پیش‌بینی وقوع امواج سهمگین مبتنی بر متغیرهای هواشناسی با استفاده از روش ترکیبی الگوریتم ضریب داده پرت محلی و شبکه عصبی دسته‌بندی کننده

کیومرث محمودی[۱]، حسن قاسمی[۲*]، ابوالحسن رزمی‌نیا[۳]

[۱] دانشجوی دکتری، دانشکده مهندسی دریا، دانشگاه صنعتی امیرکبیر؛ kumarsmahmoodi@aut.ac.ir
[۲] استاد، دانشکده مهندسی دریا، دانشگاه صنعتی امیرکبیر؛ gasemi@aut.ac.ir
[۳] دانشیار، گروه مهندسی برق، دانشگاه خلیج فارس؛ razminia@pgu.ac.ir

اطلاعات مقاله

تاریخچه مقاله:
تاریخ دریافت مقاله: ۱۳۹۷/۰۹/۱۹
تاریخ پذیرش مقاله: ۱۳۹۸/۰۷/۰۷

کلمات کلیدی:
پیش‌بینی امواج سهمگین
داده‌های هواشناسی
شبکه عصبی مصنوعی دسته‌بندی کننده
روش ضریب داده پرت محلی.

چکیده

به دلیل ماهیت پیچیده تصادفی و غیر خطی امواج، پیش‌بینی وقوع امواج سهمگین دشوار است. پیش‌بینی وقوع امواج سهمگین می‌تواند تا حد زیادی از بروز خسارات جانی و مالی بکاهد. هدف از این تحقیق ارائه یک روش ترکیبی برای پیش‌بینی وقوع امواج سهمگین از روی متغیرهای هواشناسی با استفاده از روش‌های داده‌کاوی ضریب داده پرت محلی و شبکه عصبی مصنوعی دسته‌بندی کننده است. برای بررسی کارایی مدل ارائه شده از داده‌های تاریخی متغیرهای هواشناسی برداشت شده طی سه سال از ایستگاه‌های شماره ۴۱۰۴۱ و ۴۱۰۰۴ دو طوفان مشهور Dean (2007) و Irene (2011) استفاده شده است. ابتدا با استفاده از روش ضریب داده پرت محلی امواج سهمگین شناسایی شده، سپس از خروجی‌های این روش برای آموزش شبکه عصبی استفاده شده است. از ۸۰ درصد داده‌های جمع آوری شده برای مرحله آموزش و از مابقی برای مرحله آزمون مدل بکار رفته است. کارایی روش ارائه شده با استفاده از معیارهای متداول بررسی عملکرد روی داده‌های آموزشی و آزمون بررسی شد؛ نتایج بیانگر عملکرد مناسب روش پیشنهادی در پیش‌بینی وقوع امواج سهمگین از روی متغیرهای هواشناسی است به طوری که برای هر دو مجموعه داده مورد بررسی دقت میانگین به ۹۹٪ رسید.

Proposed a New Hybrid LOF-ANN Method to Extreme Wave Height Prediction based on Meteorological Data

Kumars Mahmoodi[1], Hassan Ghassemi [2*], Abolhassan Razminia[3]

[1] Ph. D. Student, Faculty of Marine Technology, Amirkabir University of Technology; kumarsmahmoodi@aut.ac.ir
[2] Professors, Faculty of Marine Technology, Amirkabir University of Technology; gasemi@aut.ac.ir
[3] Associate Professor, Electrical Engineering Department, School of Engineering, Persian Gulf University

ARTICLE INFO

Article History:
Received: 10 Dec. 2018
Accepted: 29 Sep. 2019

Keywords:
Extreme Wave Height Prediction
Meteorological Data
ANN Classifier
Local Outlier Factor

ABSTRACT

Extreme wave height prediction is very challenging due to its very high non-stationarity and non-linearity nature. The main aim of the present study is to propose a new hybrid method based on Local Outlier Factor and Artificial Neural Networks classifier, called LOF-ANN, to accurate prediction of extreme wave height occurrence using historical meteorological data. In this study to create models two major hurricanes Dean 2007 and Irene 2011at two locations (NDBC wave buoys stations: http://www.ndbc.noaa.gov) namely; 41004, 41041 in the Gulf of Mexico, is used. TO detect extreme waves, LOF method is used. The outputs of this method are considered as ANN targets. Extreme and normal waves are considered as Class 0 and class 1, respectively. The inputs of ANN models are historical metrological data, including: Wind direction (WDIR), Wind speed (WSPD), Sea level pressure (PRES), Air temperature (ATMP), and Sea surface temperature (WTMP). To create and evaluation of models, the input data sets are randomly divided into training (80%) and test set (20%). The performance of created models is evaluated using three popular criteria Root Mean Square Error (RMSE) and Receiver Operating Characteristic (ROC) and accuracy parameter. The experiment results show that the proposed method is able to predict the occurrence of extreme wave heights with height accuracy (up to 99%).

[Downloaded from marine-eng.ir on 2026-02-07]
[DOR: 20.1001.1.17357608.1398.15.30.2.2]
[DOI: 10.29252/marineeng.15.30.23]

۱- مقدمه

فرایند فیزیکی تولید امواج در دریا بسیار پیچیده، غیر ایستا، غیرخطی و تصادفی است. با وجود پیشرفت‌های قابل توجه در روش‌های محاسباتی، راه‌حل‌های ارائه شده کاملا دقیق نبوده و در بسیاری از موارد قابل تعمیم به نواحی دیگر با شرایط موج متفاوت نیست. به همین دلیل فرآیند تولید موج و پیش‌بینی آن به خصوص در شرایط حاد و امواج طوفانی تا کنون آن گونه که باید شناسایی نشده و ناشناخته باقی مانده است. پیش‌بینی شرایط ایجاد امواج به خصوص امواج سهمگین از چندین ساعت تا چندین روز قبل از وقوع آن بسیار حائز اهمیت بوده و می‌تواند از ایجاد خسارت‌های جانی و مالی جلوگیری کند. به همین دلیل این مسئله در دهه‌های اخیر به یکی از مهمترین مسائل تحقیقاتی در مهندسی سواحل و دریا بدل شده است.

روش‌های سنتی پیش‌گویی امواج، مبتنی بر یک سری از ساده‌سازی‌ها بوده و با استفاده از روش‌های پارامتری و یا عددی قابل انجام هستند. در روش‌های پارامتری با مشخص بودن مشخصه‌هایی نظیر سرعت باد در یک ارتفاع معین از سطح دریا، طول حوزه بادخیز، مدت زمان وزش باد، و یک سری از فرضیات دیگر ارتفاع و پریود امواج به صورت صریح قابل محاسبه است. در روش‌های عددی با حل معادلات موج و اعمال شرایط مرزی مشخصات موج محاسبه می‌شود. این روش‌ها نیازمند مشخص بودن یک سری از پارامترهای مربوط به اقیانوس‌شناسی و هواشناسی هستند. با این حال در بسیاری از مواقع دقت روش‌های عددی بیشتر است [۱]. شماری از محققان از مدل‌های طیفی برای پیش‌بینی امواج و وقوع رویدادهای حاد به خصوص شرایط بحرانی دریا استفاده کرده‌اند [۲، ۳]. اخیرا [۴، ۵] مدل‌های نسل سوم موج را برای شناسایی شرایط بحرانی اقیانوس اطلس شمالی بکار برده‌اند. محققان از روش‌های آمار کلاسیک مثل مدل خود همبسته میانگین متحرک[۱]، مدل میانگین متحرک جمع بسته خود همبسته[۲] و فیلتر کالمن[۳] نیز برای پیش‌بینی موج و تحلیل آن استفاده کرده‌اند [۶]. این روش‌ها دارای محدودیت‌هایی در مدل‌سازی خواص دینامیک و غیر خطی سری‌های زمانی موج هستند. استفاده از روش‌های محاسبات نرم که مبتنی بر عدم قطعیت هستند، جایگزین مناسبی برای روش‌های آماری کلاسیک است. روش‌های محاسبات نرم برخواسته از پیشرفت‌های علوم نوین کامپیوتر بوده که به دلیل استوار بودن بر عدم قطعیت، سادگی کاربرد و قابلیت در نظر گرفتن خواص دینامیک و غیر خطی موج در میان محققان علوم دریایی به جایگاه و محبوبیت ویژه‌ای دست یافته است. محاسبات نرم را می‌شود حاصل تلاش‌های جدید علمی دانست که مدل‌سازی، تحلیل، و در نهایت کنترل سیستم‌های پیچیده را با سهولت و موفقیت بیشتری امکان‌پذیر می‌سازد. به عنوان مهم‌ترین شاخه‌های علوم نرم، باید داده‌کاوی، یادگیری ماشین، هوش محاسباتی، منطق فازی، شبکه‌های عصبی مصنوعی، و الگوریتم‌های تکاملی را نام برد. طی دو دهه گذشته محققان زیادی از شبکه‌های عصبی برای تحلیل و پیش‌بینی امواج بهره گرفته‌اند [۷-۱۰]. در [۱۱] مروی جامع بر کاربردهای شبکه عصبی در مهندسی دریا ارائه شده است. شماری از محققان روش ترکیبی شبکه‌های عصبی با با یک یا چند روش دیگر مثل الگوریتم‌های تکاملی، سیستم‌های فازی و غیره را برای مدل‌سازی مشخصه‌های موج بکار برده‌اند [۱۲-۱۵]. تمامی این تحقیقات بیانگر توانایی مناسب روش‌های محاسبات نرم در مدل‌سازی شرایط پیچیده امواج هستند.

همانگونه که به طور مختصر در بالا بیان شد، تا کنون توجه بیشتر تحقیقات انجام شده متکی بر پیش‌بینی و مدل‌سازی شرایط امواج در حالت عادی و شرایط نرمال بوده است. این موضوع بدین علت است که پیش‌بینی و مدل‌سازی امواج در شرایط نرمال بسیار ساده‌تر از شرایط بحرانی مثل امواج طوفانی است. شناخت ماهیت امواج طوفانی به دلیل افزایش خاصیت تصادفی و غیر خطی این امواج افزایش دشوارتر از شرایط نرمال است و در تحقیقات بسیار کمتر به آن توجه شده است. وقوع امواج سهمگین و طوفانی می‌تواند برای سازه‌های ساحلی و دور از ساحل و شناورها بسیار خطرناک باشد. از اینرو ارائه راهکاری برای پیش‌بینی وقوع آن‌ها ضروری است. با بررسی داده‌های تاریخی هواشناسی و ارتفاع امواج یک منطقه خاص می‌توان وقوع امواج سهمگین را تا حد قابل قبولی پیش‌بینی کرد. هدف از این تحقیق ارائه یک روش ترکیبی برای پیش‌بینی وقوع امواج سهمگین از روی روی متغیرهای هواشناسی با استفاده از روش‌های داده‌کاوی ضریب داده پرت محلی[۴] (LOF) [۱۶، ۱۷، ۱۸] و شبکه عصبی مصنوعی دسته‌بندی کننده[۵] (ANN) [۱۹] است. روش LOF یکی از معروف‌ترین الگوریتم‌های شناسایی داده نامتعارف در یادگیری ماشین است که از یادگیری نظارت نشده برای شناسایی الگوهای غیر متعارف در داده‌ها استفاده می‌کند. در این تحقیق از این روش برای شناسایی امواج سهمگین استفاده شده است. موج سهمگین موجی در نظر گرفته شده که به طور قابل توجهی از الگوی نرمال جامعه آماری طبعیت نکرده به گونه‌ای که اینگونه برداشت شود که با یک روند متفاوت از سایر داده‌ها تولید شده است. بنابراین با توجه به این تعریف امواجی که در فاصله قابل توجه دورتری از سایر امواج قرار گرفته باشند و تراکم همسایگی آن‌ها ناچیز باشد، به عنوان کاندیدای موج سهمگین در نظر گرفته شده‌اند. ورودی‌های الگوریتم LOF داده‌های ارتفاع موج مشخصه از نواحی مورد بررسی است. با استفاده از این روش داده‌های ارتفاع امواج در دو کلاس سهمگین (کلاس اول) و کلاس عادی (کلاس دوم) دسته‌بندی می‌شوند. از خروجی‌های روش LOF به عنوان داده‌های هدف در شبکه عصبی دسته‌بندی کننده

[Downloaded from marine-eng.ir on 2026-02-07]
[DOR: 20.1001.1.17357608.1398.15.30.2.2]
[DOI: 10.29252/marineeng.15.30.23]

استفاده شده است. هدف از دسته‌بندی قرار دادن نمونه‌های یک مجموعه داده در گروه‌های (کلاس‌های) از پیش تعیین شده (هدف) است. در مسائل دسته‌بندی بر خلاف مسائل پیش‌بینی، هدف پیش‌بینی مقدار یک متغیر گسسته است. از شبکه‌های عصبی هم می‌توان برای پیش‌بینی مسائل پیوسته و هم مسائل گسسته استفاده کرد. مدل‌سازی با استفاده از شبکه عصبی نیازمند آموزش شبکه است. آموزش شبکه عصبی نیازمند تعیین پارامترهای شبکه و مجموعه داده آموزشی است که خود به داده‌های ورودی و هدف تقسیم‌بندی می‌شود. پارامترهای شبکه معمولا با آزمون و خطا انتخاب می‌شوند. در تمامی مدل‌های ایجاد شده داده‌های هواشناسی (جهت باد، سرعت باد، ارتفاع موج مشخصه، فشار سطح دریا، دمای هوا و دمای سطح دریا) به عنوان ورودی‌های شبکه و کلاس‌های امواج (سهمگین و عادی) به عنوان خروجی شبکه در نظر گرفته شده‌اند. پس از آموزش شبکه به طور بهینه، می‌توان از آن برای پیش‌بینی وقوع امواج سهمگین استفاده کرد، بدین صورت که داده‌های هواشناسی جدید به عنوان ورودی به شبکه داده شده و شبکه تعیین می‌کند که این داده‌ها می‌تواند مربوط به وقوع یک موج سهمگین و یا یک موج عادی باشد. در این تحقیق برای مدل-سازی از داده‌های تاریخی هواشناسی برداشت شده طی سه سال از ایستگاه‌های شماره ۴۱۰۴۱ و ۴۱۰۰۴ دو طوفان مشهور Dean (2007) و Irene (2011) استفاده شده است.

رئوس مطالب این مقاله بدین صورت است: در بخش بعد جزئیات منطقه مورد مطالعه و داده‌ها تشریح می‌شود. سپس در بخش ۳ جزئیات روش‌های داده‌کاوی ضریب داده پرت محلی و شبکه عصبی مصنوعی دسته‌بندی کننده ارائه شده است. نتایج مدل‌های ایجاد شده و بحث در خصوص آن‌ها در بخش ۴ بیان شده و در نهایت تحقیق با ارائه نتایج در بخش نتیجه‌گیری به پایان رسیده است.

۲- منطقه مورد مطالعه و داده‌ها

همانگونه که در بخش مقدمه بیان شد، تمرکز تحقیق حاضر بر پیش‌بینی امواج طوفانی و سهمگین (ارتفاع موج مشخصه) است. بدین منظور دو طوفان مشهور Dean (2007) و Irene (2011) در خلیج مکزیک مورد توجه قرار گرفته است. بیشترین ارتفاع موج مشخصه ثبت شده در تمامی آب‌های آمریکا در این طوفان‌ها ثبت شده است، به همین دلیل در این تحقیق این دو طوفان مورد توجه قرار گرفته است. مسیر گذر این طوفان‌ها در سایت NOAA [۲۰] ترسیم شده است که نشان‌دهنده این است که این مسیر به ایستگاه‌های اندازه‌گیری شماره ۴۱۰۰۴ و ۴۱۰۴۱ در خلیج مکزیک بسیار نزدیک است. به همین دلیل در تحقیق حاضر داده‌های برداشت شده از این دو ایستگاه استفاده شده است. جزئیات طوفان-ها و مکان بویه‌ها در جدول ۱ و شکل ۱ ارائه شده است. از داده‌های تاریخی هواشناسی (جهت باد (WDIR)، سرعت باد (WSPD)، ارتفاع موج مشخصه (WVHT)، فشار سطح دریا (PRES)، دمای هوا (ATMP) و دمای سطح دریا (WTMP)) ثبت شده ساعتی در طول سه سال برای آموزش و آزمون مدل‌ها استفاده شده است. در جدول ۲ مشخصه‌های آماری متغیرهای هواشناسی و ارتفاع امواج مورد بررسی برای داده‌های آموزشی و داده‌های آزمون ارائه شده است.

جدول ۱- مشخصات بویه‌ها و طوفان‌های مورد بررسی [۲۰]

ردیف	طوفان	بازه وقوع	ایستگاه	مکان ایستگاه	بازه اندازه‌گیری
۱	Dean	۱۳ تا ۲۳ آگوست ۲۰۰۷	۴۱۰۴۱	N ۴۶/۰۸۲ W (۱۴°۱۹'۴۳" N ۴۶°۴'۵۵" W) ۱۴/۳۲۹	۲۰۰۵ تا ۲۰۰۷
۲	Irene	۲۰ تا ۲۸ آگوست ۲۰۱۱	۴۱۰۰۴	N ۷۹/۰۹۹ W (۳۲°۳۰'۲" N ۷۹°۵'۵۸" W) ۳۲/۵۰۱	۲۰۰۹ تا ۲۰۱۱

جدول ۲- مشخصه‌های آماری متغیرهای هواشناسی مورد بررسی

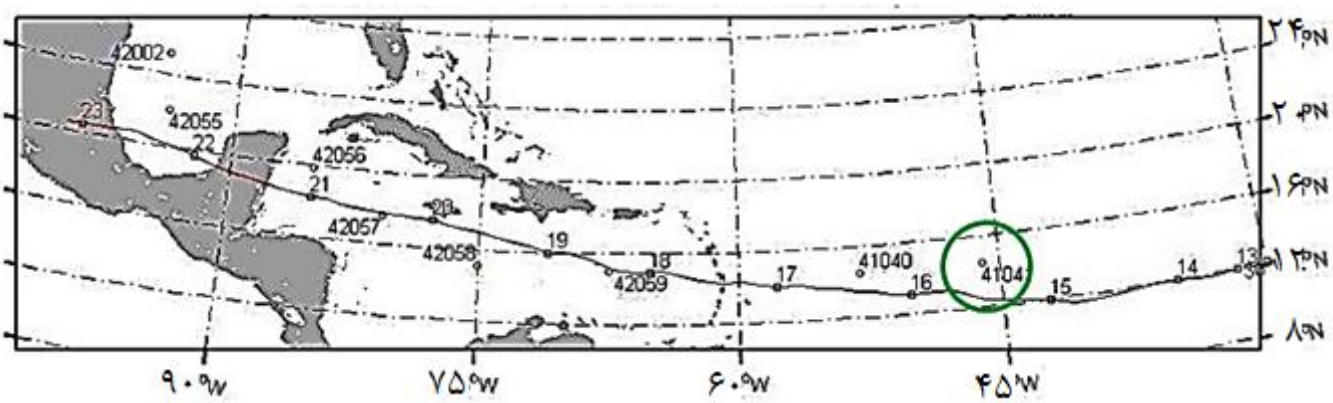
(الف): طوفان Dean (2007)

متغیر	مجموعه داده آموزشی (۴۸۳۵ نمونه)				مجموعه داده آزمایشی (۳۸۰۹ نمونه)			
	کمترین	بیشترین	میانگین	انحراف معیار	کمترین	بیشترین	میانگین	انحراف معیار
جهت باد (درجه)	۰/۰	۳۵۹	۸۶/۶۴	۳۳/۰۵	۰/۰	۳۵۵	۸۶/۴۷	۳۳/۳۴
سرعت باد (متر بر ثانیه)	۰/۱	۱۹/۳	۷/۰۲	۱/۹۵	۰/۰	۱۸/۹	۶/۹۷	۱/۹۹
ارتفاع موج مشخصه (متر)	۰/۹۲	۶/۷	۲/۰۰	۰/۴۹	۰/۹۶	۶/۸۷	۱/۹۹	۰/۴۹
فشار سطح دریا (هکتو پاسکال)	۱۰۰۵/۷	۱۰۲۰/۵	۱۰۱۴/۳۳	۱/۹۶	۱۰۰۵/۸	۱۰۲۰/۸	۱۰۱۴/۲۷	۱/۹۴
دمای هوا (سلسیوس)	۲۲/۲	۲۹/۳	۲۶/۴۰	۱/۰۲	۲۱/۷	۲۹/۵	۲۶/۴۳	۱/۰۰
دمای سطح دریا (سلسیوس)	۲۵	۲۹/۸	۲۷/۱۰	۰/۹۲	۲۵	۳۰	۲۷/۱۳	۰/۸۹

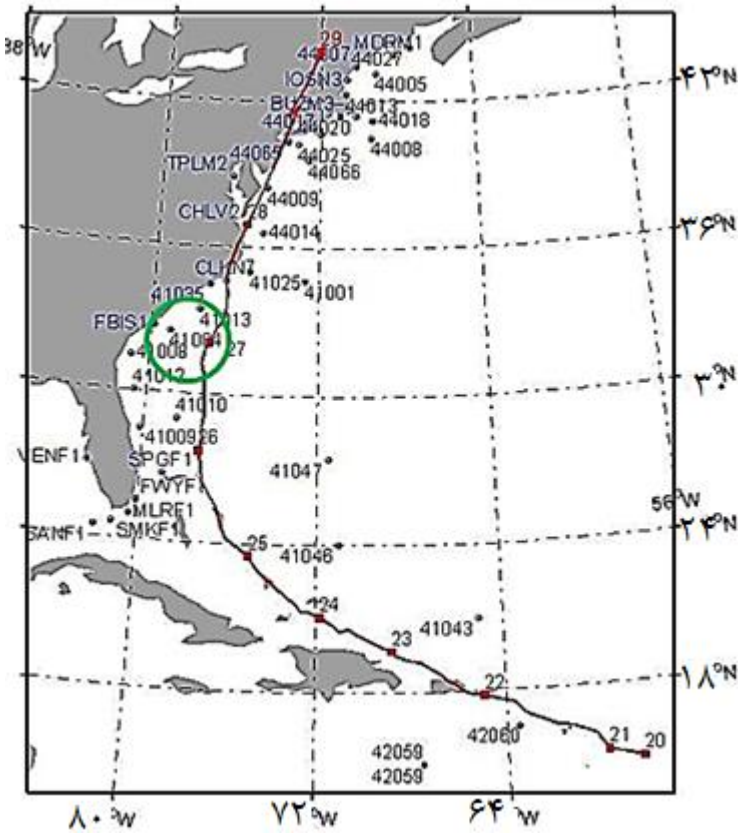
[Downloaded from marine-eng.ir on 2026-02-07]
[DOR: 20.1001.1.17357608.1398.15.30.2.2]
[DOI: 10.29252/marineeng.15.30.23]

(ب): طوفان Irene (2011)

متغیر	مجموعه داده آموزشی (۳ نمونه)				مجموعه داده آزمایشی (۳ نمونه)			
	کمترین	بیشترین	میانگین	انحراف معیار	کمترین	بیشترین	میانگین	انحراف معیار
جهت باد (درجه)	۱	۳۶۰	۱۸۰/۷۱	۹۷/۷۷	۱	۳۶۰	۱۸۱/۱۸	۹۷/۸۷
سرعت باد (متر بر ثانیه)	۰	۲۱/۱	۶/۴۵	۳/۱۲	۰	۲۱/۸	۶/۴۵	۳/۱۲
ارتفاع موج مشخصه (متر)	۰/۲۹	۷/۴۸	۱/۲۵	۰/۶۶	۰/۲۸	۶/۲۹	۱/۲۵	۰/۶۵
فشار سطح دریا (هکتو پاسکال)	۹۹۴/۹	۱۰۳۵/۱	۱۰۱۷/۱۹	۵/۸۲	۹۹۲/۷	۱۰۳۴/۷	۱۰۱۷/۱۹	۵/۸۲
دمای هوا (سلسیوس)	۰/۵	۳۰/۹	۲۰/۷۰	۶/۵۵	۰/۵	۳۰/۷	۲۰/۷۰	۶/۵۵
دمای سطح دریا (سلسیوس)	۱۴/۳	۳۱/۵	۲۳/۰۳	۴/۳۹	۱۴/۳	۳۲/۲	۲۳/۰۳	۴/۳۹



(الف): مسیر عبور طوفان Dean (2007)



(ب): مسیر عبور طوفان Irene (2011)

شکل ۱- مکان بویه‌ها (مشخص شده با علامت دایره) و مسیر عبور طوفان‌های مورد بررسی [۲۰]

[Downloaded from marine-eng.ir on 2026-02-07]
[DOR: 20.1001.1.17357608.1398.15.30.2.2]
[DOI: 10.29252/marineeng.15.30.23]

۳- روش‌ها

۳-۱- روش ضریب داده پرت محلی

شناسایی داده‌های نامتعارف (پرت، ناهنجار یا غیر نرمال) در یک پایگاه داده شاخه‌ای پرکاربرد از علم داده‌کاوی است. محققان تا کنون روش‌های متعددی را برای شناسایی این دسته از داده‌ها ارائه کرده‌اند. در یک دسته‌بندی این روش‌ها به روش‌های آماری، مبتنی بر فاصله، مبتنی بر خوشه‌بندی و مبتنی بر چگالی تقسیم‌بندی می‌شوند. روش‌های آماری مبتنی بر برآورد پارامترهای آماری مثل میانگین، میانه، و توزیع داده‌ها هستند. این روش‌ها معمولا با مجموعه داده‌های تک بعدی کاربرد دارند و با افزایش ابعاد و اندازه مجموعه داده کارایی خود را از دست می‌دهند. روش‌های مبتنی بر فاصله با اندازه‌گیری فاصله بین داده‌ها نرمال و غیر نرمال بودن آن-ها را تعیین می‌کنند. روش‌های مبتنی بر خوشه‌بندی داده‌ها را در تعدادی خوشه قرار می‌دهند. خوشه‌ها به نحوی ایجاد می‌شوند که داده‌های واقع شده در یک خوشه بیشترین شباهت را به یکدیگر، و کمترین شباهت را با خوشه‌های دیگر داشته باشند. داده‌هایی که به هیچ یک از خوشه‌ها تعلق نداشته و یا عضوی از خوشه‌های با تعداد عناصر اندک (نسبت به سایر خوشه‌ها) باشند به عنوان داده‌های غیر نرمال در نظر گرفته می‌شوند. روش‌های مبتنی بر تخمین چگالی با اندازه‌گیری چگالی همسایگی هر داده نرمال و یا غیر نرمال بودن آن‌ها را مشخص می‌کنند. در این روش‌ها هر چه تمرکز همسایگی یک داده بیشتر باشد و آن داده در یک فضای چگال‌تر (تعداد داده-های واقع شده در همسایگی آن بیشتر باشد) قرار گرفته باشد، احتمال نرمال بودن آن افزایش می‌یابد. عکس این مطلب نیز برقرار است. روش‌های مبتنی بر چگالی معمولا قابل کاربرد با مجموعه داده‌های چند بعدی بوده و افزایش اندازه مجموعه داده تاثیر آنچنانی بر عملکرد آن‌ها ندارد.

روش ضریب داده پرت محلی [۱۶، ۱۷، ۱۸] یکی از روش‌های معروف تشخیص داده نامتعارف مبتنی بر تخمین چگالی محلی در یادگیری ماشین است. این روش قادر به شناسایی داده‌های نامتعارف در مجموعه داده‌های چند بعدی است. روش ضریب داده پرت محلی به هر عضوی از مجموعه داده درجه‌ای با توجه به چگالی همسایه‌های محلی آن نسبت می‌دهد. این درجه ضریب پرت بودن داده (LOF) نامیده می‌شود. این ضریب به میزان ایزوله بودن یک داده وقتی که با همسایه‌های محلی آن مقایسه می‌شود، بستگی دارد. در حالت کلی اگر ضریب LOF محاسبه شده برای یک داده زیاد باشد (معمولا بیشتر از ۱)، آن داده می‌تواند کاندیدای یک داده غیر نرمال باشد.

قبل از تشریح الگوریتم ضریب داده پرت محلی، پارامترهای مورد استفاده معرفی می‌شوند. D مجموعه داده، o، p و q چند داده نمونه که در D واقع شده‌اند و k یک عدد صحیح مثبت است که توسط کاربر انتخاب می‌شود. $d(p,q)$ تابع اندازه‌گیری فاصله بین داده‌های p و q است که می‌تواند هر تابع اندازه‌گیری فاصله باشد. در این تحقیق از تابع اقلیدسی استفاده شده است. مراحل محاسبه ضریب LOF داده p به شرح زیر است:

گام اول: محاسبه فاصله k امین همسایه p

برای هر داده p، $k_{dist}(p)$ فاصله k امین همسایه نزدیک p از آن است. برای محاسبه این پارامتر می‌توان ابتدا فاصله k تا همسایه نزدیک p را پیدا کرد، سپس k امین فاصله را به عنوان $k_{dist}(p)$ انتخاب کرد. k تا همسایه‌های یک داده، داده‌هایی هستند که در فاصله نزدیک‌تری نسبت به آن داده واقع شده باشند. در واقع داده‌هایی که در k تا فاصله همسایگی p واقع می‌شوند دارای دو ویژگی هستند:

الف) برای حداقل k تا داده $o' \in D \setminus \{p\}$، $d(p,o') \leq d(p,o)$.

ب) برای حداکثر $k-1$ داده $o' \in D\setminus\{p\}$، $d(p,o') < d(p,o)$.

بنابراین $k_{dist}(p)$ تخمینی از چگالی همسایگی اطراف p را مشخص می‌کند.

گام دوم: پیدا کردن همسایگی k امین فاصله p [۶]

هر داده‌ای که فاصله آن از p، از $k_{dist}(p)$ کمتر و یا مساوی باشد، در k امین فاصله همسایگی p قرار می‌گیرد. این تعریف به صورت زیر ارائه می‌شود:

$$N_k(p) = \{q \in D \setminus \{p\} \mid d(p,q) \leq k_{dist}(p)\} \qquad (۱)$$

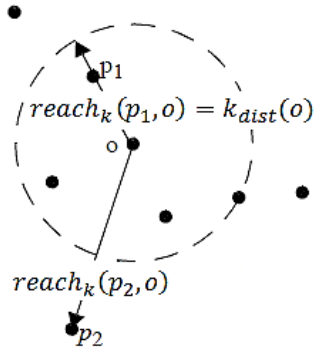
گام سوم: محاسبه چگالی دسترسی داده p نسبت به داده o [۷]

برای هر داده o واقع شده در k تا فاصله همسایگی p، چگالی دسترسی p نسبت به داده o به صورت زیر تعریف می‌شود:

$$reach_k(p,o) = max\{k_{dist}(o), d(p,o)\} \qquad (۲)$$

شکل ۲ نشان دهنده مثالی از فاصله دسترسی به ازای $k = 4$ است. همان طور که در این شکل نشان داده شده است، اگر داده p خارج از $k_{dist}(o)$ واقع شده باشد (p_2 در شکل)، چگالی دسترسی برابر فاصله بین آنها است، یعنی $d(o,p_2)$. حال اگر فاصله آنها از $k_{dist}(o)$ کمتر باشد (p_1 در شکل)، چگالی دسترسی برابر $k_{dist}(o)$ است.

[Downloaded from marine-eng.ir on 2026-02-07]
[DOR: 20.1001.1.17357608.1398.15.30.2.2]
[DOI: 10.29252/marineeng.15.30.23]



شکل ۲- مثالی از چگالی دسترسی به ازای $k = 4$

گام چهارم: محاسبه چگالی دسترسی محلی p [۸]

چگالی دسترسی محلی p، معکوس میانگین چگالی دسترسی k تا همسایه‌های نزدیک p است:

$$lrd_k(p) = \left[\frac{|N_k(p)|}{\sum_{o\epsilon N_{k(p)}} reach_k(p,o)}\right] \quad (۳)$$

برای محاسبه چگالی داده p از چگالی دسترسی محلی آن استفاده می‌شود. با استفاده از چگالی دسترسی محلی می‌توان ضریب LOF را محاسبه کرد.

گام پنجم: محاسبه ضریب داده پرت محلی p

از ضریب LOF برای تشخیص نرمال و یا عدم نرمال بودن یک داده استفاده می‌شود. $LOF(p)$ میانگین نسبت‌های چگالی دسترسی محلی p و k تا همسایه‌های آن است:

$$LOF_k(p) = \frac{\sum_{o\epsilon N_k(p)} \frac{lrd_k(o)}{lrd_k(p)}}{|N_k(p)|} \quad (۴)$$

هرچه ضریب LOF محاسبه شده برای یک داده عددی بزرگتر از ۱ شود، احتمال غیر نرمال بودن آن داده افزایش می‌یابد.

۳-۲- شبکه‌های عصبی مصنوعی دسته‌بندی کننده

شبکه‌های عصبی مصنوعی از پرکاربردترین و عملی‌ترین روش‌های مدل‌سازی مسائل پیچیده و غیر خطی هستند که می‌توان از آن‌ها برای مسائل دسته‌بندی (که خروجی یک دسته است) یا مسائل رگرسیون (که خروجی یک مقدار عددی است) مورد استفاده قرار گیرند. اساس عملکرد الگوریتم‌های مبتنی بر شبکه عصبی، بهره‌گیری از سیستم عصبی مغز انسان است. این سیستم‌ها از شمار زیادی عناصر بهم پیوسته که وظیفه پردازش اطلاعات را بر عهده دارند، تشکیل شده‌اند که به این عناصر نورون گفته می‌شود. پردازش اطلاعات توسط نرون‌ها بوسیله تابع فعال‌سازی انجام می‌شود. تابع فعال‌سازی بر اساس نوع مسأله‌ای که قرار است بوسیله شبکه عصبی حل شود، از سوی طراح شبکه انتخاب می‌شود. مجموعه‌ای از نرون‌ها یک لایه را شکل می‌دهند. به طور کلی تقریباً هر شبکه عصبی از سه لایه تشکیل شده است [۲۱]:

- لایه ورودی: داده‌های ورودی مسئله در این لایه قرار می‌گیرند. هر گره در این لایه معادل یکی از متغیرهای ورودی است.
- لایه میانی: تمامی گره‌های لایه ورودی به گره‌های لایه میانی وصل می‌شوند. و تمامی گره‌های لایه میانی می‌توانند به یک و یا چند لایه میانی دیگر و یا به گره‌های لایه خروجی وصل شوند. به لایه میانی لایه نهان نیز گفته می‌شود.
- لایه خروجی: داده‌های خروجی مسئله در این لایه قرار می‌گیرند. هر گره در این لایه معادل یکی از متغیرهای خروجی است.

از پرکاربردترین معماری شبکه‌های عصبی، شبکه‌های چند لایه پیش‌خور[۹] است. شبکه‌های پیش‌خور به شبکه‌هایی گفته می‌شود که در آنها هیچ برگشتی از یک نرون وجود ندارد و هر نورون پس از دریافت بردار ورودی، یک خروجی مشخص ارائه می‌کند. به عبارت دیگر، در شبکه‌های پیش‌خور مسیر پاسخ همواره رو به جلو پردازش شده و به نرون‌های لایه یا لایه‌های قبل برنمی‌گردد. در این نوع شبکه، تعداد نرون‌های لایه ورودی برابر تعداد متغیرهای ورودی، و تعداد نرون‌های لایه خروجی برابر تعداد متغیرهای خروجی مسئله است. تعداد گره‌های لایه‌های نهان و همچنین تعداد لایه‌های نهان با توجه به ماهیت مسئله و از طریق آزمون و خطا بدست می‌آید. در حقیقت تعداد مناسب گره‌ها و لایه‌های نهان وقتی بدست می‌آید که شبکه بهترین جواب را ارائه کند. هنگام طراحی شبکه باید پارامترهای ساختار شبکه، نوع الگوریتم آموزش، نرخ یادگیری، تعداد لایه‌های شبکه، تعداد نرون‌ها در هر لایه و تعداد تکرارها برای هر الگو در خلال آموزش را مد نظر قرار داد [۲۱]. نوع آموزش در شبکه‌های پیش‌خور از نوع یادگیری نظارت شده است. در این نوع آموزش، داده‌ها از پیش برچسب‌گذاری شده و خروجی متناسب با ورودی‌ها از قبل تعیین می‌شود. سپس شبکه تحت داده‌های علامت‌گذاری شده آموزش داده شده تا میزان خطا به حداقل ممکن کاهش یابد. برای مسائل دسته‌بندی می‌توان از شبکه‌های عصبی پیش‌خور با یک لایه نهان، تابع فعال‌سازی سیگموئید در لایه نهان و تابع فعال‌سازی سیگموئید در لایه خروجی استفاده کرد. بنابراین در این تحقیق از این ساختار شبکه برای پیش‌بینی وقوع امواج سهمگین استفاده شده است.

شبکه‌های عصبی، سیستم‌های پویای هوشمند آزاد از مدل هستند که اساس عملکرد آنها استفاده از داده‌های تجربی گردآوری شده از سابقه سیستم مورد مطالعه است. این سیستم‌ها با پردازش داده‌های تجربی، دانش و یا قانون نهفته در ورای داده‌ها را کشف کرده و با استفاده از تکنیک‌ها و الگوریتم‌های هوش مصنوعی، روابط بین متغیرها را هرچقدر هم که پیچیده باشند فرا گرفته و از آن برای

[Downloaded from marine-eng.ir on 2026-02-07]
[DOR: 20.1001.1.17357608.1398.15.30.2.2]
[DOI: 10.29252/marineeng.15.30.23]

پیش‌بینی مقادیر آتی متغیرها استفاده می‌کنند. شبکه‌های عصبی اقتباسی بسیار ساده از هوش انسان بوده و می‌توانند در مسائلی مثل شناخت الگو، خوشه‌بندی، دسته‌بندی، تخمین تابع، مدل-سازی، کنترل و به طور کلی در هر جا که نیاز به یادگیری یک نگاشت خطی و یا غیر خطی باشد، مورد استفاده قرار گیرند. این قابلیت‌ها سبب شده است که محققان برای مدل‌سازی رفتار مصالح که وابسته به خصوصیات فیزیکی و شیمیایی اجزای تشکیل دهنده‌ی آنها هستند، به استفاده از شبکه‌های عصبی روی آورند.

۳-۳- معیارهای ارزیابی عملکرد

در این تحقیق برای بررسی عملکرد مدل‌های شبکه عصبی ایجاد شده از معیارهای زیر استفاده شده است:

- جزر میانگین مربعات خطا ($RMSE$):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \quad (۵)$$

در رابطه (۵)، x_i داده‌های اندازه‌گیری شده (مقادیر واقعی)، y_i داده‌های پیش‌بینی شده (خروجی‌های شبکه)، $\bar{x}$ میانگین داده‌های اندازه‌گیری شده و $\bar{y}$ میانگین داده‌های پیش‌بینی شده است.

- ماتریس در هم ریختگی[۱۰]: این ماتریس یکی از روش‌های رایج بررسی صحت عملکرد الگوریتم‌هایی است که از یادگیری نظارت شده استفاده می‌کنند. ماتریس در هم ریختگی یک ماتریس مربعی $N \times N$ است که N بیانگر تعداد کلاس‌ها است. در این ماتریس روی محور افقی کلاس‌های واقعی، و روی محور عمودی کلاس‌های تخمینی الگوریتم نشان داده می‌شود. اعداد روی قطر اصلی ماتریس نشان دهنده تعداد کلاس بندی‌های انجام شده به صورت صحیح است. بنابراین در صورتی که تمام اعداد قرار گرفته روی درایه‌های غیر قطر اصلی صفر باشند، الگوریتم دارای دقت حداکثر است.
- دقت: پارامتر دقت به صورت نسبت تعداد نمونه‌های کلاس‌بندی شده به صورت صحیح بر تعداد کل نمونه‌های مجموعه داده تعریف می‌شود.

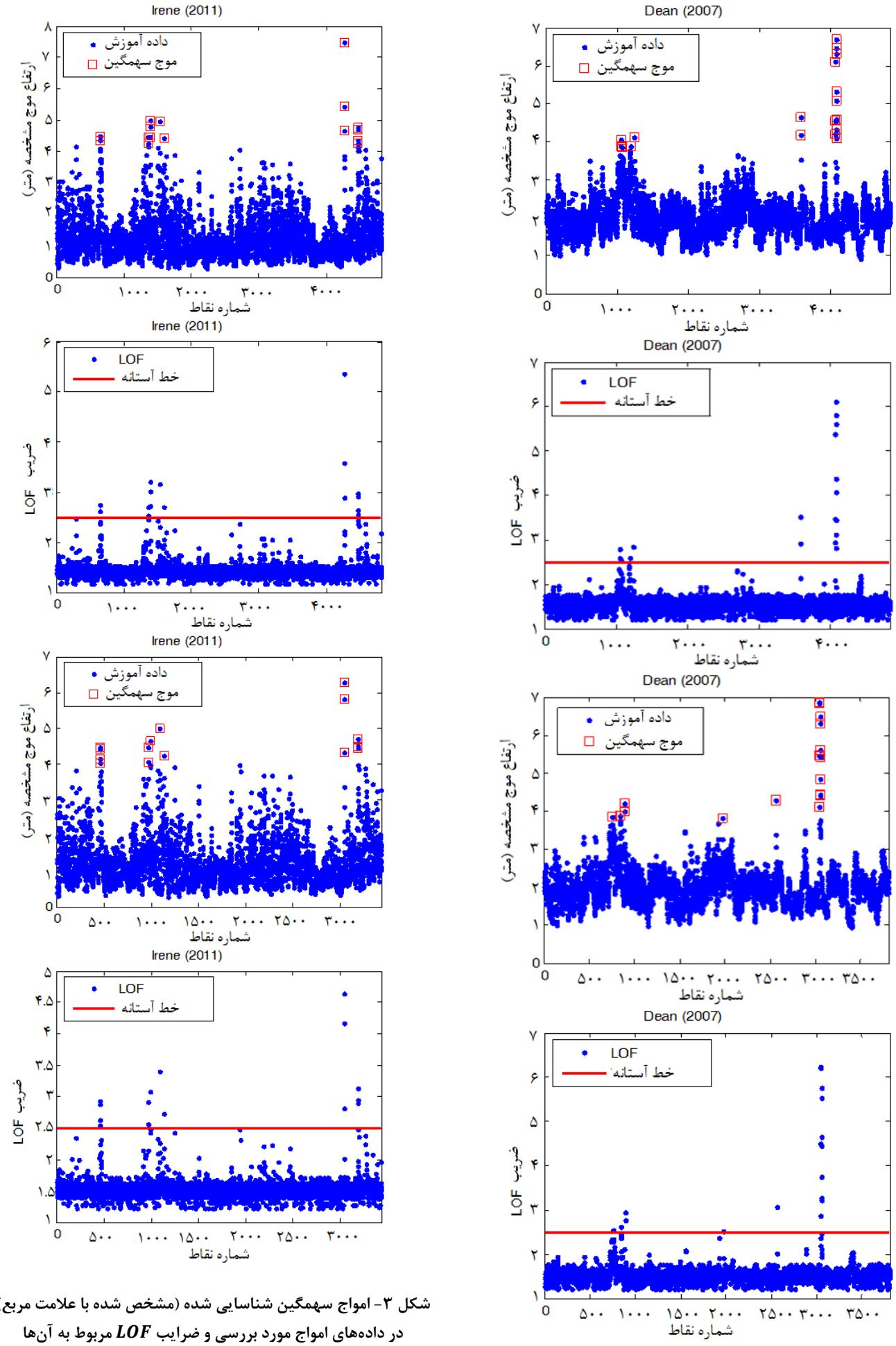
منحنی ROC[10]: این منحنی نیز یکی از شاخص‌های ارزیابی عملکرد الگوریتم‌هایی است که از یادگیری نظارت شده و یادگیری بدون ناظر استفاده می‌کنند. در این نمودار هر چه نقاط به سمت بالا و چپ نزدیک‌تر باشند، مدل ایجاد شده دارای دقت بیشتری بوده و به حالت ایده‌آل خود نزدیک‌تر است. مختصات نقطه‌ی (۱و۰) حالت ایده‌آل است. این نقطه بدین معنا است که خروجی-های مدل با مقادیر واقعی تطابق کامل دارد. هرچه مساحت زیر نمودار منحنی‌های ROC بیشتر باشد، مدل ایجاد شده دارای عملکرد بهتری است.

۴- نتایج

در این بخش عملکرد روش ترکیبی LOF-ANN برای پیش‌بینی تولید امواج سهمگین با استفاده از متغیرهای هواشناسی بررسی شده است. در این تحقیق برای انجام تمامی تجزیه و تحلیل‌ها از نرم‌افزار MATLAB استفاده شده است.

برای پیش‌بینی تولید امواج سهمگین ابتدا باید امواج سهمگین از امواج غیر سهمگین تمییز داده شود. این کار با استفاده از الگوریتم LOF قابل انجام است. اجرای این روش نیازمند تعیین پارامترهای آن، یعنی تعداد همسایه‌های انتخابی (k) و پارامتر آستانه (t) است. تا کنون هیچ قانون مشخص و پذیرفته شده‌ای برای انتخاب این پارامترها ارائه نشده است و اگر ارائه شده باشد نیز نسبی است و در تمامی حالات معتبر نیست. معمولا مقدار k بر حسب ماهیت داده-های مورد بررسی و با روش آزمون و خطا انتخاب می‌شود. در حالت کلی می‌توان گفت اگر تعداد همسایه‌های انتخابی به تعداد قابل توجهی کم انتخاب شود آنگاه دقت روش در تشخیص نمونه‌های غیر نرمال پایین خواهد آمد و ممکن است شماری از داده‌های نرمال به طور اشتباه به عنوان غیر نرمال تشخیص داده شوند و بالعکس. اگر تعداد همسایه‌های انتخابی از یک حد مشخص بالاتر انتخاب شود آنگاه دقت روش تغییری چندانی نخواهد کرد و فقط حجم محاسبات و زمان اجرای برنامه کامپیوتری به طور قابل ملاحظه‌ای افزایش خواهد یافت. بنابراین مقدار k نباید خیلی کم و نه خیلی زیاد انتخاب شود. مقدار t نیز باید عددی بزرگتر از ۱ انتخاب شود. در اینجا با روش آزمون و خطا، مقدار k برابر ۱/۲۰ تعداد اعضای مجموعه داده، و مقدار t برابر ۲/۵ انتخاب شده است. در شکل ۳ نتایج اجرای روش LOF روی مقادیر ارتفاع موج مشخصه مجموعه داده‌های آموزشی و آزمایشی طوفان‌های Dean و Irene نشان داده شده است. داده‌های شناسایی شده به عنوان امواج سهمگین با علامت مربع در شکل نشان داده شده‌اند. همانطور که از این شکل مشخص است، داده‌های کاندیدای امواج سهمگین به طور قابل ملاحظه‌ای متناقض با سایر امواج هستند. در زیر هر شکل نمودار ضرایب LOF محاسبه شده برای نقاط آن مجموعه داده نیز نشان داده شده است. همانطور که از این شکل‌ها مشخص است در مواقعی که یک موج سهمگین رخ داده، ضرایب LOF مربوط به آن‌ها نیز به طور قابل توجهی نسبت به سایر ضرایب افزایش یافته است. از خروجی‌های روش LOF به عنوان داده‌های هدف (کلاس‌های امواج) در روش شبکه عصبی مصنوعی دسته-بندی کننده استفاده می‌شود. کلاس امواج به کلاس‌های سهمگین (معرفی شده با ۱) و نرمال (معرفی شده با ۲) تقسیم‌بندی شده است.

[Downloaded from marine-eng.ir on 2026-02-07]
[DOR: 20.1001.1.17357608.1398.15.30.2.2]
[DOI: 10.29252/marineeng.15.30.23]



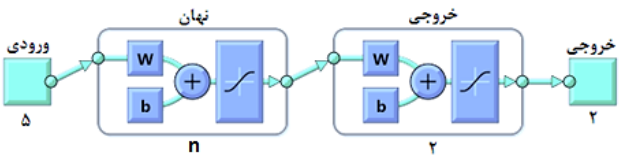
شکل ۳- امواج سهمگین شناسایی شده (مشخص شده با علامت مربع) در داده‌های امواج مورد بررسی و ضرایب *LOF* مربوط به آن‌ها

[Downloaded from marine-eng.ir on 2026-02-07]
[DOR: 20.1001.1.17357608.1398.15.30.2.2]
[DOI: 10.29252/marineeng.15.30.23]

پس از مشخص شدن کلاس داده‌ها، می‌توان شبکه عصبی را طراحی کرد. از آنجایی که یک شبکه سه لایه قادر به حل هرگونه مسئله دسته‌بندی است، بنابراین در این تحقیق از یک شبکه سه لایه تغذیه رو به جلو برای پیش‌بینی وقوع امواج سهمگین استفاده شده است. برای طراحی شبکه، باید آن را تحت پارامترهای موثر بر عملکرد آن تحت آموزش و یادگیری قرار داد و در نهایت بهینه‌ترین پارامترها (که منجر به بهترین پاسخ می‌شوند) را انتخاب کرد. برای مدل‌سازی داده‌های امواج طوفان‌های Dean و Irene دو مدل مختلف شبکه عصبی طراحی شده است. ساختار این دو شبکه یکسان بوده و تفاوت آنها در وزن‌های ارتباطی بین گره‌ها است، چون مجموعه داده‌های آموزشی هر شبکه متفاوت است. هر مدل دارای ۵ متغیر ورودی (جهت باد (WDIR)، سرعت باد (WSPD)، فشار سطح دریا (PRES)، دمای هوا (ATMP) و دمای سطح دریا (WTMP)) و ۲ متغیر خروجی (کلاس امواج) است. برای ایجاد شبکه با بهترین کارایی، باید آن را با پارامترهای مختلف شبکه آموزش داد. این کار با روش آزمون و خطا قابل انجام است. در جدول ۳ مشخصات ساختار شبکه‌های طراحی شده، و در شکل ۴ معماری شبکه بهینه ارائه شده است. در این شکل W بیانگر وزن‌ها، b بیانگر بایاس، و n بیانگر تعداد نرون‌های لایه میانی است. در اینجا برای ایجاد شبکه مطلوب، تمامی پارامترهای شبکه‌ها ثابت در نظر گرفته شده، فقط تعداد نرون‌های لایه میانی از ۶ تا ۱۲ تغییر کرده است.

جدول ۳- مشخصات ساختار شبکه‌های عصبی طراحی شده

پارامتر	مقدار
نوع شبکه	سه لایه تغذیه‌رو به جلو
تعداد متغیرهای ورودی	۵
تعداد نرون‌های لایه ورودی	۵
تعداد نرون‌های لایه میانی	متغیر از ۶ تا ۱۲
تعداد متغیرهای خروجی	۲
تعداد نرون‌های لایه خروجی	۲
تعداد داده‌های آموزش[۱۲]	۷۰ درصد کل داده‌ها
تعداد داده‌های اعتبار سنجی[۱۳]	۱۵ درصد کل داده‌ها
تعداد داده‌های آزمون[۱۴]	۱۵ درصد کل داده‌ها
الگوریتم آموزش	Scaled conjugate gradient
تابع فعال‌سازی لایه نهان	Hyperbolic tangent sigmoid
تابع فعال‌سازی لایه خروجی	Hyperbolic tangent sigmoid
تابع اندازه‌گیری خطا	میانگین مربعات خطا (MSE)
نرخ آموزش	۰/۰۵
حداکثر تعداد تکرارها	۱۰۰۰



شکل ۴- معماری شبکه عصبی بهینه طراحی شده برای مدل‌ها

در جدول ۴ و جدول ۵ نتایج مدل‌های ایجاد شده با استفاده از شبکه عصبی دسته‌بندی کننده به ازای تعداد نرون‌های مختلف روی مجموعه داده‌های آموزشی و آزمایشی ارائه شده است. در این جدول مدل‌های بهینه با رنگ تیره‌تر مشخص شده‌اند. هرچه مقدار پارامتر دقت به ۱۰۰ (بر حسب درصد)، و خطا به ۰ نزدیک‌تر باشد، شبکه طراحی شده از کارایی مطلوب‌تری برخوردار است. برای کسب مطلوب‌ترین مدل به ازای هر نرون، شبکه چندین بار تحت پارامترهای شبکه مورد آموزش و یادگیری قرار گرفته و در نهایت مناسب‌ترین مدل که منجر به بهترین پاسخ شده به عنوان مدل نهایی در نظر گرفته شده است. بهترین مدل برای طوفان Dean و Irene به ترتیب به ازای تعداد نرون‌های ۱۰ و ۱۲ حاصل شده است. در شکل‌های ۷ و ۸، نمودارهای عملکرد شبکه بهینه طراحی شده برای مجموع داده‌های مورد بررسی ارائه شده است. نمودارها شامل موارد زیر است:

- شکل الف: نمودار در هم ریختگی داده‌های آموزش، داده‌های اعتبارسنجی، داده‌های آزمون و تمامی داده‌ها برای مجموعه داده آموزشی. در این شکل تعداد نمونه‌هایی که به صورت صحیح دسته‌بندی شده‌اند روی قطر اصلی ماتریس و در خانه-های سبز رنگ قرار می‌گیرند. خانه واقع شده در پایین و سمت راست ماتریس (مشخص شده با رنگ آبی) بیانگر میزان دقت کلی مدل است.
- شکل ب: منحنی ROC داده‌های آموزش، داده‌های اعتبارسنجی، داده‌های آزمون و تمامی داده‌ها برای مجموعه داده آموزشی. در این شکل، خطوط رنگی در هر محور بیانگر منحنی‌های ROC برای هر کلاس است. هرچه این خطوط به سمت چپ گوشه بالا نمودار متمرکز شده باشند، بدین معنا است که عملکرد مدل ایجاد شده مناسب‌تر است.
- شکل ج: منحنی ROC مجموعه داده آزمایشی. مدل‌های ایجاد شده با استفاده از مجموعه داده آموزشی باید روی نمونه‌های جدید نیز کارایی خوبی داشته باشند و قادر به پیش‌بینی شرایط جدید و نمونه‌های ناشناخته باشند، در غیر این صورت مدل ایجاد شده مطلوب نیست.
- شکل د: نمودار عملکرد آموزش شبکه عصبی برای مجموعه داده آموزشی. این نمودار بیانگر روند آموزش مدل ایجاد شده

[Downloaded from marine-eng.ir on 2026-02-07]
[DOR: 20.1001.1.17357608.1398.15.30.2.2]
[DOI: 10.29252/marineeng.15.30.23]

است. تعداد تکرارهای کل و تعداد تکرار بهینه (حصول کمترین خطا) در این نمودار مشخص است.

همانطور که از نتایج جداول ۴ و ۵ و شکل‌های ۷ و ۸ مشخص است، تمامی مدل‌های تهیه شده برای مجموع داده‌های مورد بررسی با دقت قابل قبولی ایجاد شده‌اند. عملکرد مدل‌ها روی داده-های آزمایشی نیز مطلوب بوده است. بنابراین از مدل‌های ایجاد شده می‌توان با دقت مناسبی برای پیش‌بینی وقوع امواج سهمگین در موارد آتی استفاده کرد.

در جدول ۶ خلاصه عملکرد بهترین مدل‌های ایجاد شده در تشخیص امواج سهمگین از غیر سهمگین (نرمال) بر اساس نتایج حاصل از ماتریس در هم ریختگی نشان داده شده است. با توجه به نتایج این جدول می‌توان گفت به طور کلی دقت مدل‌ها در تشخیص امواج غیر سهمگین از امواج سهمگین بیشتر است. علت این است که تعداد نمونه‌های آموزشی امواج غیر سهمگین بیشتر از امواج سهمگین بوده است و مدل توانسته همبستگی بیشتری بین امواج غیر سهمگین پیدا کند. همچنین یکی دیگر از علت‌های اینکه مدل‌ها در شناسایی برخی از امواج سهمگین ناموفق بوده‌اند این است که حین برچسب‌گذاری امواج با استفاده از الگوریتم LOF، برخی از امواجی که به امواج غیر سهمگین شباهت زیادی داشته‌اند به عنوان موج سهمگین در نظر گرفته شده‌اند. به عنوان مثال همانطور که از شکل ۳ مشخص است، در داده‌های آموزشی طوفان Dean شماری از امواج با ارتفاع تقریبی ۴ متر توسط الگوریتم LOF به عنوان موج سهمگین شناسایی شده‌اند. ارتفاع ۴ به توجه به توزیع آماری داده‌ها با احتمال زیادی منطقی و غیر سهمگین به نظر می‌رسد. به همین دلیل مدل ارائه شده برای داده‌های این طوفان، داده‌های با ارتفاع تقریبی ۴ متر را به عنوان موج نرمال قلمداد کرده است. بنابراین با افزایش مقدار پارامتر آستانه در روش LOF می‌توان حساسیت این روش به داده‌های غیر نرمال را کاهش داد، در نتیجه داده‌هایی که اختلاف آن‌ها با سایر اعضا به طور قابل ملاحظه‌ای بیشتر باشد به عنوان موج سهمگین در نظر گرفته می‌شوند، در نتیجه با این کار دقت مدل‌های تولیدی افزایش می‌یابد. به عنوان مثال برای داده‌های آموزشی طوفان Dean، در روش LOF اگر مقدار آستانه از ۲/۵ به ۳/۵ افزایش یابد، داده‌های با ارتفاع تقریبی ۴ متر دیگر به عنوان موج سهمگین در نظر گرفته نمی‌شوند (شکل ۹) و کلاس‌بندی توسط مدل‌ها با دقت بیشتری انجام خواهد شد. در شکل ۱۰ نمودارهای ماتریس در هم ریختگی و منحنی ROC شبکه بهینه طراحی شده برای داده‌های طوفان Dean به ازای مقدار پارامتر آستانه ۳/۵ و تعداد نرون‌های ۱۰، نشان داده شده است. با توجه به این نمودارها می‌توان گفت که دقت مدل ایجاد شده افزایش یافته و در داده‌های آموزشی و آزمایشی از ۸ موج سهمگین موجود در داده‌ها، به ترتیب ۲ و ۱ موج به طور اشتباه به عنوان موج غیر سهمگین در نظر گرفته شده است.

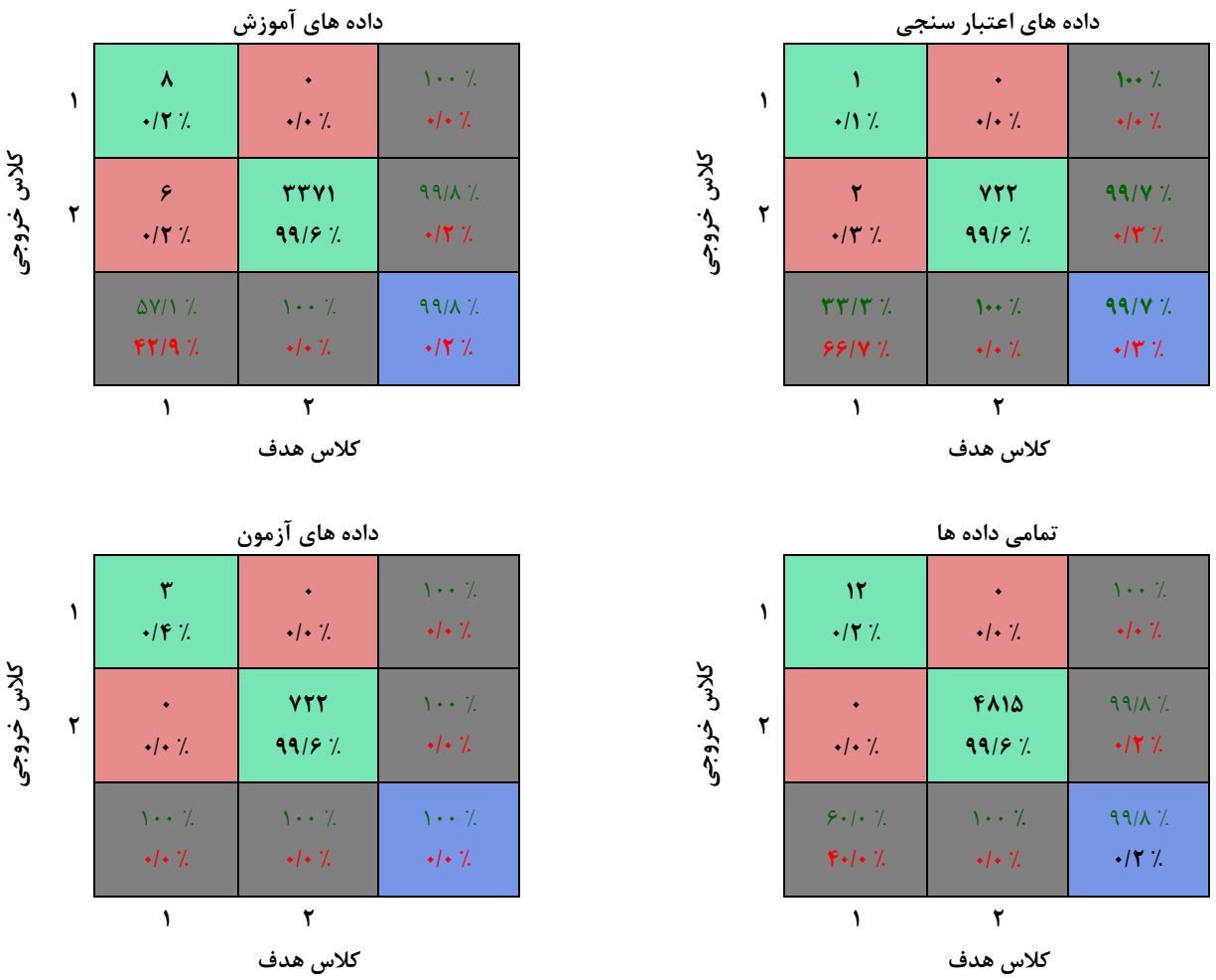
جدول ۴- نتایج مدل‌سازی پیش‌بینی امواج سهمگین طوفان Dean (2007) با استفاده از شبکه عصبی مصنوعی دسته‌بندی کننده

تعداد نرون-ها	مجموعه داده آموزشی				مجموعه داده آزمایشی	
	تعداد تکرارها	تعداد تکرار بهینه	خطا	دقت (درصد)	خطا	دقت (درصد)
۶	۲۵	۱۹	۰/۰۴۴	۹۹/۸۰	۰/۰۴۵	۹۹/۸۰
۷	۴۵	۳۹	۰/۰۴۰	۹۹/۸۰	۰/۰۴۲	۹۹/۸۰
۸	۵۸	۵۲	۰/۰۴۳	۹۹/۹۰	۰/۰۴۲	۹۹/۸۰
۹	۱۷	۱۱	۰/۰۵۹	۹۹/۶۰	۰/۰۶۰	۹۹/۶۰
۱۰	۳۸	۳۲	۰/۰۴۱	۹۹/۸۰	۰/۰۴۱	۹۹/۸۰
۱۱	۲۴	۱۸	۰/۰۴۵	۹۹/۸۰	۰/۰۴۸	۹۹/۷۰
۱۲	۳۷	۳۱	۰/۰۳۹	۹۹/۹۰	۰/۰۴۴	۹۹/۸۰

جدول ۵- نتایج مدل‌سازی پیش‌بینی امواج سهمگین طوفان Irene (2007) با استفاده از شبکه عصبی مصنوعی دسته‌بندی کننده

تعداد نرون-ها	مجموعه داده آموزشی				مجموعه داده آزمایشی	
	تعداد تکرارها	تعداد تکرار بهینه	خطا	دقت (درصد)	خطا	دقت (درصد)
۶	۴۱	۳۵	۰/۰۳۵	۹۹/۸۰	۰/۰۴۴	۹۹/۷۰
۷	۲۱	۱۴	۰/۰۵۶	۹۹/۷۰	۰/۰۶۵	۹۹/۶۰
۸	۱۶	۹	۰/۰۵۱	۹۹/۷۰	۰/۰۵۸	۹۹/۶۰
۹	۴۴	۳۸	۰/۰۴۸	۹۹/۷۰	۰/۰۵۶	۹۹/۶۰
۱۰	۶۲	۵۶	۰/۰۳۵	۹۹/۹۰	۰/۰۴۳	۹۹/۸۰
۱۱	۴۰	۳۴	۰/۰۳۶	۹۹/۹۰	۰/۰۴۷	۹۹/۸۰
۱۲	۶۴	۵۸	۰/۰۳۶	۹۹/۹۰	۰/۰۳۹	۹۹/۹۰

[Downloaded from marine-eng.ir on 2026-02-07]
[DOR: 20.1001.1.17357608.1398.15.30.2.2]
[DOI: 10.29252/marineeng.15.30.23]



(الف): نمودار در هم ریختگی داده‌های آموزش، داده‌های اعتبارسنجی، داده‌های آزمون و تمامی داده‌ها برای مجموعه داده آموزشی

[Downloaded from marine-eng.ir on 2026-02-07]

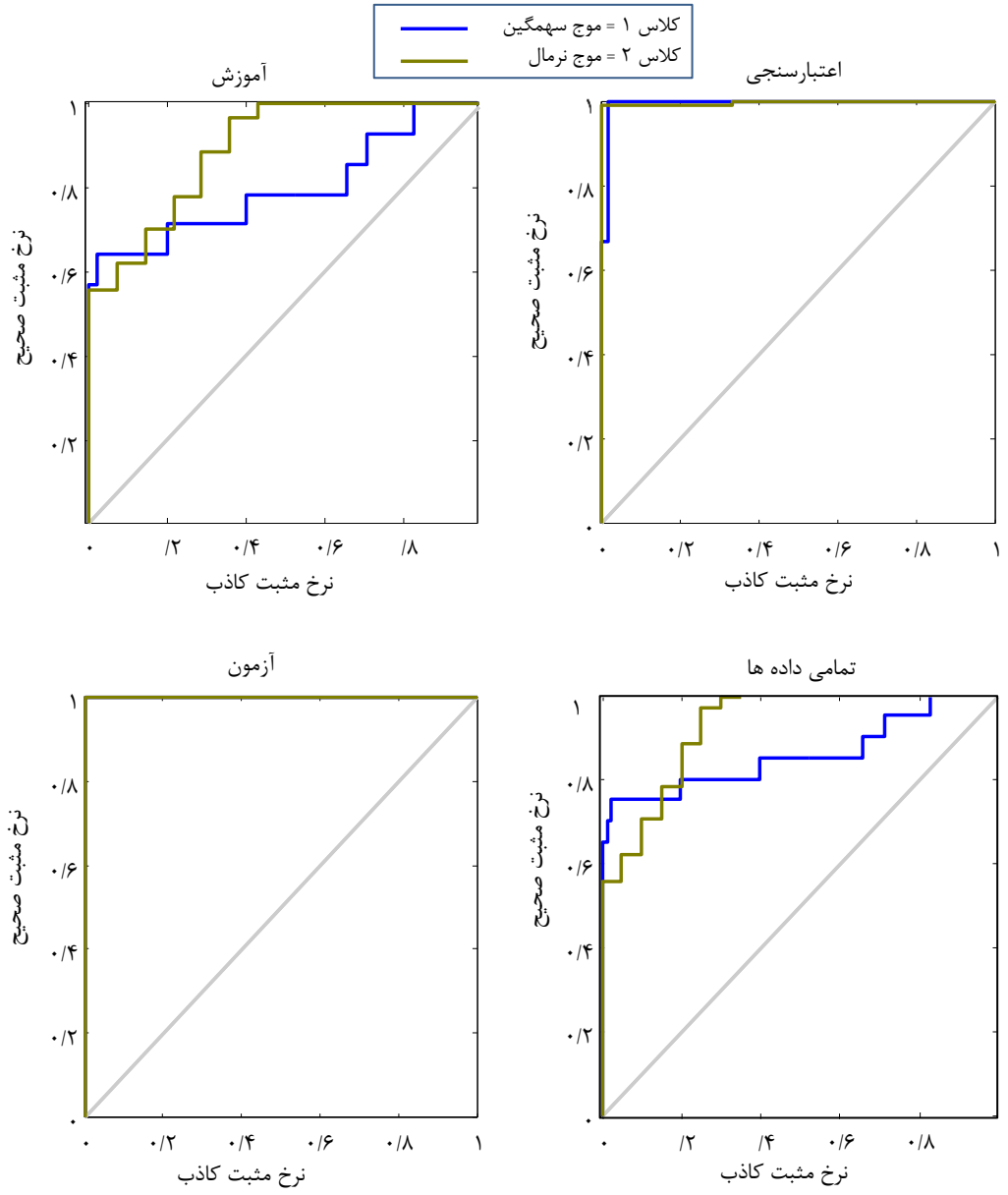
[DOR: 20.1001.1.17357608.1398.15.30.2.2]

[DOI: 10.29252/marineeng.15.30.23]

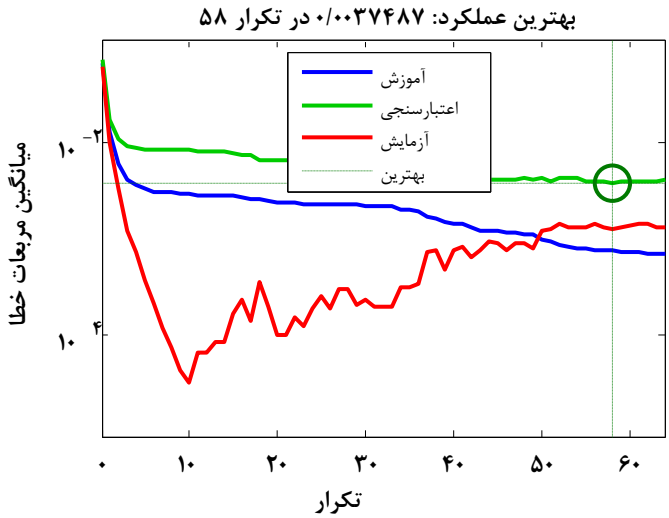
[Downloaded from marine-eng.ir on 2026-02-07]

[DOR: 20.1001.1.17357608.1398.15.30.2.2]

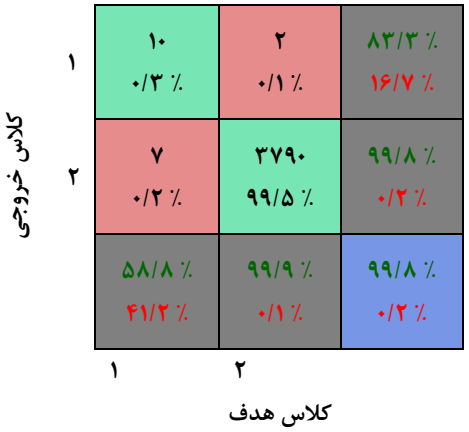
[DOI: 10.29252/marineeng.15.30.23]



(ب): منحنی ROC داده‌های آموزش، داده‌های اعتبارسنجی، داده‌های آزمون و تمامی داده‌ها برای مجموعه داده آموزشی

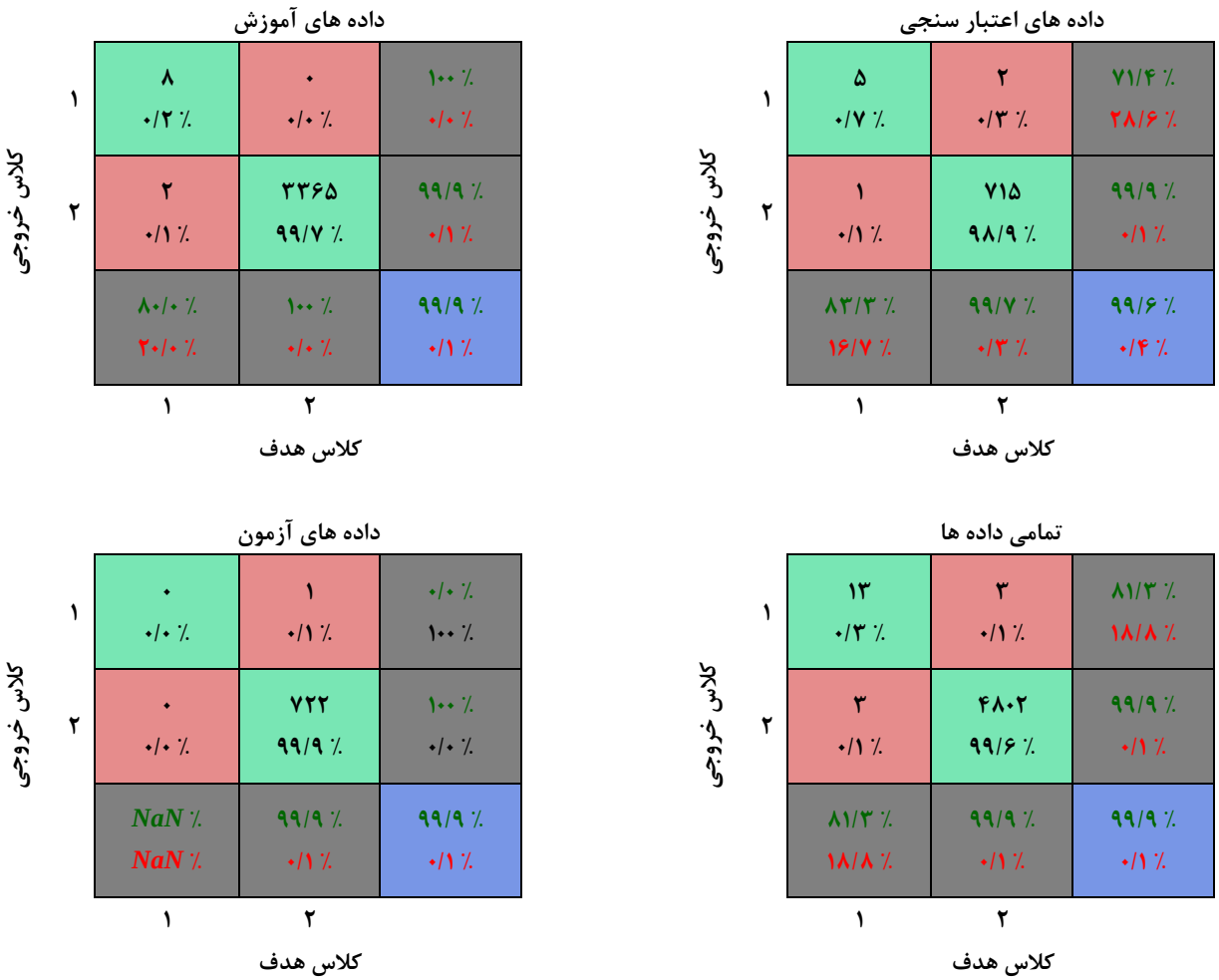


(د): نمودار عملکرد آموزش شبکه عصبی برای مجموعه داده آموزشی



(ج): منحنی در هم ریختگی مجموعه داده آزمایشی

شکل ۷- نمودارهای عملکرد شبکه بهینه طراحی شده برای مجموعه داده طوفان Dean (2007)



(الف): نمودار در هم ریختگی داده‌های آموزش، داده‌های اعتبارسنجی، داده‌های آزمون و تمامی داده‌ها برای مجموعه داده آموزشی

[Downloaded from marine-eng.ir on 2026-02-07]

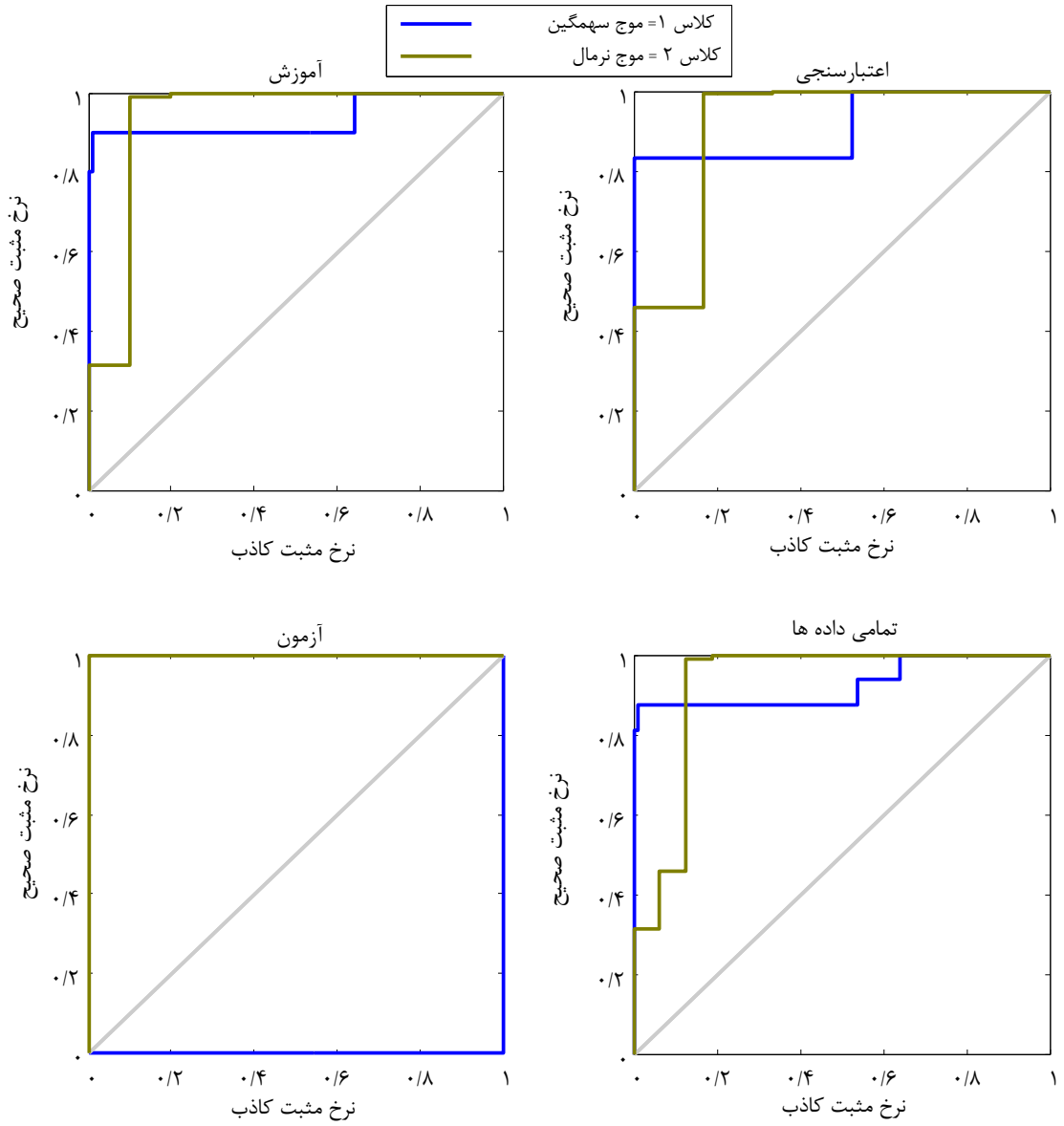
[DOR: 20.1001.1.17357608.1398.15.30.2.2]

[DOI: 10.29252/marineeng.15.30.23]

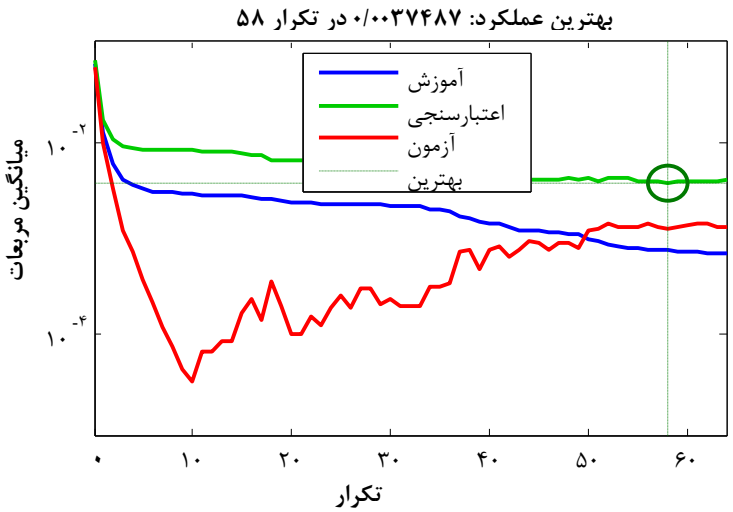
[Downloaded from marine-eng.ir on 2026-02-07]

[DOR: 20.1001.1.17357608.1398.15.30.2.2]

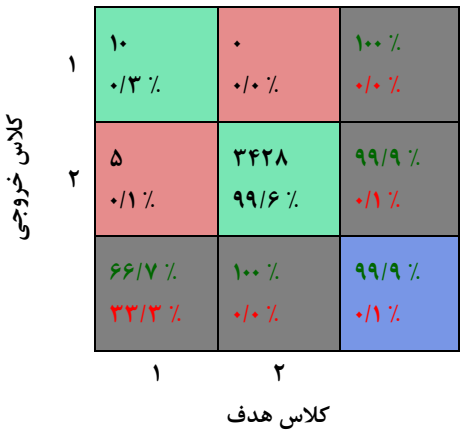
[DOI: 10.29252/marineeng.15.30.23]



(ب): منحنی ROC داده‌های آموزش، داده‌های اعتبارسنجی، داده‌های آزمون و تمامی داده‌ها برای مجموعه داده آموزشی



(د): نمودار عملکرد آموزش شبکه عصبی برای مجموعه داده آموزشی

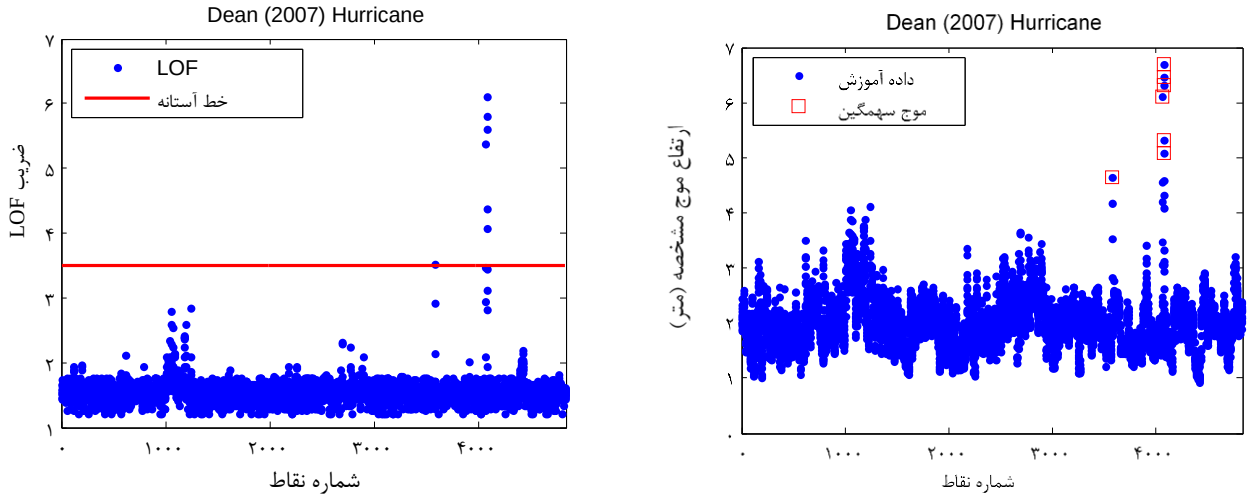


(ج): منحنی در هم ریختگی مجموعه داده آزمایشی

شکل ۸- نمودارهای عملکرد شبکه بهینه طراحی شده برای مجموعه داده طوفان Irene (2011)

جدول ۶- خلاصه عملکرد بهترین مدل‌های ایجاد شده برای امواج مورد بررسی در تشخیص امواج سهمگین از غیر سهمگین (نرمال)

طوفان	مجموعه داده	تعداد امواج نرمال شناسایی شده به عنوان نرمال	تعداد امواج نرمال شناسایی شده به عنوان سهمگین	تعداد امواج سهمگین شناسایی شده به عنوان سهمگین	تعداد امواج سهمگین شناسایی شده به عنوان نرمال	تعداد پیش-بینی‌های صحیح	تعداد پیش-بینی‌های اشتباه
Dean	آموزشی	۴۸۱۵	۰	۱۲	۸	۴۸۲۷	۸
	آزمایشی	۳۷۹۰	۲	۱۰	۷	۳۸۰۰	۹
Irene	آموزشی	۴۸۰۲	۳	۱۳	۳	۴۸۱۵	۶
	آزمایشی	۳۴۲۸	۰	۱۰	۵	۳۴۳۸	۵

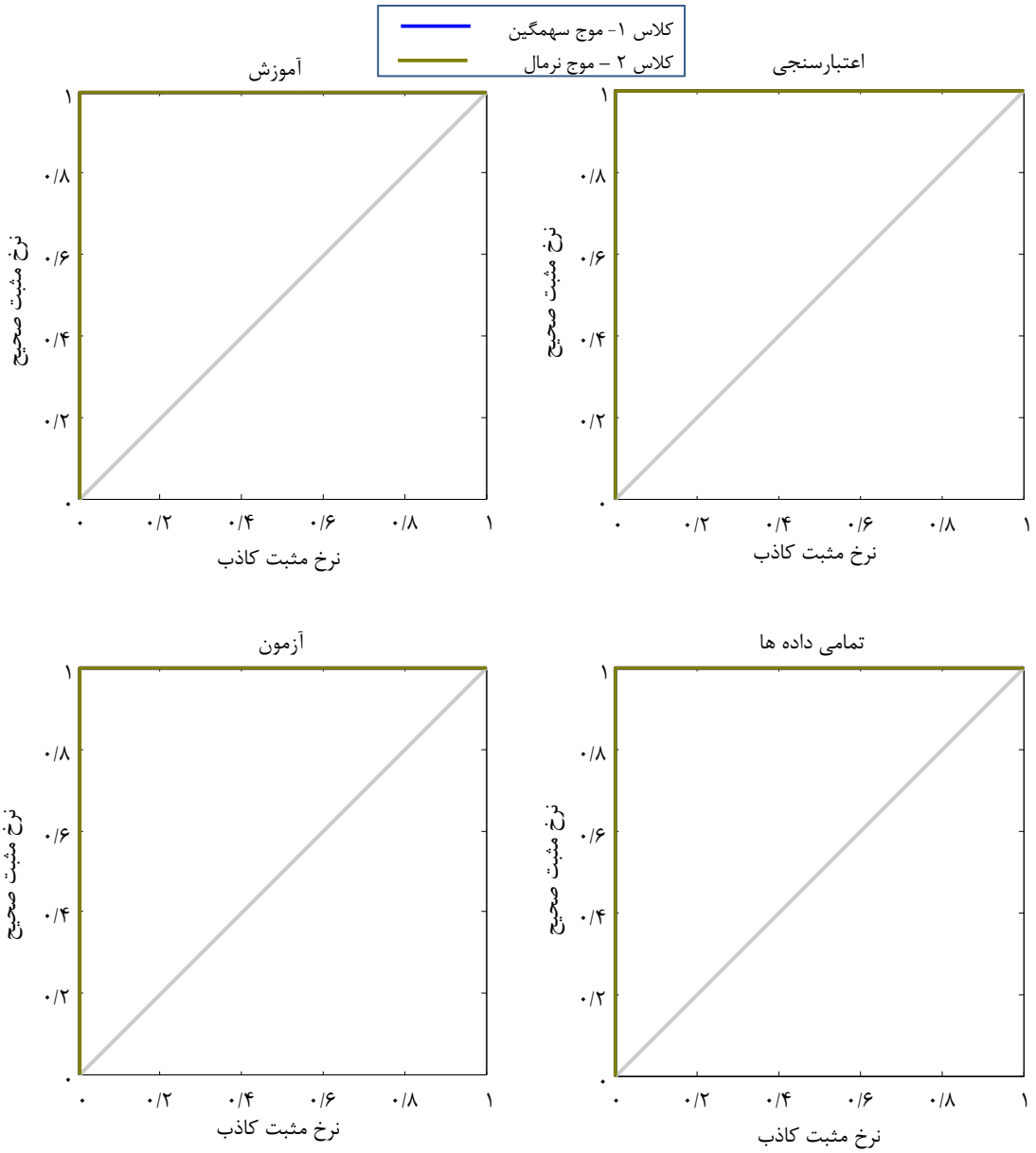


(الف): نمودار امواج سهمگین و غیر سهمگین (ب): نمودار ضرایب **LOF**

شکل ۹- امواج سهمگین شناسایی شده (مشخص شده با علامت مربع) در داده‌های آموزشی طوفان Dean به ازای مقدار آستانه ۳/۵

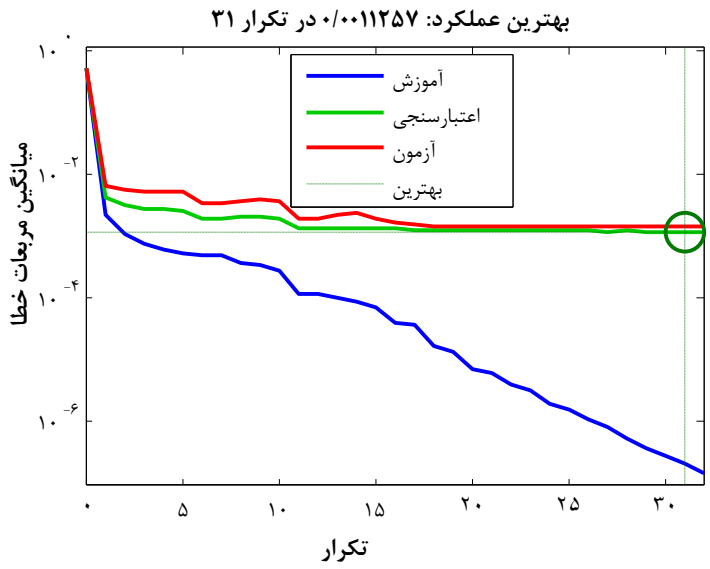
(الف): نمودار در هم ریختگی داده‌های آموزش، داده‌های اعتبارسنجی، داده‌های آزمون و تمامی داده‌ها برای مجموعه داده آموزشی

[Downloaded from marine-eng.ir on 2026-02-07]
[DOR: 20.1001.1.17357608.1398.15.30.2.2]
[DOI: 10.29252/marineeng.15.30.23]

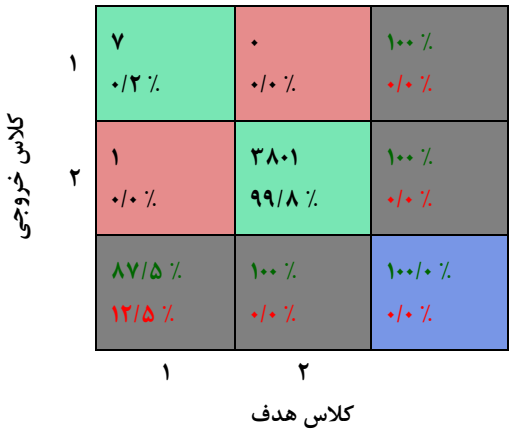


(ب): منحنی ROC داده‌های آموزش، داده‌های اعتبارسنجی، داده‌های آزمون و تمامی داده‌ها برای مجموعه داده آموزشی

[Downloaded from marine-eng.ir on 2026-02-07]
[DOR: 20.1001.1.17357608.1398.15.30.2.2]
[DOI: 10.29252/marineeng.15.30.23]



(د): نمودار عملکرد آموزش شبکه عصبی برای مجموعه داده آموزشی



(ج): منحنی در هم ریختگی مجموعه داده آزمایشی

شکل ۱۰- نمودارهای عملکرد شبکه بهینه طراحی شده برای مجموعه داده آموزشی طوفان Dean به ازای مقدار آستانه ۳/۵

پس از ایجاد مدل‌ها، می‌توان از آن‌ها برای پیش‌بینی وقوع امواج سهمگین با استفاده از متغیرهای هواشناسی موجود استفاده کرد، در نتیجه منجر به صرفه‌جویی در هزینه‌ها و زمان شده و در مواقعی که فقط اطلاعات مربوط به متغیرهای هواشناسی در دسترس باشد می‌توان بدون صرف هزینه و در کمترین زمان ممکن وقوع امواج سهمگین را پیش‌بینی کرد.

۵- نتیجه‌گیری

امواج طوفانی می‌تواند برای سازه‌های دریایی و شناورها بسیار خطرناک باشد. به دلیل ماهیت بسیار پیچیده و غیر خطی این امواج، معمولا ارائه مدلی که بتواند وقوع امواج طوفانی را شناسایی کند دشوار است. در این تحقیق با استفاده از مدل ترکیبی ضریب داده پرت محلی و شبکه‌های عصبی دسته‌بندی کننده سه لایه تغذیه‌رو به جلو (LOF-ANN) مدل‌هایی بر اساس داده‌های امواج دو طوفان مشهور Dean (2007) و Irene (2011) طی سه سال، برای پیش‌بینی وقوع امواج سهمگین با استفاده از متغیرهای هواشناسی تهیه شد. متغیرهای هواشناسی مورد استفاده در مدل-سازی‌ها عبارتند از: جهت باد، سرعت باد، ارتفاع موج مشخصه، فشار سطح دریا، دمای هوا و دمای سطح دریا. مدل‌های ایجاد شده توانستند با دقت خوبی رابطه بین ورودی‌ها و خروجی‌ها را تخمین بزنند به طوری که برای هر دو مجموعه داده مورد بررسی دقت میانگین به بالای ۹۹٪ رسید. برای مجموعه داده Dean، بهترین مدل به ازای تعداد نرون‌های لایه میانی برابر ۱۰ حاصل شد و مقدار خطای داده‌های آموزشی و آزمایشی ۰/۰۴۱ بدست آمد. همچنین برای مجموعه داده Irene، بهترین مدل به ازای تعداد نرون‌های ۱۲ بدست آمد که مقدار خطای داده‌های آموزشی و آزمایشی به ترتیب ۰/۰۳۶ و ۰/۰۳۹ حاصل شد. بنابراین با استفاده از روند معرفی شده در این تحقیق می‌توان با صرف کمترین هزینه و زمان و با دقت خوبی وقوع امواج سهمگین را از روی متغیرهای هواشناسی پیش-بینی کرد. صحت عملکرد روش ارائه شده در تشخیص موج سهمگین از غیر سهمگین تا حد زیادی وابسته به صحت اجرای الگوریتم LOF است. صحت اجرای این الگوریتم وابسته به انتخاب صحیح پارامترهای آن است (تعداد همسایه‌ها (k) و پارامتر آستانه (t)) است. انتخاب صحیح این دو پارامتر وابسته به ماهیت داده‌های مورد بررسی بوده و بر حسب تجربه شخصی و شناخت قبلی از داده‌ها تعیین می‌شود. با توجه به نتایج این تحقیق و به طور شهودی پیشنهاد می‌شود مقدار k عددی صحیح بین ۱/۲۰ تا ۱/۱۰ تعداد داده‌های مورد بررسی انتخاب شود. مقدار t نیز عددی بزرگتر از ۲ انتخاب شود. هرچه t بزرگتر انتخاب شود، امواجی که تناقض بیشتری با سایر امواج دارند به عنوان امواج سهمگین برچسب‌گذاری می‌شوند. این عامل باعث ایجاد مدل‌هایی با دقت بالاتر می‌شود.

زیرنویس‌ها

1. Auto Regressive Moving Average (ARMA)
2. Auto Regres-sive Integrated Moving Average (ARIMA)
3. Kalman Filter
4. Local Outlier Factor (LOF)
5. Artificial Neural Network (ANN) classifier
6. k–distance neighborhood of p
7. Reachability distance of p w.r.t object o
8. Local reachability density of p
9. Feed-Forward
10. Confusion matrix
11. Receiver Operating Characteristic (ROC)
12. Training data

[Downloaded from marine-eng.ir on 2026-02-07]
[DOR: 20.1001.1.17357608.1398.15.30.2.2]
[DOI: 10.29252/marineeng.15.30.23]

13. Validation data
14. Test data

۶ – مراجع

1- Deo, M.C., Jha, A., Chaphekar, A.S., and Ravikant, K, (2001), *Wave prediction using neuralnetworks*, Journal of Ocean Engineering, Vol. 28, No. 7, pp. 889–898.

2- Komen, G.J., Cavaleri, L., Donelan, M., Hasselmann, K., and Hasselmann, S., P. Janssen, (1994), *Dynamics and Modelling of Ocean Waves*, Cambridge University Press.

3- Londhe, S.N., and Panchang, V.G., (2006), *One-day wave forecasts based on artificial neural networks*, Journal of Atmospheric and Oceanic Technology. AMS, Vol. 23, No. 11, pp. 1593–1603.

4- Ponce de León, S., and Guedes Soares, C. (2015), *Hindcast of extreme sea states in North Atlantic Extratropical storms*, Jourbal of Ocean Dynamics, Vol. 65, No. 2, pp. 241–254.

5- Tolman, H.L., (2014), *The WAVEWATCH III Development Group*, User manual and system documentation version 4018, Tech. Note 316, NOAA/NWS/NCEP/MMAB. 282.

6- Jain, P., and Deo, M.C., (2007), *Real-time wave forecasts off The Western Indian Coast*, Journal of Applied Ocean Research. Vol. 29, pp. 72–79.

7- Mahmoodi, K., Ghassemi, H., and Nowruzi, H., (2017), *Data mining models to predict ocean wave energy flux in the absence of wave records*, Science Journal of Maritime University of Szczecin, Vol. 49, No. 49, pp. 119-129.

8- Berbić, J., Ocvirk, E., Carević, D., and Lončar, G., (2017), *Application of neural networks and support vector machine for significant wave height prediction*, Journal of Oceanologia, Vol. 59, pp. 331—349.

9- Nikoo, M. R., Kerachian, R., Alizadeh, M. R., (2018), A fuzzy KNN-based model for significant wave height prediction in large lakes, Journal of Oceanologia, Vol. 60, No. 2, pp. 153-168.

10- Krishna Kumar, N., Savitha, R., Al Mamun, A., (2018), Ocean wave height prediction using ensemble of Extreme Learning Machine, Journal of Neurocomputing, Vol. 277, pp. 12-20.

11- Jain, P., and Deo M.C., (2006), *Neural networks in ocean engineering*, International Journal of Ships and Offshore Structures, Vol. 1, pp. 25–35.

12- Dixit, P., and Londhe, S. (2016), *Prediction of extreme wave heights using neuro wavelet technique*, Journal of Applied Ocean Research, Vol. 58, pp. 241-252.

13- Krishna Kumar, N., Savitha, R., Al Mamun, A., (2017), Regional ocean wave height prediction using sequential learning neural networks, Journal of Ocean Engineering, Vol. 129, pp. 605-612.

14- Shukla, R. P., Kinter, J. L., Shin, C., S. (2018), Sub-seasonal prediction of significant wave heights over the Western Pacific and Indian Oceans, part II: The impact of ENSO and MJO, Journal of Ocean Modelling, Vol. 123, pp. 1-15

15- 10- Krishna Kumar, N., Savitha, R., Al Mamun, A., (2018), *Ocean wave characteristics prediction and its load estimation on marine structures: A transfer learning approach*, Journal Marine Structures, Vol. 61, pp. 202-219.

16- Breuning, M., Kriegel, H-P., Ng, R., and Sander, J., (2000), *LOF: Identifying Density Based Local Outliers*, In Proceeding of ACM SIGMOD International Conference on Management of Data (SIGMOD'00), Dallas, Texas, pp. 93-104.

17- Vaghefi, M., Mahmoodi, K., and Akbari, M (2018), *A Comparison among Data Mining Algorithms for Outlier Detection using Flow Pattern Experiments*. Scientia Iranica. doi: 10.24200/sci.2017.4182

18- Vaghefi, M., Mahmoodi, K., and Akbari, M. (2019). *Detection of Outlier in 3D Flow Velocity Collection in an Open-Channel Bend Using Various Data Mining Techniques*. Iranian Journal of Science and Technology, Transactions of Civil Engineering, Vol. 43, pp. 197–214. doi: 10.1007/s40996-018-0131-2

19- Schmidhuber, J. (2015), *Deep Learning in Neural Networks: An Overview*, Neural Networks, Vol. 61, pp. 85–117. arXiv:1404.7828.

20- http://www.ndbc.noaa.gov

21- Rasheed, S., and Sasikumar, K. (2015), *Modelling Vertical Infiltration in an Unsaturated Porous Media Using Neural Network Architecture*, Aquatic Procedia, Vol. 4, pp. 1008-1015.

[Downloaded from marine-eng.ir on 2026-02-07]
[DOR: 20.1001.1.17357608.1398.15.30.2.2]
[DOI: 10.29252/marineeng.15.30.23]