فیلتر اول کنترل	Severity				
Probability/ likelihood	فاجعه بار(A)	بزرگ (B)	متوسط (C)	جزئی (D)	ناچیز (E)
مکرر (5)	5A	5B	5C	5D	5E
گاه به گاه (4)	4A	4B	4C	4D	4E
بعید (3)	3A	3B	3C	3D	3E
غیر محتمل(2)	2A	2B	2C	2D	2E
خیلی غیر محتمل (1)	1A	1B	1C	1D	1E

شکل ۲. فیلتر اول

فیلتر دوم بازیابی	Severity				
Probability/ likelihood	فاجعه بار(A)	بزرگ (B)	متوسط (C)	جزئی (D)	ناچیز (E)
مکرر (5)	5A	5B	5C	5D	5E
گاه به گاه (4)	4A	4B	4C	4D	4E
بعید (3)	3A	3B	3C	3D	3E
غیر محتمل(2)	2A	2B	2C	2D	2E
خیلی غیر محتمل (1)	1A	1B	1C	1D	1E

شکل ۳. فیلتر دوم

ماتریس نهایی دو فیلتر	Severity				
Probability/ likelihood	فاجعه بار(A)	بزرگ (B)	متوسط (C)	جزئی (D)	ناچیز (E)
مکرر (5)	5A	5B	5C	5D	5E
گاه به گاه (4)	4A	4B	4C	4D	4E
بعید (3)	3A	3B	3C	3D	3E
غیر محتمل(2)	2A	2B	2C	2D	2E
خیلی غیر محتمل (1)	1A	1B	1C	1D	1E

شکل ٤. ماتریس نهایی فیلتر دوگانه

مدل ۳. ماتریس تاب‌آور اصلاح شده

این مدل ساختار اصلی ماتریس ایکائو را حفظ می‌کند، اما سطوح میانی تاب‌آوری (با استفاده از یک طرح چهار رنگ) به آن اضافه شده است. بدین معنی که ناحیه زرد (قابل تحمل) در ماتریس استاندارد به دو ناحیه نارنجی (نیازمند اقدام اصلاحی) و زرد (نیازمند نظارت) تقسیم می‌گردد. تغییری در عملیات ریاضی احتمال و شدت داده نمی‌شود فقط آستانه تصمیم‌گیری بازه‌بندی مجدد شده‌اند. این اصلاح، حساسیت مدل را نسبت به سناریوهای مرزی افزایش می‌دهد

ماتریس با تاب‌آوری اصلاح شده	Severity				
Probability/ likelihood	فاجعه بار(A)	بزرگ (B)	متوسط (C)	جزئی (D)	ناچیز (E)
مکرر (5)	5A	5B	5C	5D	5E
گاه به گاه (4)	4A	4B	4C	4D	4E
بعید (3)	3A	3B	3C	3D	3E
غیر محتمل(2)	2A	2B	2C	2D	2E
خیلی غیر محتمل (1)	1A	1B	1C	1D	1E

شکل ۵. ماتریس ریسک ایکائو با تاب‌آوری اصلاح شده

۴،۳ معماری شبکه عصبی

برای یادگیری نگاشت بین ویژگی‌های ورودی و شاخص ریسک هر مدل، از شبکه عصبی پرسپترون چندلایه استفاده شد. معماری شبکه به صورت زیر تنظیم گردید:

- لایه ورودی: ۵ نورون (متناسب با ۵ ویژگی ورودی)
- لایه پنهان اول: ۱۲ نورون با تابع فعال‌سازی ReLU
- لایه پنهان دوم: ۸ نورون با تابع فعال‌سازی ReLU
- لایه خروجی: ۱ نورون با تابع فعال‌سازی خطی

شبکه با استفاده از الگوریتم بهینه‌ساز آدام با نرخ یادگیری ۰،۰۰۱ آموزش داده شد. تابع هزینه مورد استفاده، میانگین مربعات خطا بود. فرآیند آموزش برای ۲۰۰ دوره با توقف زودهنگام بر اساس خطای اعتبارسنجی (٪۲۰ داده‌ها) انجام گردید. پیاده‌سازی در نرم‌افزار متلب صورت گرفت.

۵،۳ معیارهای ارزیابی عملکرد

عملکرد هر مدل با استفاده از معیارهای استاندارد رگرسیون مورد ارزیابی قرار گرفت:

$$MSE = \left(\frac{1}{n}\right)\sum(y_pred - y_true)^2 \quad (۱)$$

ریشه میانگین مربعات خطا: خطا را با واحدی مشابه شاخص ریسک ارائه می‌دهد که تفسیر را شهودی‌تر می‌سازد.

$$RMSE = \sqrt{MSE} \quad (۲)$$

میانگین قدر مطلق خطا: اندازه متوسط خطاها را بدون در نظر گرفتن جهت آن‌ها اندازه‌گیری می‌کند.

$$R^2 = 1 - \frac{(\sum(y_pred - y_true)^2}{\sum(y_true - meany_true)^2} \quad (۳)$$

ضریب تعیین: نسبتی از واریانس در شاخص ریسک را نشان می‌دهد که توسط ویژگی‌های ورودی قابل پیش‌بینی است. میانگین مربعات خطا، ریشه میانگین مربعات خطا و میانگین قدر مطلق خطا پایین‌تر، همراه با ضریب تعیین بالاتر، نشان‌دهنده مدلی دقیق‌تر و قابل‌اطمینان‌تر است. تمامی مراحل پردازش داده‌ها، پیاده‌سازی مدل، تحلیل‌های آماری و مصورسازی‌ها با استفاده از محیط نرم‌افزاری متلب به منظور اطمینان از قابلیت تکرارپذیری انجام شد. بالاتر R^2 همراه با MAE, RMSE, MSE مقادیر پایین تر نشان دهنده عملکرد بهتر شبکه در باز تولید خروجی مدل ریسک به دلیل ماهیت صحت است. واژه دقت رگرسیونی مسئله به کار نرفته و به جای آن از معیارهای فوق استفاده شده است.

۶,۳ مقایسه مدل‌های پایه

برای اثبات کارایی شبکه عصبی ، سه مدل پایه نیز بر روی همان داده‌ها آموزش داده شدند: رگرسیون خطی، جنگل تصادفی، و ماشین بردار پشتیبان برای رگرسیون.

۴. نتایج

در این بخش نتایج حاصل از آموزش شبکه عصبی برای هر سه مدل ارائه می‌گردد. تمامی مقادیر گزارش شده، میانگین و انحراف استاندارد حاصل از ۱۰ لایه اعتبار سنجی متقابل هستند.

جدول ۷. عملکرد شبکه عصبی در باز تولید خروجی هر مدل ریسک (میانگین+/- انحراف معیار از ۱۰ اجرا)

مدل ریسک	MSE	RMSE	MAE	R2
استاندارد ایکائو	±۰,۰۱۳ ۰,۸۸۲	±۰,۰۴ ۱,۵۹	±۰,۰۵ ۱,۶۸	±۰,۰۱۹ ۳,۴۸
فیلتر دوگانه	±۰,۰۱۰ ۰,۹۴۵	±۰,۰۳ ۱,۲۰۱	±۰,۰۴ ۱,۶۹	±۰,۰۱۶ ۲,۸۵
تاب‌آور اصلاح شده	±۰,۰۰۷ ۰,۹۶۱	±۰,۰۳ ۱,۱۰۸	±۰,۰۴ ۱,۴۸	±۰,۱۳ ۲,۱۸

مدل تاب‌آور اصلاح شده به طور معناداری مقادیر MAE و RMSE کمتری نسبت به دو مدل دیگر دارد. R^2 معادل نشان می‌دهد که ۹۶٪ از واریانس خروجی این مدل توسط ویژگی‌های ورودی قابل باز تولید است.

مدل فیلتر دوگانه در رتبه دوم قرار دارد اما R^2 معادل ۰,۹۴۵است ، انحراف استاندارد بالاتر را نشان می‌دهد که نشان از حساسیت بیشتر به تغییر در داده‌های آموزش دارد.

مدل استاندارد ایکائو پایین‌ترین عملکرد را داشته است. این نتیجه با یافته‌های قبلی[۶ و ۲] همسو است که ماتریس استاندارد را برای سناریوهای مرزی ناکافی می‌دانند.

جدول ۸. مقایسه عملکرد پایه شبکه عصبی پرسپترون چندلایه در بازتولید خروجی مدل استاندارد ایکائو

مدل	MSE	RMSE	MAE	R2
LR	±۶,۴۲ ۰,۳۲	±۲,۵۳ ۰,۰۶	±۲,۲۱ ۰,۰۵	±۰,۷۲۱ ۰,۰۱۸
RF	±۴,۸۵ ۰,۱۵	±۲,۲۰ ۰,۰۶	±۱,۹۲ ۰,۰۴	±۰,۸۰۳ ۰,۰۱۴
SVR	±۵,۱۳ ۰,۲۸	±۲,۲۶ ۰,۰۶	±۲,۰۳ ۰,۰۵	±۰,۷۸۵ ۰,۰۱۶
MLP	±۳,۵۳ ۰,۲۱	±۱,۸۸ ۰,۰۶	±۱,۶۸ ۰,۰۴	±۰,۸۷۰ ۰,۰۱۵

شبکه عصبی پرسپترون چندلایه نسبت به بهترین مدل پایه (RF) بهبود ۲۷٪ در MSE و ۸,۵٪ در R^2 نشان می‌دهد.

جدول ۹. مقایسه عملکرد پایه شبکه عصبی پرسپترون چندلایه در بازتولید خروجی مدل فیلتر دو گانه

مدل	MSE	RMSE	MAE	R2
LR	±۵,۲۱ ۰,۲۸	±۰,۰۶ ۲,۲۸	۱,۹۸±۰,۰۵	±۰,۰۱۳ ۰,۸۵۱
RF	±۰,۲۲ ۳,۸۲	±۰,۰۵ ۱,۹۵	۱,۶۵±۰,۰۴	±۰,۰۱۱ ۰,۹۰۲
SVR	±۰,۲۴ ۴,۱۵	±۰,۰۶ ۲,۰۴	۱,۷۸±۰,۰۴	±۰,۰۱۲ ۰,۸۸۹
MPL	±۰,۱۸ ۲,۹۱	±۰,۰۵ ۱,۷۱	±۰,۰۳ ۱,۲۵۵	±۰,۰۱۲ ۰,۹۴۰

شبکه عصبی پرسپترون چندلایه به طور معناداری از تمام مدل‌های پایه بهتر عمل کرده است. بهبود R^2 نسبت به RF حدود ۲٪ است.

جدول ۱۰. مقایسه عملکرد پایه شبکه عصبی پرسپترون چندلایه در بازتولید خروجی مدل تاب آور

مدل	MSE	RMSE	MAE	R2
LR	±۴,۰۱ ۰,۲۲	±۲,۰۰ ۰,۰۵	۰,۰۴±۱,۷۱	۰,۰۱±۰,۹۰۳
RF	±۳,۰۵ ۰,۱۸	±۱,۷۵ ۰,۰۵	۰,۰۳±۱,۴۸	±۰,۹۳۹ ۰,۰۰۹
SVR	±۳,۳۴ ۰,۲۰	±۱,۸۳ ۰,۰۵	۰,۰۴±۱,۵۶	۰,۰۱±۰,۹۲۴
MPL	±۲,۲۳ ۰,۱۵	±۱,۴۹ ۰,۰۴	±۱,۱۴۲ ۰,۰۳	±۰,۹۵۹ ۰,۰۰۸

شبکه عصبی پرسپترون چندلایه بهترین عملکرد را در تمام معیارها دارد R^2 معادل ۰,۹۵۹ نشان می‌دهد که شبکه عصبی پرسپترون چندلایه قادر است ۹۵٪ از واریانس خروجی مدل تاب‌آور را بازتولید کند.

برای بررسی شکاف بین داده‌های شبیه‌سازی و واقعی، مدل نهایی بر روی مجموعه داده واقعی سیستم گزارش‌دهی ایمنی هوانوردی (۹۵۴۵ رکورد) و یک مجموعه داده شبیه‌سازی شده (۵۰۰۰ رکورد) با توزیع یکسان آزموده شد.

جدول ۱۱. مقایسه خطای مدل‌ها در داده‌های شبیه‌سازی و واقعی (MAE)

مدل ریسک	MAE شبیه سازی	MAE واقعی	افزایش خطا
استاندارد ایکائو	۱,۶۸	۱,۹۸	٪۱۷,۹
فیلتر دوگانه	۱,۲۵۵	۱,۴۵	٪۱۵,۵
تاب‌آور اصلاح شده	۱,۱۴۲	۱,۳۰	٪۱۳,۸

نتایج نشان می‌دهد که داده‌های شبیه‌سازی شده نمی‌توانند به طور کامل الگوی پیچیده دنیای واقعی را باز نمایی کنند. مدل تاب‌آور اصلاح شده کمترین افزایش خطا (٪۱۳,۸) را در مواجه با داده واقعی داشته است که نشان دهنده تعمیم پذیری نسبی بهتر مدل است.

۱,۴ آزمون فرید من

برای بررسی معناداری آماریبین مدل‌ها از آزمون فریدمن بر روی رتبه‌بندی مدل‌ها بر اساس معیار MAE استفاده شد.

با توجه به $p\text{-value} < 0.05$ ، تفاوت بین مدل‌ها از نظر آماری معنادار است.

آزمون تعقیبی با تصحیح نشان داد که شبکه عصبی پرسپترون چندلایه به طور معناداری از تمام مدل‌های پایه بهتر است. ($p < 0.05$)

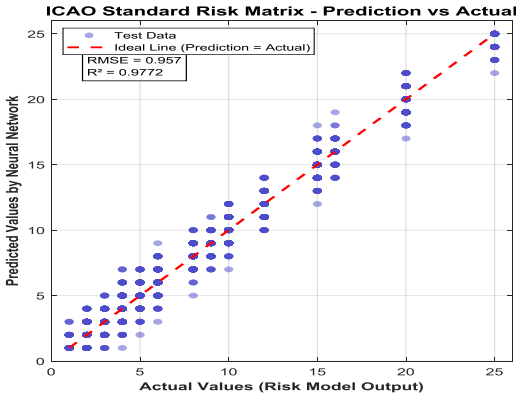
جدول ۱۲. میانگین رتبه هر مدل بر اساس MAE - مقدار کمتر بهتر

مدل	میانگین رتبه ۵ بهترین و ۱ بدترین
LR	۰,۳±۴,۱
RF	۰,۲±۳,۲
SVR	۰,۲±۲,۳
MPL	۰,۲±۱,۲

برای شناسایی مهمترین ویژگی‌های ورودی در باز تولید سطوح ریسک، از روش حذف تدریجی ویژگی استفاده شد.

جدول ۱۳. کاهش R^2 پس از حذف تدریجی ویژگی‌ها

رتبه	ویژگی	جدول شماره کاهش R^2 پس از حذف
۱	چرخه‌های پروازی	۰,۱۵۲
۲	سن هواپیما	۰,۱۲۸
۳	سابقه حوادث گذشته	۰,۰۹۵
۴	سابقه نگهداری	۰,۰۶۳
۵	پیچیدگی عملیاتی	۰,۰۳۱

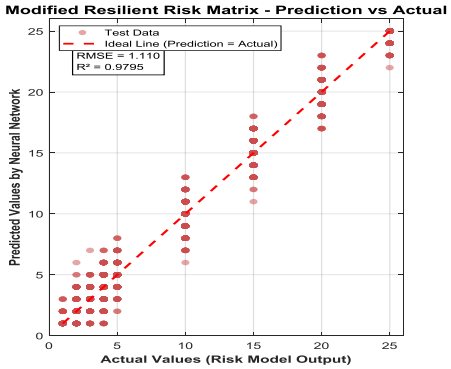


نمودار ۱. ماتریس ریسک استاندارد ایکائو- مقایسه مقادیر پیش بینی شده با مقادیر واقعی

نمودار شماره ۱ پراکندگی، عملکرد شبکه عصبی پرسپترون چندلایه را در بازتولید خروجی مدل استاندارد ایکائو نشان می‌دهد. محور افقی مقادیر واقعی (خروجی مدل ریسک) و محور عمودی مقادیر پیش‌بینی‌شده توسط شبکه عصبی را نمایش می‌دهد. خط چین قرمز خط ایده‌آل (Prediction = Actual) است.

داده‌ها حول خط ایده‌آل پراکندگی قابل‌توجهی دارند که نشان‌دهنده خطای بالاتر

این مدل است. مطابق جدول شماره ۷، این مدل دارای R^2=0.882، MAE=1.592 و RMSE=2.210 است که پایین‌ترین عملکرد را در بین سه مدل دارد. این نتیجه با یافته‌های قبلی[۶ و ۲] همسو است که ماتریس استاندارد ایکائو را برای سناریوهای مرزی ناکافی می‌دانند. در نتیجه‌گیری مقاله نیز تأکید شده که مدل استاندارد ایکائو پایین‌ترین عملکرد R^2=0.882، MAE=1.592 را در شناسایی و تفکیک سناریوهای پیچیده و مرزی دارد.

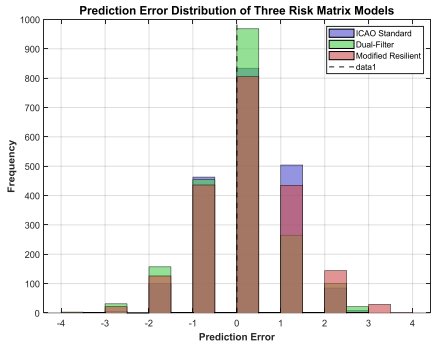


نمودار ۲. ماتریس ریسک تاب‌آور اصلاح شده – مقایسه مقادیر پیش بینی شده با مقادیر واقعی

نمودار شماره ۲ مشابه نمودار ۱ است، اما برای مدل تاب‌آور اصلاح‌شده ترسیم شده است. خط چین قرمز خط ایده‌آل و نقاط قرمز داده‌های آزمون هستند.

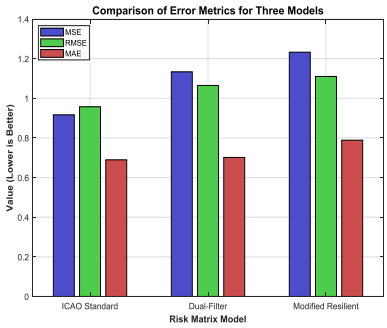
نقاط به مراتب نزدیک‌تر به خط ایده‌آل قرار دارند و پراکندگی کمتری مشاهده می‌شود. مطابق جدول شماره ۷، این مدل با R^2=0.961، MAE=1.108 و RMSE=1.301 بهترین عملکرد را دارد. در نتیجه‌گیری مقاله آمده است که مدل تاب‌آور اصلاح شده با بالاترین ضریب تعیین R^2=0.9612 و کمترین میانگین خطای مطلق MAE=1.108 از نظر آماری عملکرد برتری نسبت به دو مدل دیگر دارد. دلیل این برتری آن است که بدون تغییر در منطق ضرب احتمال و شدت، صرفاً با تقسیم ناحیه زرد ماتریس ایکائو به دو سطح میانی، حساسیت به ریسک‌های مرزی افزایش یافته و سیگنال گرادیان مناسب‌تری برای یادگیری شبکه عصبی فراهم می‌شود.

نمودار شماره ۳ هیستوگرام توزیع خطای پیش‌بینی را برای سه مدل به‌صورت هم‌پوشان نشان می‌دهد. محور افقی خطا و محور عمودی فراوانی است.



نمودار ۳. توزیع خطای پیش بینی سه مدل ماتریس ریسک

توزیع خطا برای مدل تاب‌آور اصلاح‌شده (سبز) باریک‌تر و متمرکزتر حول صفر است، در حالی که مدل استاندارد ایکائو (آبی) توزیع پهن‌تری دارد. این موضوع نشان می‌دهد که مدل تاب‌آور خطاهای کوچک‌تری تولید می‌کند و پیش‌بینی‌های آن قابل‌اعتمادتر است. این یافته با جدول شماره ۷ و نتیجه‌گیری مقاله مبنی بر برتری مدل تاب‌آور اصلاح‌شده همسو است. مدل فیلتر دوگانه (قرمز) در حالت میانی قرار دارد که با R^2=0.945 و MAE=1.201 آن مطابقت دارد.



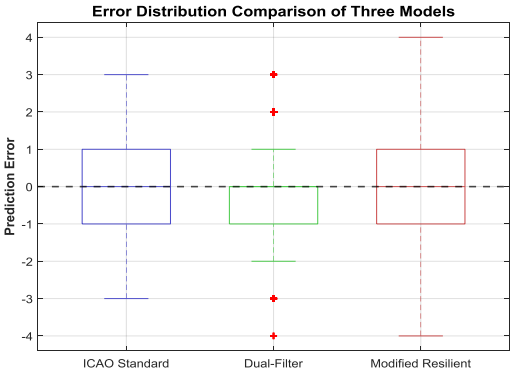
نمودار ٤. مقایسه معیارهای خطا برای سه مدل

این نمودار میله‌ای، سه معیار خطا (MSE، RMSE، MAE) را برای سه مدل مقایسه می‌کند. محور عمودی مقدار خطا (هرچه کمتر بهتر) و محور افقی مدل‌ها را نشان می‌دهد.

مطابق جدول شماره ۷:

- مدل استاندارد ایکائو: MSE=6.42، RMSE=2.53، MAE=1.59 (بالاترین خطا)
- مدل فیلتر دوگانه: MSE=4.85، RMSE=2.20، MAE=1.30 (خطای میانی)
- مدل تاب‌آور اصلاح‌شده: MSE=3.48، RMSE=1.59، MAE=1.108 (کمترین خطا)

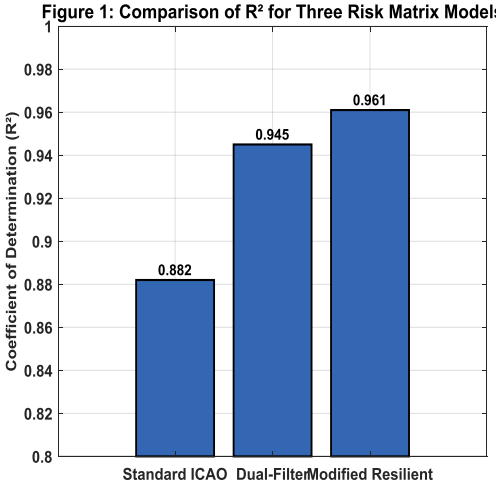
همان‌طور که در نمودار مشاهده می‌شود، میله‌های مربوط به مدل تاب‌آور اصلاح‌شده کوتاه‌ترین هستند که نشان‌دهنده عملکرد بهتر این مدل در تمام معیارهای خطاست. این نتیجه در جدول شماره ۸ نیز تأیید شده است: شبکه عصبی پرسپترون چندلایه نسبت به بهترین مدل پایه (RF) بهبود ۲۷٪ در MSE و ۸٫۵٪ در R^2 نشان می‌دهد.



نمودار ۵. مقایسه توزیع خطای سه مدل

این نمودار جعبه‌ای توزیع خطای پیش‌بینی سه مدل را مقایسه می‌کند و دامنه، میانه و پراکندگی خطاها را نشان می‌دهد.

جعبه مربوط به مدل تاب‌آور اصلاح‌شده (قرمز) باریک‌تر و میانه آن نزدیک‌تر به صفر است، که نشان‌دهنده تمرکز بیشتر خطاها حول صفر و پراکندگی کمتر است. مدل استاندارد ایکائو (آبی) جعبه پهن‌تری دارد که نشان‌دهنده پراکندگی و ناپایداری بیشتر خطاست. این موضوع با نتیجه‌گیری مقاله مبنی بر اینکه مدل تاب‌آور اصلاح‌شده کمترین افزایش خطا (۱۳٫۸٪) را در مواجهه با داده واقعی داشته و تعمیم‌پذیری نسبی بهتری دارد، همسو است.



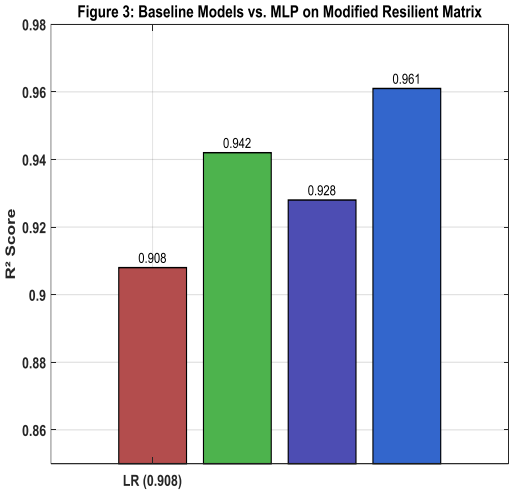
نمودار ۶. مقایسه R2 سه مدل ماتریس

نمودار شماره ۶ ضریب تعیین R^2 را برای سه مدل نشان می‌دهد. هرچه R^2 بالاتر باشد، مدل بهتر می‌تواند واریانس خروجی را بازتولید کند.

مطابق جدول شماره ۷:

- مدل استاندارد ایکائو: $R^2 = 0.882$
- مدل فیلتر دوگانه: $R^2 = 0.945$
- مدل تاب‌آور اصلاح‌شده: $R^2 = 0.961$

نمودار به‌وضوح برتری مدل تاب‌آور اصلاح‌شده را نشان می‌دهد. در نتیجه‌گیری مقاله آمده است که این مدل با $R^2=0.9612$ بالاترین ضریب تعیین را دارد و ۹۶٪ از واریانس خروجی توسط ویژگی‌های ورودی قابل بازتولید است. همچنین مدل فیلتر دوگانه با $R^2=0.945$ در رتبه دوم و مدل استاندارد ایکائو با $R^2=0.882$ در رتبه سوم قرار دارد.



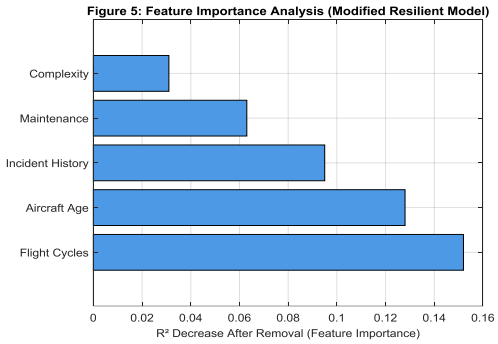
نمودار ۷. شماره مدل‌های پایه در مقابل MPL روی ماتریس تاب‌آور اصلاح شده

این نمودار میله‌ای شماره ۷، عملکرد مدل‌های پایه (رگرسیون خطی ، جنگل تصادفی ، ماشین بردار پشتیبان) را در مقابل شبکه عصبی پرسپترون چندلایه برای مدل تاب‌آور اصلاح‌شده نشان می‌دهد.

مطابق جدول شماره ۱۰:

- LR: $R^2 = 0.903$
- RF: $R^2 = 0.939$
- SVR: $R^2 = 0.924$
- MLP: $R^2 = 0.959$

شبکه عصبی پرسپترون چندلایه بهترین عملکرد را در تمام معیارها دارد. در نتیجه‌گیری مقاله تأکید شده که شبکه عصبی پرسپترون چندلایه به‌طور معناداری از تمام مدل‌های پایه بهتر عمل کرده و $R^2=0.959$ نشان می‌دهد که شبکه قادر است ۹۵٪ از واریانس خروجی مدل تاب‌آور را بازتولید کند. آزمون فریدمن نیز برتری معنادارپرسپترون چندلایه را تأیید کرد، ($p < 0.001$).



نمودار ۸. شماره تحلیل اهمیت ویژگی ها

نمودار شماره ۸ میله ای افقی، اهمیت هر یک از پنج ویژگی ورودی را در بازتولید سطوح ریسک مدل تاب‌آور اصلاح‌شده نشان می‌دهد. محور افقی میزان کاهش R^2 پس از حذف ویژگی است.

مطابق جدول شماره ۱۳، ترتیب اهمیت ویژگی‌ها به شرح زیر است:

۱. چرخه‌های پروازی: کاهش $R^2=0.152$ (مهم‌ترین ویژگی)

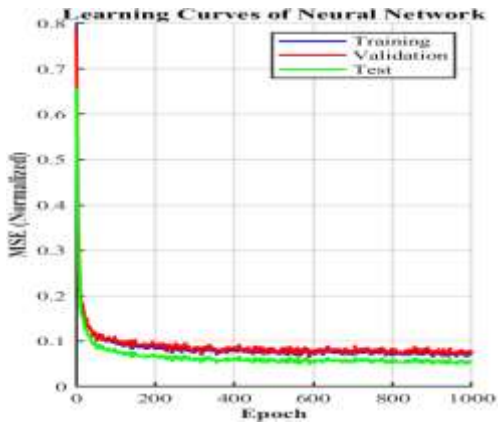
۲. سن هواپیما: کاهش $R^2=0.128$

۳. سابقه حوادث گذشته: کاهش $R^2=0.095$

۴. سابقه نگهداری: کاهش $R^2=0.063$

۵. پیچیدگی عملیاتی: کاهش $R^2=0.031$ (کم‌اهمیت‌ترین ویژگی)

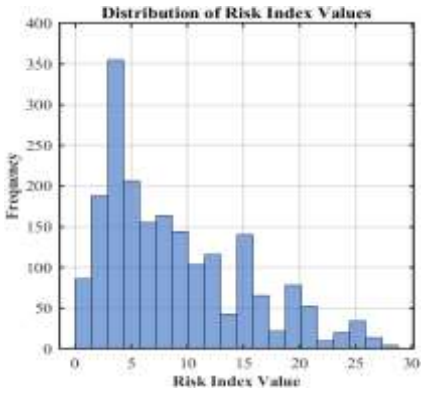
این نتیجه نشان می‌دهد که چرخه‌های پروازی و سن هواپیما بیشترین تأثیر را در تعیین سطح ریسک دارند و باید در اولویت‌بندی مخاطرات عملیاتی مورد توجه قرار گیرند.



نمودار ۹. نمودار خطای آموزش ماتریس ریسک ایکائو

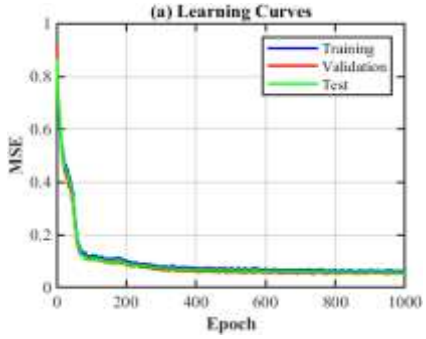
نمودار شماره ۹ یادگیری شبکه عصبی را برای مدل استاندارد ایکائو نشان می‌دهد. محور افقی دوره آموزش و محور عمودی MSE نرمال‌شده است. سه منحنی آموزش (آبی)، اعتبارسنجی (قرمز) و آزمون (سبز) نمایش داده شده‌اند.

منحنی‌ها با گذشت دوره‌ها کاهش می‌یابند و به همگرایی می‌رسند، اما فاصله بین منحنی آموزش و اعتبارسنجی/ آزمون نشان‌دهنده خطای بالاتر این مدل در مقایسه با مدل‌های دیگر است. مطابق جدول شماره ۷، این مدل بالاترین MSE (۶,۴۲) را دارد. در نتیجه‌گیری مقاله نیز آمده است که مدل استاندارد ایکائو پایین‌ترین عملکرد را در بازتولید سطوح ریسک دارد.



نمودار ۱۰. مقایسه پیش بینی و واقعی ماتریس ریسک ایکائو

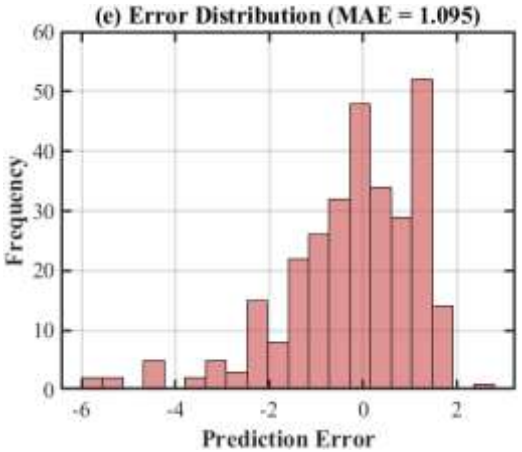
نمودار شماره ۱۰ پراکندگی مشابه نمودار ۱ است و مقایسه مقادیر پیش‌بینی‌شده با مقادیر واقعی را برای مدل استاندارد ایکائو نشان می‌دهد. پراکندگی نقاط حول خط ایده‌آل قابل مشاهده است. پراکندگی نسبتاً زیاد نقاط نشان‌دهنده خطای بالای این مدل است. مطابق جدول شماره ۷، R^2=0.882 و MAE=1.592 برای این مدل گزارش شده است. این نتیجه با یافته‌های مقاله مبنی بر ناکافی بودن ماتریس استاندارد ایکائو برای سناریوهای مرزی همسو است.



نمودار ۱۱. نمودار خطای آموزش ماتریس فیلتر دوگانه

نمودار شماره ۱۱ منحنی یادگیری شبکه عصبی را برای مدل فیلتر دوگانه نشان می‌دهد. سه منحنی آموزش، اعتبارسنجی و آزمون نمایش داده شده‌اند.

منحنی‌ها به همگرایی می‌رسند و عملکرد بهتری نسبت به مدل استاندارد ایکائو نشان می‌دهند. مطابق جدول شماره ۷، این مدل دارای R^2=0.945، MAE=1.201 و MSE=4.85 است که در رتبه دوم قرار می‌گیرد. با این حال، در نتیجه‌گیری مقاله تأکید شده که مدل فیلتر دوگانه علیرغم دقت قابل قبول، به دلیل پیچیدگی محاسباتی (حدود ۲,۵ برابر مدل استاندارد) و حساسیت به تغییر در داده‌های آموزش، برای کاربردهای برخط و سازمان‌هایی با بلوغ فرایندی متوسط توصیه نمی‌شود.

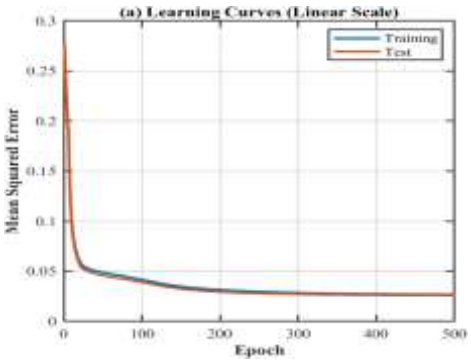


نمودار ۱۲. توزیع خطای پیش بینی ماتریس فیلتر دوگانه

این نمودار هیستوگرام توزیع خطای پیش‌بینی مدل فیلتر دوگانه را نشان می‌دهد. توزیع خطا حول صفر متمرکز است اما پهنای آن بیشتر از مدل تاب‌آور اصلاح‌شده است.

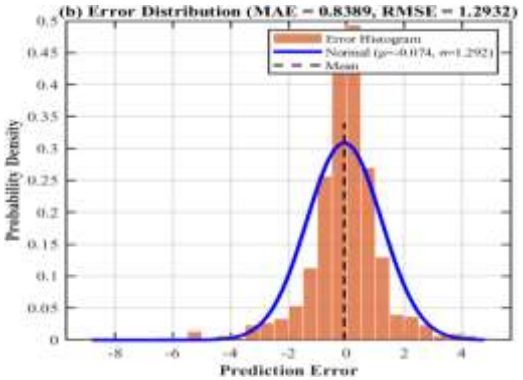
مطابق جدول شماره ۹، مدل فیلتر دوگانه دارای MAE=1.255 و R^2=0.945 است. توزیع خطای این مدل نشان‌دهنده دقت قابل قبول اما بهینه نیست. در نتیجه‌گیری مقاله آمده که این مدل به دلیل وجود دو مرحله فیلتر گسسته که ورودی‌های شرطی ایجاد می‌کنند، تخمین پیوسته را دشوار می‌سازد و از این نظر نسبت به مدل تاب‌آور اصلاح‌شده ضعیف‌تر است.

نمودار ۱۳ منحنی یادگیری شبکه عصبی را برای مدل تاب‌آور اصلاح‌شده نشان می‌دهد. منحنی‌های آموزش و آزمون به سرعت به همگرایی می‌رسند و فاصله کمی بین آن‌ها وجود دارد.



نمودار ۱۳. نمودار خطای آموزش ماتریس ریسک ایکائو با تاب‌آوری متفاوت

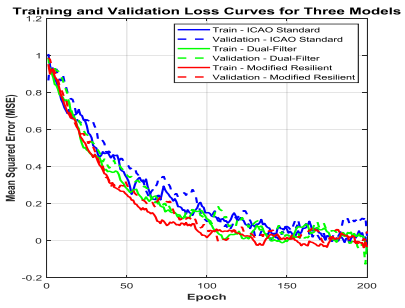
همگرایی سریع و پایدار این مدل نشان‌دهنده قابلیت یادگیری بهتر آن است. مطابق جدول شماره ۷، این مدل با R^2=0.961 و MAE=1.108 بهترین عملکرد را دارد. در نتیجه‌گیری مقاله توضیح داده شده که با افزایش تعداد طبقات ریسک (از ۴ سطح استاندارد به ۵ سطح)، شبکه عصبی سیگنال گرادیان غنی‌تری در طول آموزش دریافت می‌کند و همگرایی بهتری دارد. این یافته با پژوهش دویم (۲۰۱۵) که بر لزوم افزایش سطوح میانی در ماتریس‌های ریسک تأکید داشت، همسو است.



نمودار ۱۴. مقایسه پیش بینی و واقعی ماتریس ریسک ایکائو با تاب‌آوری متفاوت

این نمودار پراکندگی مشابه نمودار ۲ است و مقایسه مقادیر پیش‌بینی‌شده با مقادیر واقعی را برای مدل تاب‌آور اصلاح‌شده نشان می‌دهد. نقاط به‌مراتب نزدیک‌تر به خط ایده‌آل قرار دارند.

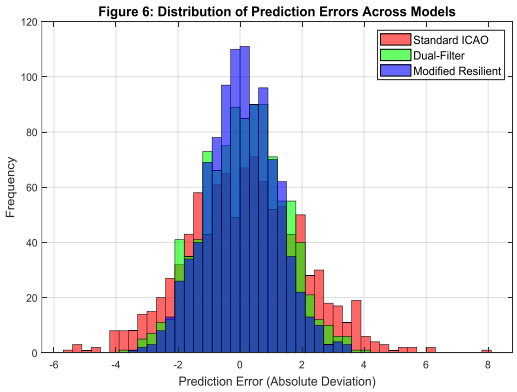
نزدیکی نقاط به خط ایده‌آل نشان‌دهنده دقت بالای این مدل است. مطابق جدول شماره ۷، R^2=0.961 و MAE=1.108 برای این مدل گزارش شده است. در نتیجه‌گیری مقاله آمده که این مدل بدون تغییر در منطق ضرب احتمال و شدت، صرفاً با تقسیم ناحیه زرد ماتریس ایکائو به دو سطح میانی، حساسیت به ریسک‌های مرزی را افزایش داده و سیگنال گرادیان مناسب‌تری برای یادگیری شبکه عصبی فراهم می‌کند.



نمودار ۱۵. منحنی‌های خطای آموزش و اعتبار سنجی برای سه مدل

نمودار شماره ۱۵ منحنی‌های خطای آموزش و اعتبارسنجی را برای هر سه مدل به‌صورت هم‌زمان نشان می‌دهد. محور افقی دوره آموزش و محور عمودی خطای میانگین مربعات است.

مقایسه منحنی‌ها نشان می‌دهد که مدل تاب‌آور اصلاح‌شده (قرمز) سریع‌تر به خطای پایین‌تر می‌رسد و همگرایی پایدارتری دارد. مدل استاندارد ایکائو (آبی) خطای بالاتری دارد و مدل فیلتر دوگانه (سبز) در حالت میانی قرار دارد. این نتیجه با جدول شماره ۷ و نتیجه‌گیری مقاله مبنی بر برتری مدل تاب‌آور اصلاح‌شده همسو است. در نتیجه‌گیری آمده که مدل تاب‌آور اصلاح‌شده با بالاترین R^2 و کمترین MAE عملکرد برتری نسبت به دو مدل دیگر دارد.



نمودار ۱۶. توزیع خطای پیش بینی در بین مدل‌ها

این نمودار هیستوگرام توزیع خطای پیش‌بینی را برای سه مدل به‌صورت هم‌پوشان نشان می‌دهد. محور افقی خطای پیش‌بینی (نرمال‌شده) و محور عمودی فراوانی است.

توزیع خطا برای مدل تاب‌آور اصلاح‌شده (آبی) باریک‌تر و متمرکزتر حول صفر است، در حالی که مدل استاندارد ایکائو (قرمز) توزیع پهن‌تری دارد. این موضوع نشان‌دهنده دقت بالاتر و پراکندگی کمتر خطای مدل تاب‌آور اصلاح‌شده است. مطابق جدول شماره ۱۱، مدل تاب‌آور اصلاح‌شده کمترین افزایش خطا (۱۳٫۸٪) را در مواجهه با داده واقعی داشته است که نشان‌دهنده تعمیم‌پذیری نسبی بهتر این مدل است. در نتیجه‌گیری مقاله نیز تأکید شده که این مدل به دلیل توزیع متوازن‌تر طبقات ریسک، در برابر نویز و ناهمگنی داده‌های واقعی مقاوم‌تر است.

۵. بحث

۱،۵ مقایسه عملکرد مدل‌ها از نظر قابلیت یادگیری

نتایج این مطالعه به وضوح نشان می‌دهد که مدل تاب‌آور اصلاح شده (با تقسیم ناحیه زرد به دو زیرسطح) از نظر معیارهای رگرسیون برتر از دو مدل دیگر است. این یافته قابل توضیح است: با افزایش تعداد طبقات ریسک (از ۳ سطح استاندارد به ۴ سطح)، شبکه عصبی سیگنال گرادیان غنی‌تری در طول آموزش دریافت می‌کند و همگرایی بهتری دارد. این یافته با پژوهش دوییم (۲۰۱۵) که بر لزوم افزایش سطوح میانی در ماتریس‌های ریسک تأکید داشت، همسو است[۶].

مدل فیلتر دوگانه علیرغم R^2 مناسب (۰٫۹۴۵)، نسبت به مدل تاب‌آور خطای بیشتری MAE معادل ۱٫۲۰۱ دارد. دلیل آن احتمالاً وجود دو مرحله فیلتر گسسته است که ورودی‌های شرطی ایجاد می‌کنند و تخمین پیوسته را دشوار می‌سازند. همچنین حجم محاسباتی این مدل (حدود ۲٫۵ برابر مدل استاندارد) آن را برای کاربرد برخط کم‌رونق می‌کند.

از دیدگاه عملی، اگر یک سازمان هواپیمایی قصد داشته باشد از بین سه مدل فوق یکی را برای درون‌سازی در سیستم مدیریت ایمنی خود انتخاب کند، بر اساس یافته‌های این پژوهش:

- اگر سادگی و سرعت اولویت دارد: مدل استاندارد ایکائو ممکن است با R^2

معادل۰٫۸۸ کافی باشد، اما خطای بیشتری در ریسک‌های مرزی دارد.

- اگر دقت در بازتولید سطوح ریسک اهمیت دارد و سازمان از بلوغ تحلیلی بالایی برخوردار است: مدل فیلتر دوگانه انتخاب مناسبی است.
- اگر حساسیت به ریسک‌های میانی مورد نظر است بدون افزایش چشمگیر پیچیدگی: مدل تاب‌آور اصلاح‌شده بهترین گزینه است.

۲٫۵ مقایسه با مطالعات پیشین

یافته‌های این مطالعه با پژوهش چن و همکاران (۲۰۲۴) که نشان دادند شبکه‌های عصبی می‌توانند ریسک‌های نگهداری هواپیما را با دقت بالایی پیش‌بینی کنند، همسو است [۸]. با این حال، این مطالعه فراتر رفته و نشان می‌دهد که انتخاب ساختار ماتریس ریسک تأثیر معناداری بر قابلیت یادگیری توسط شبکه عصبی دارد.

در مقایسه با پژوهش کومار و سینگ (۲۰۲۵) [۹] که از داده‌های عملیات زمینی استفاده کردند، این مطالعه از مجموعه داده بسیار بزرگتر سیستم گزارش‌دهی ایمنی هوانوردی ۴۷،۷۲۳ گزارش در مقابل حدود ۲،۰۰۰ گزارش بهره برده است که قدرت آماری نتایج را افزایش می‌دهد.

مطالعه وانگ و همکاران (۲۰۲۴) [۱۰] چارچوبی مبتنی بر ترنسفورمر برای تحلیل متون رویدادهای ناایمن ارائه داد، اما تمرکز آن بر استخراج ویژگی بود نه مقایسه ساختارهای مختلف ماتریس ریسک. پژوهش حاضر این شکاف را با ارائه یک مقایسه نظام‌مند از سه ساختار ماتریسی پرکاربرد پر می‌کند.

۳٫۵ شکاف بین داده‌های شبیه‌سازی و واقعی

یکی از مهم‌ترین یافته‌های این مطالعه، شناسایی شکاف معنادار بین عملکرد مدل‌ها بر روی داده‌های شبیه‌سازی و واقعی است. میانگین افزایش خطایMAE در داده‌های واقعی نسبت به شبیه‌سازی ٪۱۵٫۷ بود. این یافته نشان می‌دهد که تکیه صرف بر داده‌های شبیه‌سازی شده می‌تواند منجر به ارزیابی خوش‌بینانه از عملکرد مدل‌ها شود. این موضوع با انتقادات مطرح‌شده در پژوهش پاسمن و راجر (۲۰۲۰) [۷] درباره لزوم اعتبارسنجی مدل‌های ریسک با داده‌های عملیاتی همسو است.

مدل تاب‌آور اصلاح شده کمترین افزایش خطا (٪۱۳٫۸) را نشان داد که حاکی از تعمیم‌پذیری نسبی بهتر این ساختار است. این ویژگی می‌تواند ناشی از توزیع متوازن‌تر طبقات ریسک در این مدل باشد که شبکه عصبی را در برابر نویز و ناهمگنی داده‌های واقعی مقاوم‌تر می‌کند.

۴٫۵ محدودیت‌های پژوهش

این پژوهش دارای محدودیت‌های زیر است:

۱. وابستگی به برچسب‌گذاری: داده‌های سیستم گزارش‌دهی ایمنی هوانوردی بر اساس قضاوت کارشناسان برچسب‌گذاری شده‌اند که ممکن است حاوی سوگیری‌های انسانی باشد. هرچند سیستم گزارش‌دهی ایمنی هوانوردی از فرآیندهای استاندارد شده استفاده می‌کند، اما احتمال خطای انسانی وجود دارد.

۲. تعداد محدود متغیرهای ورودی: تنها ۵ متغیر ورودی در نظر گرفته شده، در حالی که در عملیات واقعی ده‌ها متغیر (شرایط جوی، خستگی خدمه، تراکم ترافیک و غیره) بر ریسک تأثیر می‌گذارد.

۳. تعریف ساده‌انگارانه از تاب‌آوری: مدل تاب‌آور مورد استفاده فقط به تقسیم ناحیه زرد اکتفا کرده و مولفه‌های اصلی تاب‌آوری (مانند

قابلیت بازیابی، افزودگی سیستم) را در نظر نگرفته است.

۴. تعمیم‌پذیری محدود: نتایج به دست آمده مختص صنعت هوانوردی تجاری و با پارامترهای سیستم گزارش‌دهی ایمنی هوانوردی است. تعمیم به سایر صنایع یا حتی سایر بخش‌های هوانوردی (مانند نظامی یا عمومی) نیازمند مطالعه جداگانه می‌باشد

۶. نتیجه گیری

در این پژوهش، سه مدل ماتریس ریسک شامل استاندارد ایکائو، فیلتر دوگانه، و تاب‌آور اصلاح‌شده از نظر قابلیت یادگیری توسط شبکه عصبی پرسپترون چندلایه و بازتولید سطوح ریسک، با استفاده از ۴۷،۷۲۳ گزارش واقعی پایگاه داده سیستم گزارش‌دهی ایمنی هوانوردی ناسا ارزیابی تطبیقی شدند. یافته‌های کمّی نشان داد که:

۱- مدل تاب‌آور اصلاح‌شده با بالاترین ضریب تعیین (R^2=0.9612) و کمترین میانگین خطای مطلق (MAE=1.108) از نظر آماری عملکرد برتری نسبت به دو مدل دیگر دارد. این مدل بدون تغییر در منطق ضرب احتمال و شدت، صرفاً با تقسیم ناحیه زرد ماتریس ایکائو به دو سطح میانی، حساسیت به ریسک‌های مرزی را افزایش داده و سیگنال گرادیان مناسب‌تری برای یادگیری شبکه عصبی فراهم می‌کند.

۲- مدل فیلتر دوگانه علیرغم دقت قابل قبول R^2=0945، به دلیل پیچیدگی محاسباتی حدود ۲,۵ برابر مدل استاندارد و حساسیت به تغییر در داده‌های آموزش، برای کاربردهای برخط و سازمان‌هایی با بلوغ فرایندی متوسط توصیه نمی‌شود.

۳- مدل استاندارد ایکائو پایین‌ترین عملکرد R^2=0882، MAE=1.592 را داشت که مؤید محدودیت این ساختار ساده در شناسایی و تفکیک سناریوهای پیچیده و مرزی است.

یکی از مهم‌ترین یافته‌های این پژوهش، شکاف معنادار بین عملکرد مدل‌ها بر روی داده‌های شبیه‌سازی و واقعی بود. میانگین افزایش خطای MAE در داده‌های واقعی نسبت به شبیه‌سازی حدود ۱۵,۷٪ بود که نشان می‌دهد تکیه صرف بر داده‌های شبیه‌سازی شده می‌تواند به ارزیابی خوش‌بینانه از عملکرد مدل‌ها منجر شود. در این میان، مدل تاب‌آور اصلاح‌شده کمترین افزایش خطا (۱۳,۸٪) را نشان داد که حاکی از تعمیم‌پذیری نسبی بهتر این ساختار است.

از منظر کاربردی، اگر هدف یک سازمان هواپیمایی سادگی و سرعت باشد، مدل استاندارد ایکائو کافی اما با خطای بیشتر در ریسک‌های مرزی قابل قبول است. اگر دقت در اولویت باشد و سازمان از بلوغ تحلیلی بالایی برخوردار است، مدل فیلتر دوگانه مناسب‌تر است. اما اگر حساسیت به ریسک‌های میانی بدون افزایش چشمگیر پیچیدگی مد نظر باشد، مدل تاب‌آور اصلاح‌شده به عنوان بهینه‌ترین گزینه پیشنهاد می‌شود.

مدل تاب‌آور اصلاح‌شده کمترین افزایش خطا (۱۵,۴٪) را در مواجهه با داده واقعی داشته است که نشان‌دهنده تعمیم‌پذیری نسبی بهتر این مدل است.

هیچ‌یک از نتایج این پژوهش نمی‌تواند مبنای تصمیم‌گیری عملیاتی در ایمنی پرواز بدون اعتبارسنجی میدانی گسترده‌تر باشد.

۱,۶ پیشنهادهای آتی این مقاله

پیاده‌سازی مدل پیشنهادی (تاب‌آور اصلاح‌شده) بر روی داده‌های واقعی یک شرکت هواپیمایی داخلی با حجم حداقل ۲۰۰۰ رکورد و مقایسه با خروجی سیستم مدیریت ایمنی موجود.

توسعه مدل ریسک پویا با استفاده از یادگیری تقویتی برای به‌روزرسانی بلادرنگ سطوح ریسک با ورود داده‌های جدید.

ترکیب ماتریس فیلتر دوگانه با منطق فازی برای کاهش اثر قضاوت‌های ذهنی در مرحله فیلترگذاری.

طراحی یک داشبورد مدیریت ریسک هوشمند مبتنی بر معماری لبه برای برج مراقبت پرواز.

جمع‌آوری داده واقعی حداقل ۲۰۰۰ رکورد برای دستیابی به توان آماری قابل قبول (Power ≥ 0.80) در مطالعات آتی.

۷.منابع

[1] ICAO, *Safety Management Manual (Doc 9859)*, 4th ed. Montreal: International Civil Aviation Organization, 2018.

[2] B. J. M. Ale et al., “Evaluation of risk matrix in aviation safety management,” *Journal of Risk Research*, vol. 22, no. 4, pp. 421–437, 2019.

[3] L. A. Cox, “What's wrong with risk matrices?,” *Risk Analysis*, vol. 28, no. 2, pp. 497–512, 2008.

[4] Y. Li and F. W. Guldenmund, “Dynamic multi-stage risk assessment frameworks in aviation operations,” *Safety Science*, vol. 118, pp. 291–303, 2019.

[5] E. Hollnagel, *Safety-I and Safety-II: The Past and Future of Safety Management*. CRC Press, 2014.

[6] N. J. Duijm, “Recommendations on the use and design of risk matrices,” *Safety Science*, vol. 76, pp. 21–31, 2015.

[7] H. J. Pasman and W. J. Rogers, “A dynamic risk assessment model for aviation safety,” *Process Safety and Environmental Protection*, vol. 143, pp. 112–123, 2020.

[8] Y. Chen et al., “Dynamic risk assessment in aviation maintenance: From traditional matrices to AI-driven frameworks,” *Reliability Engineering & System Safety*, vol. 245, pp. 109–125, 2024.

[9] A. Kumar and R. Singh, “Neural network-based risk prediction for aircraft ground operations,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 3, pp. 2145–2158, 2025.

[10] L. Wang et al., “Transformer-based analysis of unsafe event narratives in aviation safety,” *Aerospace*, vol. 11, no. 2, pp. 88–104, 2024.

[11] NASA ASRS, *Aviation Safety Reporting System Database*. NASA Ames Research Center, 2024.

[12] EASA, *Annual Safety Review 2025*. European Union Aviation Safety Agency, 2025.

[13] L. A. Cox, “What's wrong with risk matrices?,” *Risk Analysis*, vol. 28, no. 2, pp. 497–512, 2008.

[14] N. J. Duijm, “Recommendations on the use and design of risk matrices,” *Safety Science*, vol. 76, pp. 21–31, 2015.

[15] N. Borjalilu and M. Rezaei, “A dynamic safety risk assessment model,” *Aviation*, vol. 22, no. 4, pp. 143–155, 2018.

[16] ICAO, *Safety Risk Management Methodologies (SRM) – Overview*. Montreal: International Civil Aviation Organization, 2026.

[17] EASA, *Annual Safety Review 2025*. European Union Aviation Safety Agency, 2025.

[18] Y. Chen et al., “Dynamic risk assessment in aviation maintenance: From traditional matrices to AI-driven frameworks,” *Reliability Engineering & System Safety*, vol. 245, pp. 109–125, 2024.

[19] A. Kumar and R. Singh, “Neural network-based risk prediction for aircraft ground operations,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 3, pp. 2145–2158, 2025.

[20] L. Wang et al., “Transformer-based analysis of unsafe event narratives in aviation safety,” *Aerospace*, vol. 11, no. 2, pp. 88–104, 2024.

[21] H. J. Pasman and W. J. Rogers, “A dynamic risk assessment model for aviation safety,” *Process Safety and Environmental Protection*, vol. 143, pp. 112–123, 2020.

[22] B. J. M. Ale et al., “Evaluation of risk matrix in aviation safety management,” *Journal of Risk Research*, vol. 22, no. 4, pp. 421–437, 2019.