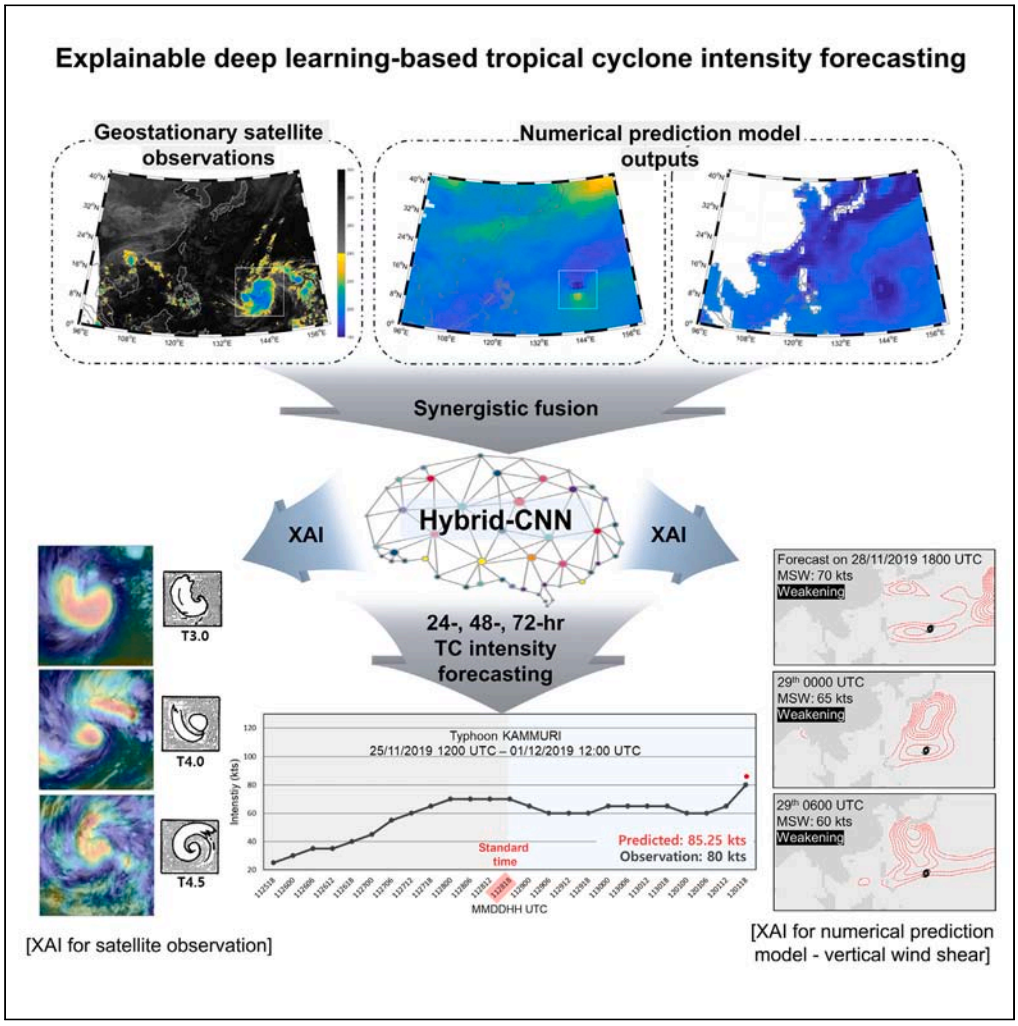
iScience

CellPress
OPEN ACCESS

Article

Enhancing tropical cyclone intensity forecasting with explainable deep learning integrating satellite observations and numerical model outputs



Juhyun Lee, Jungho Im, Yeji Shin

ersgis@unist.ac.kr

Highlights

Proposed hybrid-CNN-based TC intensity forecast models with lead times of 1–3 days

Hybrid-CNN showed significant improvements of 7–110% over operational forecasts

For rapid intensification, our model improved the performance by 50–87%

Suggested potential of hybrid-CNN in understanding TC changes using XAI

Lee et al., iScience 27, 109905
June 21, 2024
https://doi.org/10.1016/j.isci.2024.109905

iScience

CellPress
OPEN ACCESS

Article

Enhancing tropical cyclone intensity forecasting with explainable deep learning integrating satellite observations and numerical model outputs

Juhyun Lee,[1] Jungho Im,[1,2,3,5,*] and Yeji Shin[1,4]

SUMMARY

Tropical cyclone (TC) intensity change forecasting remains challenging due to the lack of understanding of the interactions between TC changes and environmental parameters, and the high uncertainties resulting from climate change. This study proposed hybrid convolutional neural networks (hybrid-CNN), which effectively combined satellite-based spatial characteristics and numerical prediction model outputs, to forecast TC intensity with lead times of 24, 48, and 72 h. The models were validated against best track data by TC category and phase and compared with the Korea Meteorological Administrator (KMA)-based TC forecasts. The hybrid-CNN-based forecasts outperformed KMA-based forecasts, exhibiting up to 22%, 110%, and 7% improvement in skill scores for the 24-, 48-, and 72-h forecasts, respectively. For rapid intensification cases, the models exhibited improvements of 62%, 87%, and 50% over KMA-based forecasts for the three lead times. Moreover, explainable deep learning demonstrated hybrid-CNN's potential in predicting TC intensity and contributing to the TC forecasting field.

INTRODUCTION

Ongoing climate change increases the unpredictability of tropical cyclones (TCs), leading to increased damage.[1–5] Accurate TC forecasts can help mitigate and prepare for damage. Although advances in numerical models have improved TC forecasting, predicting TC intensity remains challenging. This challenge is mainly because of the lack of clarity regarding TC development mechanisms.[6–9] The horizontal structure and inner core of a TC are significantly related to its intensity.[10–14] Consequently, data fusion and assimilation using TC observations play crucial roles in enhancing TC intensity forecasts. Several objective and subjective data fusion techniques have been used to forecast TC intensity. Regional meteorological centers issue TC forecasts operationally using both approaches: subjective data fusion, which relies on forecasters' knowledge and experience, and the objective approach, which often performs worse than the former. For example, official guidance-based TC intensity forecasts improved by 35%, 29%, and 25% compared with ensemble model forecasts.[15] However, excluding subjectivity is difficult, leading to discrepancies in TC forecasts among typhoon centers; thus, objective data fusion must be improved.

Studies have investigated TC intensity forecasting based on the data fusion of statistical-numerical models and TC observations. For example, Velden and Lewis (2017)[16] used two geostationary satellite-based atmospheric motion vectors and the Hurricane Weather Research and Forecasting model, Honda et al. (2018)[17] used Himawari-8-based water vapor (WV) observations and a statistical TC forecasting system, and Yin et al. (2021)[18] used FY-4A based IR observations and the Global/Regional Assimilation and Prediction System. These studies have demonstrated that satellite-based TC observations can improve intensity forecasts. Because objective satellite data integration approaches have used pixel-based brightness temperatures observed by satellite sensors (e.g., average and minimum brightness temperatures near a TC center), they cannot fully simulate the structural characteristics of TCs directly related to their intensity. Thus, imitating forecasters' use of intensity-wise structural characteristics when using satellite observations is necessary.

Research has employed deep learning to characterize TC structure and identify the relationship between TC structure and intensity. Pradhan et al. (2017)[19] proposed an automatic TC intensity estimation model through deep learning using single infrared-window channel data observed by a geostationary satellite sensor. Lee et al. (2020)[20] proposed a multi-dimensional convolutional neural network (CNN)-based TC intensity estimation model, demonstrating that multi-infrared observations in a deep learning framework can improve the accuracy of intensity estimation. Using a deep learning visualization approach, they confirmed the presence of intensity-specific structural characteristics in satellite-based TC observations, particularly in the vicinity of the inner core and the outer rainband. In addition, each infrared channel played

[1]Department of Civil, Urban, Earth, and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea
[2]Graduate School of Carbon Neutrality, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea
[3]Artificial Intelligence Graduate School, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea
[4]Market Intelligence Team, Purchasing Strategy Unit, CJ CheilJedang Corporation, Seoul, South Korea
[5]Lead contact
*Correspondence: ersgis@unist.ac.kr
https://doi.org/10.1016/j.isci.2024.109905

Table 1. Evaluation results of hybrid-CNN, COMS only-CNN, CFSv2 only-CNN, and Intensity only-LSTM

	Hybrid-CNN		COMS-CNN		CFSv2-CNN		Intensity-LSTM	
Forecasting time	R	MAE	R	MAE	R	MAE	R	MAE
24 h	**0.79**	**15.1**	0.65	18.1	0.19	28.7	0.05	34.6
48 h	**0.72**	**20.1**	0.30	25.6	0.13	28.2	0.08	36.3
72 h	**0.47**	**23.1**	0.31	23.1	−0.02	26.7	0.02	30.1

Forecasting time-wise evaluation results (i.e., 24-, 48-, and 72-h). The best validation accuracies are bolded. The unit of MAE is kts, and R is unitless.

a unique role in determining the TC intensity. They revealed that the CNN successively simulated the relationship between satellite-based TC observations and intensity, providing its potential application in the field of intensity forecasts.

This research aims to propose a new framework for deep learning–based TC intensity forecasts with 24-, 48-, and 72-h lead times through the synergistic fusion of numerical model-based predictions and satellite observations. To effectively integrate the two data sources, this study proposed a hybrid-CNN approach. This approach consists of multi-dimensional CNNs whose model parameters are shared. The model could simulate both the structural characteristics of TCs and the interactions between the environmental predictors and their nonlinearity. The proposed approach makes three contributions to the literature: (1) it provides effective synergistic fusion of satellite observations and numerical model output using the hybrid deep learning model; (2) it produces forecasts with multiple lead times (i.e., from 24- to 72-h forecasting) through the deep learning–based fusion of numerical model data and satellite observations; and (3) its deep learning visualization method suggests that the structural features of TCs affect TC intensity forecasts.

RESULTS

Comparison of hybrid-CNN model and single data-based models

Deep learning–based 24-, 48-, and 72-h TC intensity forecasting was conducted through synergistic fusion of satellite observations (i.e., Communicate, Ocean and Meteorological Satellite, COMS) and numerical model outputs (Climate Forecasting System version 2, CFSv2) (see STAR methods part). Prior to evaluating the overall performance of the proposed model, the fusion model (i.e., hybrid-CNN) and single data-based deep learning models (i.e., COMS-CNN, CFSv2-CNN, and Intensity-long short-term memories (LSTM)) were compared to represent the impact of deep learning–based data fusion. Table 1 summarizes the 24-, 48-, 72-h TC intensity forecasting performance of the models (i.e., hybrid-CNN, COMS-CNN, CFSv2-CNN, and Intensity-LSTM). For all forecasting times, the hybrid-CNN models performed much better than the single data-based CNN models, resulting in the lowest mean absolute error (MAE) and highest Pearson's correlation (R). The hybrid-CNN models improved by MAE of 16.57% and 21.48% compared with COMS-CNN, improved by MAE of 47.36% and 48.72% compared with CFSv2-CNN, and improved by MAE of 56.36% and 44.62% compared with Intensity-LSTM in 24- and 48-h forecasting, respectively. In the 72-h forecasting, hybrid-CNN did not significantly outperform COMS-CNN in terms of MAE; however, the prediction correlation improved by 51%. These results revealed that deep learning–based data fusion significantly improved TC intensity forecasting, consistent with results in the literature on data assimilation.[21,22] These findings indicate that data fusion of multi-source datasets in a deep learning framework could synergistically improve TC intensity forecasting with lead times of 24–72 h.

Overall evaluation of hybrid-CNN models

Tables 2, 3, and 4 shows the 24-, 48-, and 72-h forecasting results of KMA (i.e., K24, K48, and K72) and the proposed hybrid-CNN models (i.e., H24, H48, and H72). H24 yielded better performance than K24 did for all categories except for category 1: SS values of 0.22, 0.14, 0.16, and 0.11 for categories 2–5, respectively. Although H24 performed worse for weak TCs with an intensity <48 kts, resulting in an SS of −0.34 in terms of

Table 2. Category-wise 24-h forecasting performance

	K24			H24			
Target category	MAE	R	MAPE	MAE	R	MAPE	Skill score
Category 1	7.5	0.1	54.2	10.1	0.2	43.0	−0.34
Category 2	13.8	0.3	45.4	10.8	0.4	19.7	0.22
Category 3	17.0	−0.1	26.6	14.7	0.2	20.8	0.14
Category 4	17.7	0.3	16.7	14.9	0.3	16.5	0.16
Category 5	19.6	0.3	15.5	17.4	0.3	14.2	0.11

Category-wise 24-h forecasting performance of KMA (K24) and hybrid-CNN (H24) using 27 typhoons in 2019. Skill score was calculated using the MAE of each model. The unit of MAE is kts, MAPE is %, and R and Skill score are unitless.

Table 3. Category-wise 48-h forecasting performance

	K48			H48			
Target category	MAE	R	MAPE	MAE	R	MAPE	Skill score
Category 1	9.9	0.27	21.8	15.8	−0.13	58.1	−0.60
Category 2	13.1	0.38	24.2	13.4	0.34	25.0	−0.02
Category 3	21.3	−0.15	29.9	10.1	0.07	14.2	1.11
Category 4	18.8	0.40	21.0	14.9	0.30	16.4	0.21
Category 5	29.8	0.25	25.3	28.2	0.11	23.8	0.05

Category-wise 48-h forecasting performance of KMA (K48) and hybrid-CNN (H48) using 27 typhoons in 2019. Skill score was calculated using the MAE of each model. The unit of MAE is kts, MAPE is %, and R and Skill score are unitless.

mean absolute error (MAE), it yielded better correlation values than K24 did. H24 performed well, resulting in a mean absolute percentage error (MAPE) < 20 for categories 2, 4, and 5; it also showed reasonable performance for categories 1 and 3. As the target TC intensity increased, the performance of K24 degraded rapidly: an MAE of 7.5 kts was observed for category 1, and the MAE doubled for category 3 and nearly tripled for category 5 (MAE of 17.0 kts and 19.6 kts for categories 3 and 5, respectively). In contrast, H24 demonstrated relatively stable performance by intensity, with an MAE of 17.4 kts for category 5 and 10.1 kts for category 1. For 48-h forecasts, H48 demonstrated a significantly superior performance, especially in strong TCs of category 3 or higher. Both models showed that the prediction error increased as the intensity of the target TC increased (K48 had MAEs of 21.3, 18.8, and 29.8 kts, and H48 had MAEs of 10.1, 14.9, and 28.2 kts for categories 3–5, respectively), H48 performed 5%–21% better than K48 did. H48 was a good, reasonable intensity forecasting model, with MAPEs of 14.2, 16.4, and 23.8% in categories 3 to 5, and K48 had MAPEs of 29.9, 21.0, and 25.3% in the same categories. For 72-h forecasting results, H72 showed a reasonable performance with MAPEs of 28.4, 28.5, 31.1, and 24.1 in categories 2–5, respectively; however, it had 8%–32% lower performance in categories 2–4 than K72 did (H72 had MAEs of 15.1, 20.4, and 18.1 kts, and K72 had MAEs of 13.9, 19.6, and 21.3 kts in categories 2–4, respectively). For category 5 TCs, H72 demonstrated a 7% performance improvement over K72. However, the proposed forecasting models (i.e., H24, H48, and H72) showed relatively low performance for weak TCs (category 1). Several reasons were inferred: 1) due to the tangled pattern of weak TCs, simulating the intensity-wise TC pattern using a CNN-based model was difficult,[20] and 2) due to the environmental perturbations that occur when TCs dissipate, predicting weak TC cases with a high probability of dissipating was especially difficult.[23,24]

Figure 1 depicts the scatterplots of the performance of all forecasting models with three lead times according to the TC phase (i.e., dissipating, sustained, and developing phases, which are the real intensity changes over the past 6 h based on targe TC intensity observations). Corresponding to the categorical performances shown in Table 1, H24 outperformed K24: K24 exhibited an R of 0.68 and a slope of 0.46, and H24 yielded an R of 0.79 and a slope of 0.60 (Figures 1A and 1B). Interestingly, K24 demonstrated phase-biased prediction results: it underestimated the intensity for developing cases and overestimated it for dissipating cases. By contrast, H24 demonstrated robust prediction results regardless of the phase. K48 resulted in an R of 0.67 and a slope of 0.51; H48 exhibited an R of 0.72 and a slope of 0.38 (Figures 1C and 1D). Slope did not significantly improve, but R improved by 5% in H48 compared to K48. For overall accuracy, K72 showed better performance than H72 did (Figures 1E and 1F). However, for strong TCs exceeding 85 kts (i.e., categories 4 and 5), H72 outperformed K72, resulting in an R of 0.64 and a slope of 0.91, and K72 had an R of 0.36 and a slope of 0.47 (Figure 2). As shown in Table 4, H72 performed poorly in forecasting the intensity of weak TCs, particularly dissipating weak TCs.

Figure 3 depicts boxplots showing the variability of two numerical forecasting factors (i.e., Pressure at mean sea level (PRES) and Temperature at low altitude (TMPL)) for all dissipating weak TC cases used for validation. H72 generally exhibited good forecasting performance when the environmental prediction factors significantly fluctuated. By contrast, when the pressure at mean sea level and temperature at

Table 4. Category-wise 72-h forecasting performance

	K72			H72			
Target category	MAE	R	MAPE	MAE	R	MAPE	Skill score
Category 1	10.8	0.45	36.7	20.5	0.20	87.8	−0.89
Category 2	13.9	0.36	26.2	15.1	0.17	28.4	−0.08
Category 3	19.6	0.00	27.5	20.4	−0.12	28.5	−0.04
Category 4	21.3	0.53	23.8	28.1	0.33	31.0	−0.32
Category 5	30.5	−0.28	25.7	28.1	0.25	24.1	0.07

Category-wise 72-h forecasting performance of KMA (K72) and hybrid-CNN (H72) using 27 typhoons in 2019. Skill score was calculated using the MAE of each model. The unit of MAE is kts, MAPE is %, and R and Skill score are unitless.

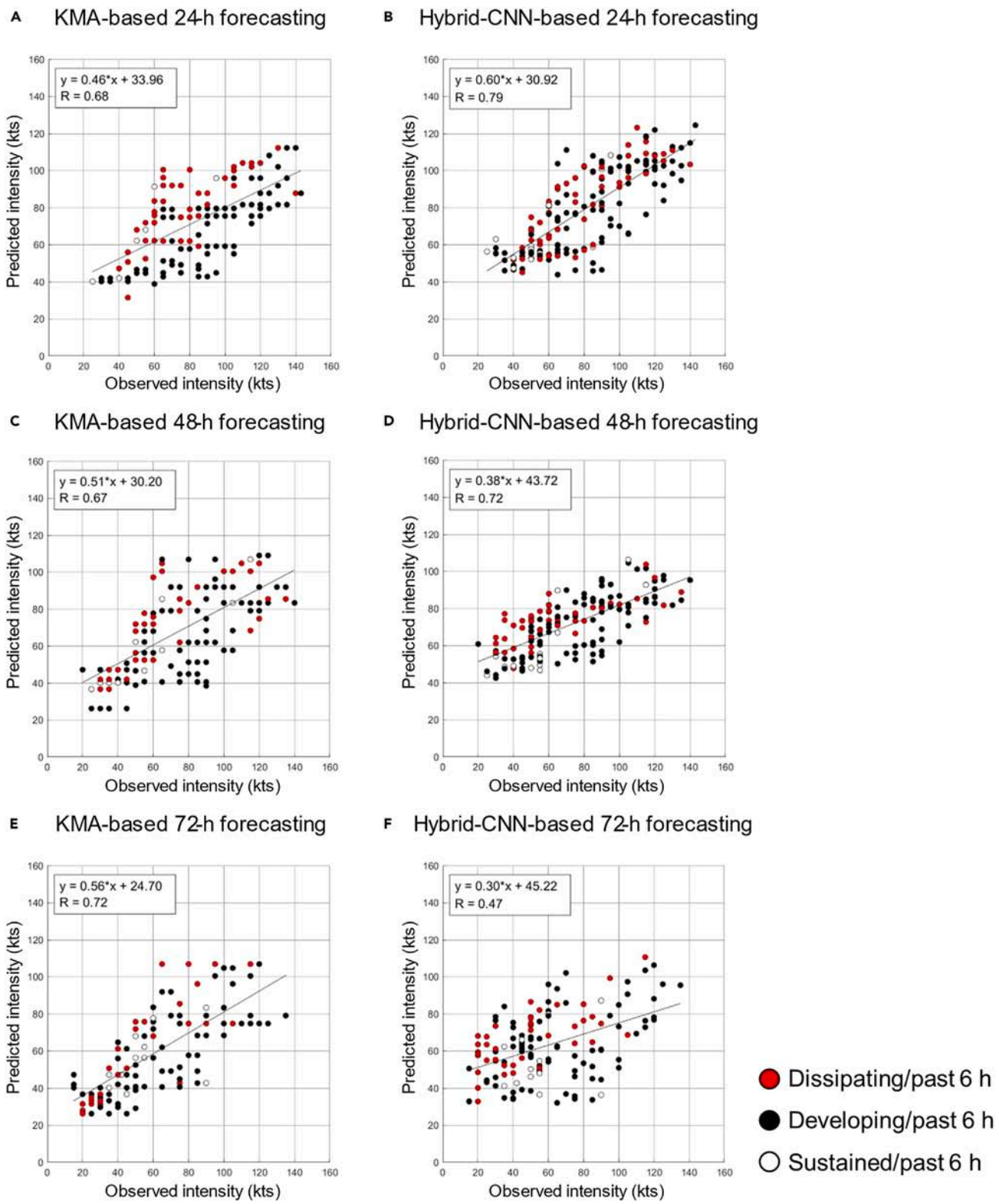


Figure 1. Scatterplots of KMA- and hybrid-CNN-based TC intensity forecasting results

(A) KMA-based forecasting results for lead time for 24 h.

(B) Hybrid-CNN-based forecasting results for lead time 24 h.

Figure 1. *Continued*

(C) KMA-based forecasting results for lead time 48 h.

(D) Hybrid-CNN-based forecasting results for lead time 48 h.

(E) KMA-based forecasting results for lead time 72 h.

(F) Hybrid-CNN-based forecasting results for lead time 72 h. The x axis indicates observed intensity (kts) reported from JTWC best track, and the y axis indicates predicted intensity (kts). The TC activity phase in the past 6 h of validation samples is marked with colors. Red, black, and white-colored scatters indicate dissipating, developing, and sustained cases, respectively.

low altitude forecasts were stable, as in cases 11–22, H72 had high forecasting errors. An exception was that cases with increased atmospheric temperatures at low altitudes (i.e., cases 1–3) were accurately predicted as weak TCs by H72, although the predicted atmospheric temperatures were stable. Because the number of weak TCs was greater than that of strong TCs, the biased forecasting errors by the target TC intensities led to a decline in the overall prediction performance of H72.

However, it should be noted that the joint typhoon warning center (JTWC) best track data used in this study require post-processing time, making it unsuitable to directly use them for operational forecasting. To overcome this limitation, the best track data might be replaced with operational reports. To verify the replaceability, the 24-h, 48-h, and 72-h intensity forecasting models using KMA-based operational real-time analysis were evaluated, resulting in skill score differences of 1.8%, 0.7%, and 0.7%, respectively, when compared to the models with JTWC best track. It implies that the proposed models can be operationally used based on the operational reports without a significant drop in performance.

Evaluation for rapid intensification (RI) of hybrid-CNN models

Although TC intensity prediction has significantly improved, accurate forecasting of RI remains challenging.[7,25,26] Because of the obscurity of TC development mechanisms, such as the interaction between environmental (atmospheric and oceanic) factors and the inter physics of TCs, whether rapid intensification (RI) occurrence is typical or exceptional remains unclear.[27] Thus, the performance of RI prediction by forecast time was evaluated using all typhoons that occurred in 2019. There were 16, 8, and 4 RI cases for 24-, 48-, and 72-h forecasts, respectively.

Figure 4 shows the prediction performances of the TC intensity change with three lead times over the previous 24 h for each RI case. Some cases were predicted to have a negative intensity gradient by both forecasting models, and H24 performed significantly better than K24 in nine cases (Figure 4A). K24 predicted cases 11–14 as a weakening or in a sustained phase, and H24 predicted them as intensifying cases. For cases 11 through 16, which traversed the Philippines, H24 exhibited greater improvement than K24 did. The TCs weakened as they made landfall in the Philippines but quickly intensified as they left land. This finding implies that H24 may mitigate the unpredictability of K24 when TCs are close to land. H48 performed better than K48 in all RI cases (Figure 4B). For cases 2, 4, and 8, K48 predicted the RI cases as a weakening or sustained phase, and H48 predicted them in an intensifying phase. In particular, for cases 6 and 7, which intensified by more than 50 kts, K48 could not simulate the intensity change, whereas H48 predicted it in an intensifying phase. The RI prediction performance of the proposed model was also noteworthy for 72-h forecasts (Figure 4C). For two of the four cases (i.e., RI cases 3 and 4), H72 demonstrated better performance than K72 did. RI cases 1 and 2, in the South China Sea and surrounded by land, were difficult to predict with both models, and cases 3 and 4, near the Philippines, performed significantly better with H72 than with K72.

Although hybrid-CNN-based models did not demonstrate high RI detection accuracy for over 30 kts/day, in terms of probability of detection (POD) rate, H24 achieved an 8% improvement compared to K24. While there was no RI forecasting hit in KMA-based forecasts, three cases were

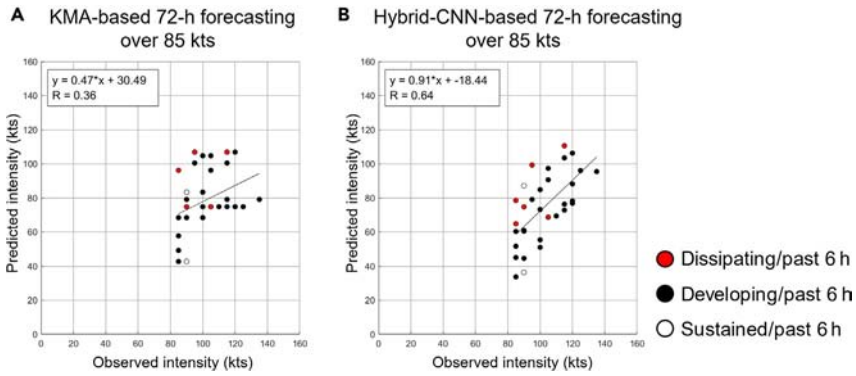


Figure 2. Scatterplots of KMA- and hybrid-CNN-based forecasting results when targeting TCs over 85 kts

(A) KMA-based 72 h forecasting results.

(B) Hybrid-CNN-based 72 h forecasting results. The x axis indicates the observed intensity (kts) reported from JTWC best track, and the y axis indicates the predicted intensity (kts). The TC activity phase in the past 6 h of validation samples is marked with colors. Red, black, and white-colored scatters indicate dissipating, developing, and sustained cases, respectively.

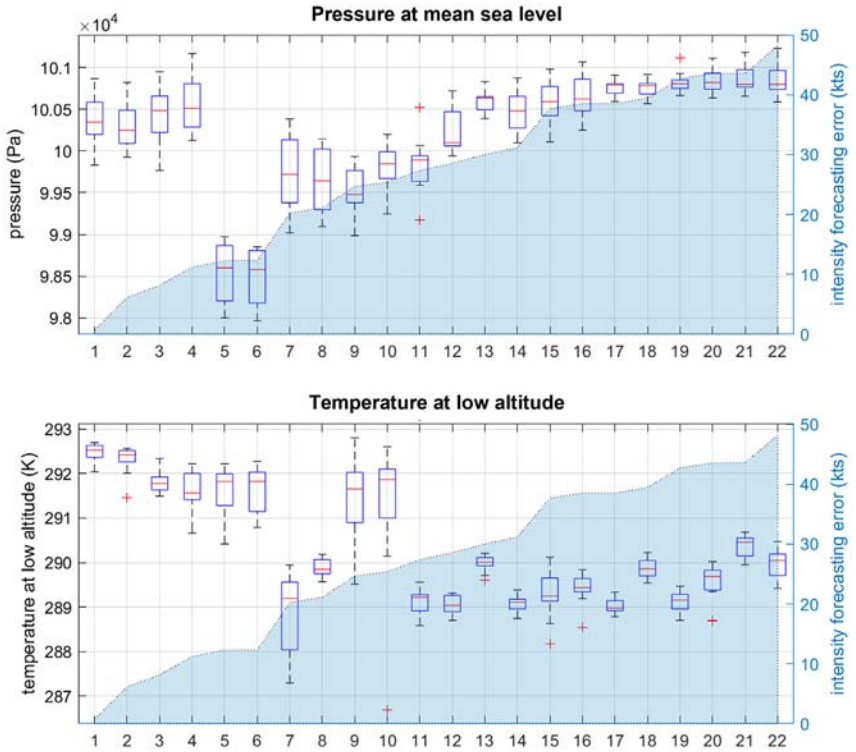


Figure 3. Distribution of CFSv2-based two atmospheric forecasts (i.e., pressure at mean sea level and temperature at low altitude) for dissipating TCs with category 1

The x axis indicates TC cases in ascending order of intensity forecasting error. The main y axis indicates the pressure (Pa) for top and temperature (K) for bottom, and the sub y axis indicates the intensity forecasting error.

detected by H24. As there are generally underpredicted trends in H48 and H72, it seems difficult to predict RI. On the contrary, the hybrid-CNN outperformed the KMA-based forecasts with regard to the gradient of intensification. Specifically, the KMA-based models failed to detect significant gradients for all forecasting lead times. For all forecasting lead times, the hybrid-CNN demonstrated a better performance than the KMA-based forecasts did, whereas the KMA-based models could not identify large gradients. In particular, the proposed model performed well even when RI occurred before or after landfall, one of the experimental limitations of the approaches in the literature.[12,28,29]

Forecasting performance for typhoon KAMMURI in 2019

Figure 5 shows the successive TC intensity forecasting results using Typhoon KAMMURI in 2019. Similar to the RI evaluation (Figure 4), the TC intensity prediction performance of the KMA models decreased during periods of RI (02/12 06:00, 12:00 and 18:00 UTC in 2019). However, the proposed model performed well during the intensifying and dissipating phases. When the TC was heading southwest of the Philippines and rapidly intensified by an increase of 55 kts from 01/12/2019 06:00 UTC to 02/12/2019 12:00 UTC, the proposed model predicted an increase of 30.7 kts in 24-h forecasting and 21.0 kts in 48-h forecasting, and the KMA models predicted them in a sustained phase. In the 72-h forecasting (Figure 5C), the intensified period of KAMMURI (i.e., from 01/12/2019 18:00 UTC to 02/12/2019 12:00 UTC) was predicted to weaken with a decrease of 5.8 kts in K72, and H72 predicted it would intensify with an increase of 21.1 kts. KMA-based models exhibited better performance in the rapid weakening phase than in the RI phase for Typhoon KAMMURI (2019). By contrast, the hybrid-CNN effectively simulated both the timing and gradient of rapid changes (i.e., intensifying and weakening).

According to the literature, simulations using the horizontal structure of TCs play a crucial role in forecasting intensity changes.[30,31] Thus, a hybrid-CNN-based objective analysis of satellite-based observations may help improve TC intensity change forecasts. To determine how well the hybrid-CNN learned satellite-based observations, this study used the heatmap approach, a CNN-based visualization method. Figure 6 depicts the visualization results based on the proposed 24-, 48-, and 72-h forecasting models (on 12/01/2019 00 UTC, 12/02/2019 00 UTC, and 12/03/2019 00 UTC, respectively) and the TC development pattern, which corresponds to the JTWC best track-based intensity. Each heatmap extracted from all the forecasting models mimicked the Dvorak technique. Specifically, the models highlighted the successive curvature of the central dense overcast (CDO) pattern around the TC centers. Because the structure of a TC cloud system represents the inner dynamics of each system, the effective analysis of successive TC observations using convolutional layers proposed in this study contributes to the forecasting of TC intensity. Although each heatmap followed a typical TC development pattern, maps varied by forecasting time. As the forecasting time increased, the concentration of high heatmap values increased in the CDO region, and their area decreased. This finding implies

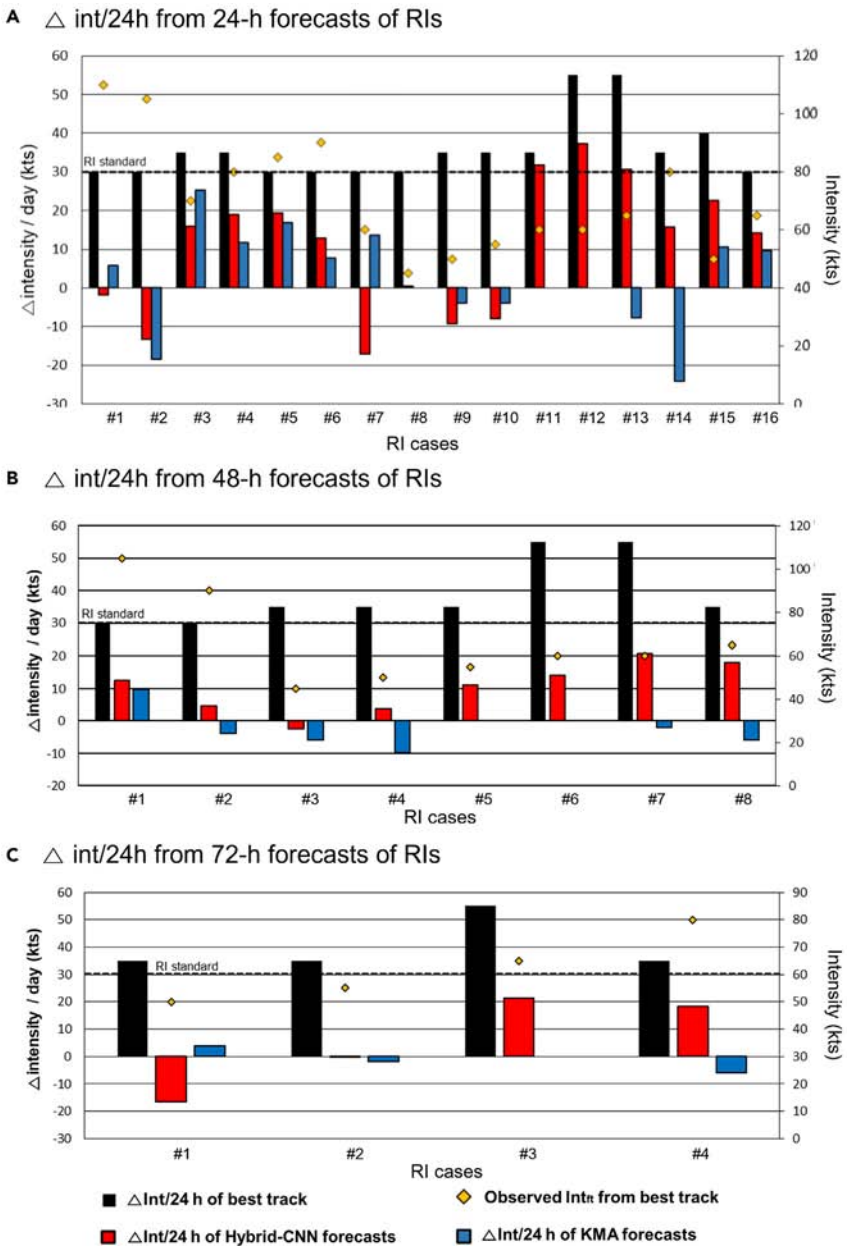


Figure 4. Gradients of intensities (kts) predicted by KMA and hybrid-CNN models when TCs were rapidly intensified (i.e., intensity gradients >30 kts in the past 24 h)

(A) The intensity gradients for the lead time of 24 h.

(B) The intensity gradients for the lead time of 48 h.

(C) The intensity gradients for the lead time of 72 h. Black, red, and blue bars indicate intensity gradients (kts, main y axis, left side of the graph) of observations from the JTWC best track and forecasting by the hybrid-CNN and KMA, respectively. The yellow diamond indicates the observed intensity (kts, secondary y axis, right side of the graph) of the target TC.

that as the lead time increased, the concentration area of the proposed models (i.e., high-value pixels in the heatmap) decreased, and the focus on a significant region over the observations was maintained.

While satellite-observed intensity-related structural characteristics of TCs were incorporated into the model, numerical model-derived environmental factors were also considered simultaneously to forecast future intensity in the hybrid-CNN. To examine the impact of environmental factors, the heatmap of vertical wind shear (VWS) was used. VWS is a well-known factor associated with the development of TCs.[32–35] VWS contributes to the weakening of TCs by decreasing vertical convective activity and forming an asymmetric inner core that makes it

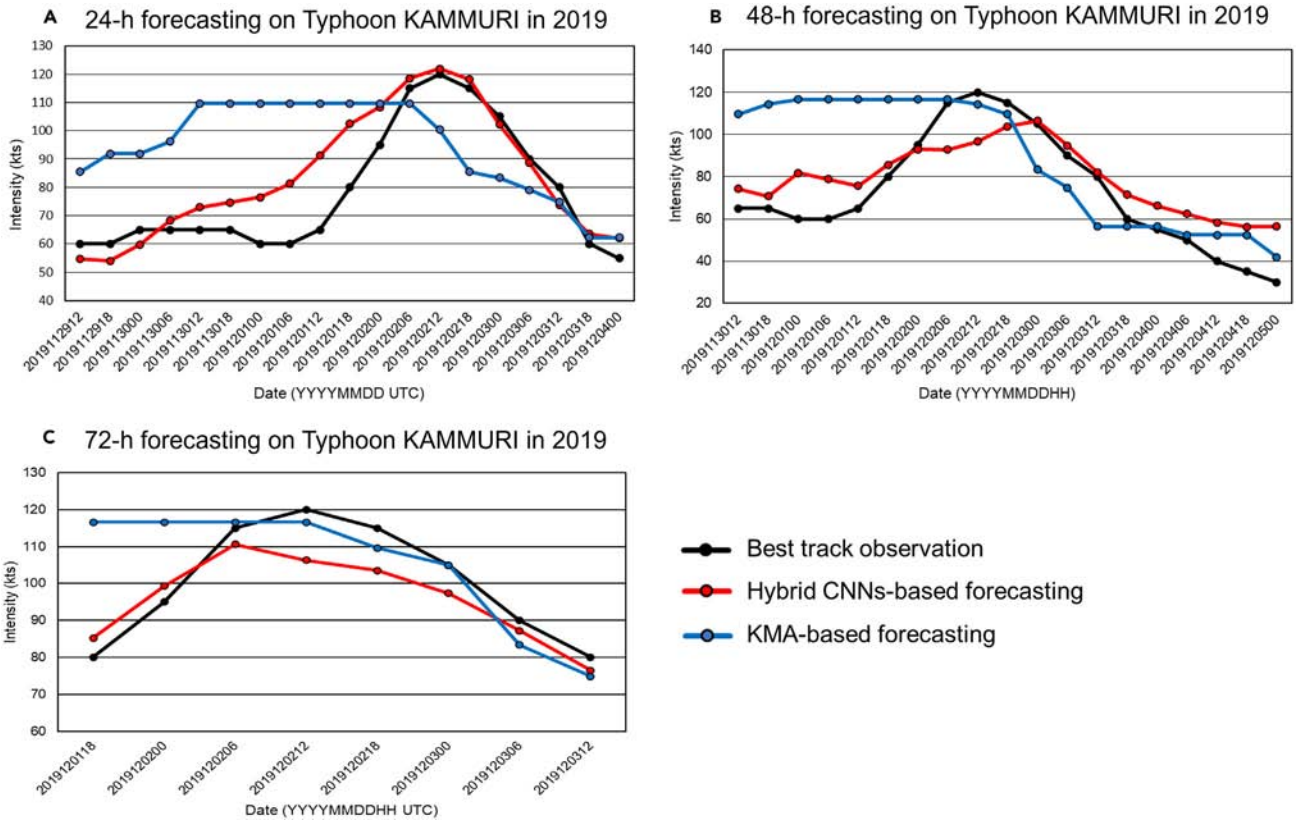


Figure 5. Intensity forecasting results by hybrid-CNN and KMA-based models for Typhoon KAMMURI in 2019

(A) Typhoon KAMMURI (2019) forecasting result for the lead times of 24 h.

(B) Typhoon KAMMURI (2019) forecasting result for the lead times of 48 h.

(C) Typhoon KAMMURI (2019) forecasting result for the lead times of 72 h..

difficult for the eyewall to retain heat and moisture.[36–39] Xu et al. (2013)[40] also revealed that the horizontal flow of environmental vertical shear can influence TC weakening. Figure 7 depicts heatmaps of VWS extracted from the hybrid-CNN-based 72-h forecasting for typhoon KAMMURI at 1800 UTC on 01/12/2019. During the TC weakening phase (Figures 7A–7C and 7I–7J), the heatmap of VWS was mostly emphasized over the TC center, forming a relatively zonal flow (east–west), which contributes to the asymmetric flows surrounding the TC. On the other hand, during the intensification (Figures 7F–7H and 7L–7M), the highlighted heatmap of VWS was away from the TC center. It implies that the proposed hybrid-CNN-based intensity forecasting model utilized the environmental factors appropriately to predict intensity.

DISCUSSION

In this study, an advanced deep learning–based data fusion model (hybrid-CNN) was proposed to improve objective TC intensity forecasting. This attempt was the first to fuse multiple data types into a deep learning framework for TC intensity forecasting. The existing objective data fusion techniques using satellite observations are limited to short-term prediction (up to 24 h), but the proposed hybrid-CNN simulates TC intensity with lead times of 24, 48, and 72 h. Overall, the hybrid-CNN models outperformed the KMA-based forecasting reports for the 24-h and 48-h TC intensity forecasts. However, in the 72-h forecasts, H72 exhibited a better performance with an SS of 7% for category 5, and the general performance of H72 was slightly less than K72 (Table 4).

When evaluating the performance at RIs, the hybrid-CNN models outperformed KMA-based forecasts. In particular, when RI was near the Philippines, the proposed models accurately predicted TC intensity in the intensifying phase, and the operational system mis-predicted the intensity in a sustained or decaying phase. For Typhoon KAMMURI in 2019, the proposed models were well predicted for both the RI and rapid weakening phases, whereas the operational forecasts were underestimated for RI. The structural characteristics of TCs from satellite observations and numerical model outputs substantially improved TC intensity forecasting, confirmed using a heatmap approach. This result implies that the hybrid-CNN model reasonably simulated the Dvorak technique, which has been used by forecasters for subjective analysis, and intelligently adopted the dynamics of environmental factors. It may play an important role in improving TC intensity forecasts operationally.

This research demonstrated the promising results of the proposed hybrid-CNN for TC intensity forecasts, but further improvements are necessary:1) to improve the training of the mechanisms of TC intensification or dissipation, hybrid-CNN models by phase can be developed

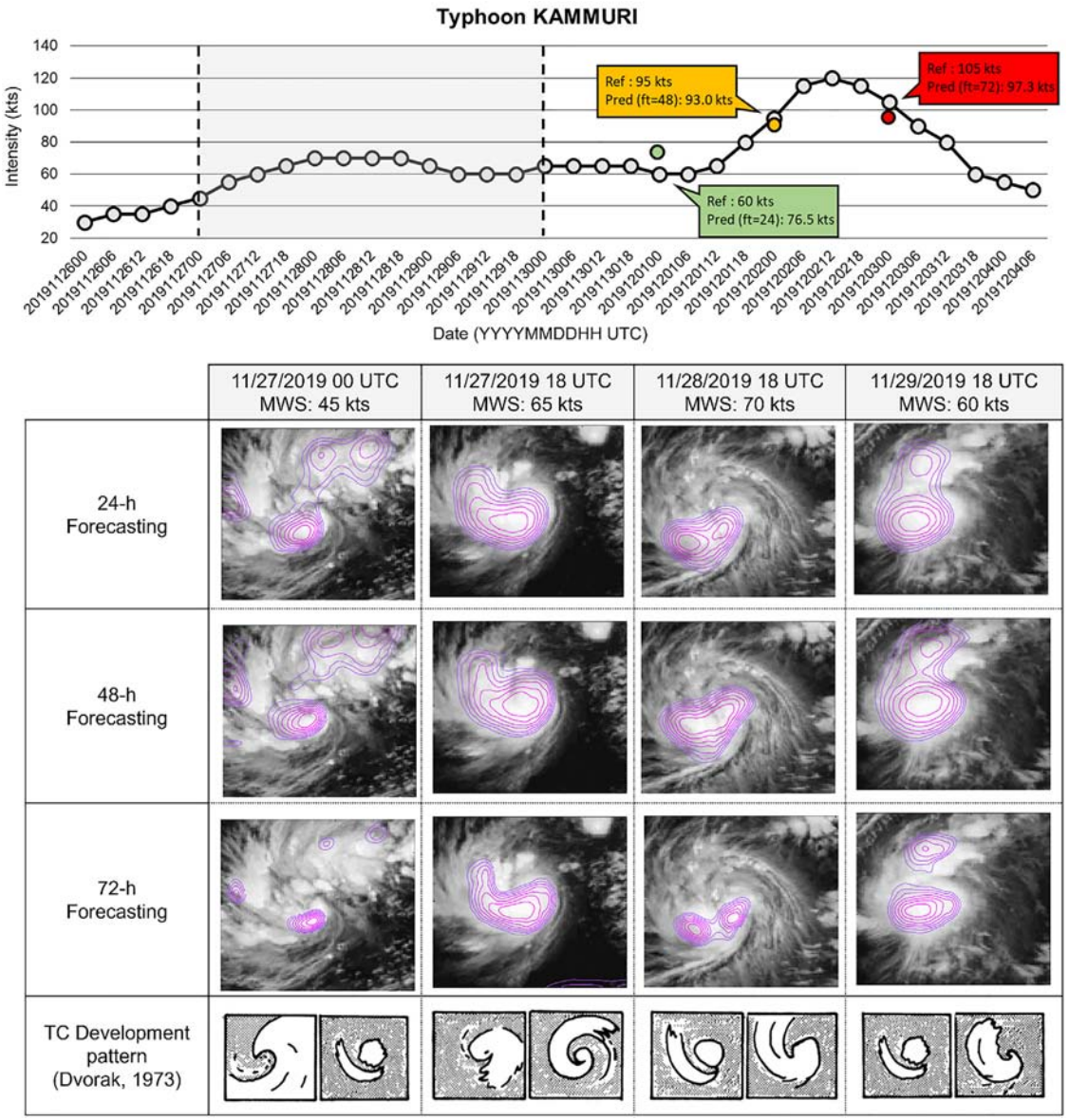


Figure 6. Model-wise CNN-based visualization results for satellite observations

Gray dots in the upper graph indicate intensity from JTWC best track. Green, yellow, and red dots show forecasted intensities at 12/01/2019 00 UTC, 12/02/2019 00 UTC, and 12/03/2019 00 UTC by the hybrid-CNN-based models with lead times 24, 48, and 72 h, respectively. 'Ref', 'Pred', and 'ft' in the graph mean the reference intensity from JTWC best track, predicted intensity by the proposed model, and lead time, respectively. The gray background image in the bottom is COMS IR1-based observations, and pink contours are drawn with the upper 50% value of each heatmap. The last row depicts the TC development pattern proposed by Dvorak (1973), of which intensity corresponds to the best track-based intensity.

and 2) multi-numerical models or an ensemble model can be adopted in the hybrid-CNN architecture to mitigate the dependency of the proposed approach on the systematic error of a single numerical model.

Limitations of the study

As deep learning–based models require a lot of data for various cases, the previous generation geostationary satellite (i.e., COMS) launched in 2010 was utilized for this study. However, COMS was decommissioned in 2020, and the next generation geostationary satellite, named GeoKompsat-2A (GK2A), has been launched in 2019. Given the limited operational period of GK2A, we were not able to use it in the present study. The continuous observations from two satellites may further enhance TC intensity forecasting via transfer learning, which enables the proposed hybrid-CNN-based forecasting models to effectively adapt to new data while retaining their previous knowledge.

Figure 7. Heatmaps of vertical wind shear (red contour) over typhoon KAMMURI (72-h forecasting for 01/12/2019 1800 UTC)

Heatmaps generated by the last convolutional layer of mini-CNNs for large-scale in the hybrid-CNN-based 72-h intensity forecasting model. The TC location is marked with a cyclone icon. (A–C) and (I–J) show the weakening phase, while (F–H) and (L–M) depict the intensifying phase.

The varied temporal resolution of the used datasets was a limitation. As the best track and numerical model-based forecasts used in this study provide 6-hourly information, we conducted hourly interpolation in order to design deep learning–based TC intensity forecasting models. The uncertainty introduced by this preprocessing may persist in the reconstructed datasets. If the best track and environmental variables were provided as hourly data, the proposed model could mitigate the interpolation-induced uncertainties. Furthermore, hourly numerical model-based forecasts with enhanced performance can also contribute to the improvement of hybrid-CNN-based TC forecasts.

The limited computational capacity is another limitation of this study. In order to handle the huge amount of data with high dimensions, it was necessary to reduce computational costs. Consequently, the satellite-based TC observations were upscaled from 4 km to 12 km in this study. While prior research has indicated that a spatial resolution of 12 km is sufficient for simulating TC intensity,[20,41] it is evident that a higher resolution can furnish a greater amount of information pertaining to TC intensity. If there is enough computer power to train the satellite-based observations at their original resolution (4 km for COMS MI IR channel-based observations), it may be possible to get more detailed spatial information about TC intensity. Furthermore, advances in satellite sensor technology enable the collection of imagery with better specifications (i.e., higher spatial, temporal, and spectral resolutions). For instance, GK2A Advanced Meteorological Imager (AMI), which succeeds COMS, has a spatial resolution of 2 km and a temporal resolution of 10 min for longwave infrared and water vapor channel observations, as well as more precise spectral resolution-based channels, which can be used to further enhance the accuracy of TC intensity forecasts.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109905.

ACKNOWLEDGMENTS

This work was funded by Korea Institute of Marine Science & Technology (KIMST) funded by the Ministry of Oceans and Fisheries (RS-2023-00256330, Development of risk managing technology tackling ocean and fisheries crisis around Korean Peninsula by Kuroshio Current) and (20210046, Development of technology using analysis of ocean satellite images), and by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea goverment (Ministry of Science and ICT (MSIT)) (No.2020-0-01336, Artificial Intelligence Graduate School Program, Ulsan National Institute of Science and Technology (UNIST)). Additionally, we sincerely appreciate all reviewers for improving the quality of the manuscript.

AUTHOR CONTRIBUTIONS

J.L.: Conceptualization, Methodology, Analysis, Writing – original draft, Writing – review and editing; J.I.: Conceptualization, Writing – review and editing, Supervision; Y.S.: Writing – review and editing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 14, 2023
Revised: November 18, 2023
Accepted: May 2, 2024
Published: May 3, 2024

REFERENCES

1. Meehl, G.A., Zwiers, F., Evans, J., Knutson, T., Mearns, L., and Whetton, P. (2000). Trends in extreme weather and climate events: issues related to modeling extremes in projections of future climate change. Bull. Am. Meteorol. Soc. *81*, 427–436.
2. Hill, K.A., and Lackmann, G.M. (2011). The impact of future climate change on TC intensity and structure: A downscaling approach. J. Clim. *24*, 4644–4661.
3. Mendelsohn, R., Emanuel, K., Chonabayashi, S., and Bakkensen, L. (2012). The impact of climate change on global tropical cyclone damage. Nat. Clim. Change *2*, 205–209.
4. Mori, N., and Takemi, T. (2016). Impact assessment of coastal hazards due to future changes of tropical cyclones in the North Pacific Ocean. Weather Clim. Extrem. *11*, 53–69.
5. Pandey, R.S., and Liou, Y.A. (2022). Typhoon strength rising in the past four decades. Weather Clim. Extrem. *36*, 100446.
6. Bauer, P., Thorpe, A., and Brunet, G. (2015). The quiet revolution of numerical weather prediction. Nature *525*, 47–55.
7. Courtney, J.B., Langlade, S., Sampson, C.R., Knaff, J.A., Birchard, T., Barlow, S., Kotal, S., Kriat, T., Lee, W., Pasch, R., and Shimada, U. (2019). Operational perspectives on tropical cyclone intensity change part 1: Recent advances in intensity guidance. Trop. Cycl. Res. Rev. *8*, 123–133.
8. Heming, J.T., Prates, F., Bender, M.A., Bowyer, R., Cangialosi, J., Caroff, P., Coleman, T., Doyle, J.D., Dube, A., Faure, G., et al. (2019). Review of recent progress in tropical cyclone track forecasting and expression of uncertainties. Trop. Cycl. Res. Rev. *8*, 181–218.
9. Wang, C.C., Lee, C.Y., Jou, B.J.D., Celebre, C.P., David, S., and Tsuboki, K. (2022). High-resolution time-lagged ensemble prediction for landfall intensity of Super Typhoon Haiyan (2013) using a cloud-resolving model. Weather Clim. Extrem. *37*, 100473.
10. DeMaria, M., Sampson, C.R., Knaff, J.A., and Musgrave, K.D. (2014). Is tropical cyclone intensity guidance improving? Bull. Am. Meteorol. Soc. *95*, 387–398.
11. Yamaguchi, M., Owada, H., Shimada, U., Sawada, M., Iriguchi, T., Musgrave, K.D., and DeMaria, M. (2018). Tropical cyclone intensity prediction in the western North Pacific basin using SHIPS and JMA/GSM. Inside Solaris *14*, 138–143.
12. Knaff, J.A., Sampson, C.R., and Musgrave, K.D. (2018). An operational rapid intensification prediction aid for the western North Pacific. Weather Forecast. *33*, 799–811.
13. Shimada, U., Owada, H., Yamaguchi, M., Iriguchi, T., Sawada, M., Aonashi, K., DeMaria, M., and Musgrave, K.D. (2018). Further improvements to the Statistical Hurricane Intensity Prediction Scheme using tropical cyclone rainfall and structural features. Weather Forecast. *33*, 1587–1603.
14. Peng, C.H., and Wu, C.C. (2020). The impact of outer-core surface heat fluxes on the convective activities and rapid intensification of tropical cyclones. J. Atmos. Sci. *77*, 3907–3927.
15. Chen, G., Zhang, X., Bai, L., and Wan, R. (2019). Verification of tropical cyclone operational forecast in 2018. In Proceedings of the ESCAP/WMO Typhoon Committee, Guangzhou, China, p. 26.
16. Velden, C., Lewis, W.E., Bresky, W., Stettner, D., Daniels, J., and Wanzong, S. (2017). Assimilation of high-resolution satellite-derived atmospheric motion vectors: Impact on HWRF forecasts of tropical cyclone track and intensity. Mon. Weather Rev. *145*, 1107–1125.
17. Honda, T., Miyoshi, T., Lien, G.Y., Nishizawa, S., Yoshida, R., Adachi, S.A., Terasaki, K., Okamoto, K., Tomita, H., and Bessho, K.

(2018). Assimilating all-sky Himawari-8 satellite infrared radiances: A case of Typhoon Soudelor (2015). Mon. Weather Rev. *146*, 213–229.

18. Yin, R., Han, W., Gao, Z., and Li, J. (2021). Impact of high temporal resolution FY-4A Geostationary Interferometric Infrared Sounder (GIIRS) radiance measurements on Typhoon forecasts: Maria (2018) case with GRAPES global 4D-Var assimilation system. Geophys. Res. Lett. *48*, e2021GL093672.
19. Pradhan, R., Aygun, R.S., Maskey, M., Ramachandran, R., and Cecil, D.J. (2018). Tropical cyclone intensity estimation using a deep convolutional neural network. IEEE Trans. Image Process. *27*, 692–702.
20. Lee, J., Im, J., Cha, D.H., Park, H., and Sim, S. (2019). Tropical cyclone intensity estimation using multi-dimensional convolutional neural networks from geostationary satellite data. Rem. Sens. *12*, 108.
21. DeMaria, M., Mainelli, M., Shay, L.K., Knaff, J.A., and Kaplan, J. (2005). Further improvements to the statistical hurricane intensity prediction scheme (SHIPS). Weather Forecast. *20*, 531–543.
22. Jones, T.A., Knopfmeier, K., Wheatley, D., Creager, G., Minnis, P., and Palikonda, R. (2016). Storm-scale data assimilation and ensemble forecasting with the NSSL experimental Warn-on-Forecast system. Part II: Combined radar and satellite data experiments. Weather Forecast. *31*, 297–327.
23. Gopalakrishnan, S.G., Goldenberg, S., Quirino, T., Zhang, X., Marks, F., Yeh, K.S., Atlas, R., and Tallapragada, V. (2012). Toward improving high-resolution numerical hurricane forecasting: Influence of model horizontal grid resolution, initialization, and physics. Weather Forecast. *27*, 647–666.
24. Emanuel, K., and Zhang, F. (2016). On the predictability and error sources of tropical cyclone intensity forecasts. J. Atmos. Sci. *73*, 3739–3747.
25. Mundell, D.B. (1990). Prediction of tropical cyclone rapid intensification events (AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH).
26. Krishnamurti, T.N., Pattnaik, S., Stefanova, L., Kumar, T.S.V.V., Mackey, B.P., O'Shay, A.J., and Pasch, R.J. (2005). The hurricane intensity issue. Mon. Weather Rev. *133*, 1886–1912.
27. Kowch, R., and Emanuel, K. (2015). Are special processes at work in the rapid intensification of tropical cyclones? Mon. Weather Rev. *143*, 878–882.
28. Wang, B., and Zhou, X. (2008). Climate variation and prediction of rapid intensification in tropical cyclones in the western North Pacific. Meteorol. Atmos. Phys. *99*, 1–16.
29. Ryglicki, D.R., Cossuth, J.H., Hodyss, D., and Doyle, J.D. (2018). The unexpected rapid intensification of tropical cyclones in moderate vertical wind shear. Part I: Overview and observations. Mon. Weather Rev. *146*, 3773–3800.
30. Leroux, M.D., Wood, K., Elsberry, R.L., Cayanan, E.O., Hendricks, E., Kucas, M., and Yu, Z. (2018). Recent advances in research and forecasting of tropical cyclone track, intensity, and structure at landfall. Trop. Cycl. Res. Rev. *7*, 85–105.
31. Sinclair, V.A., Rantanen, M., Haapanala, P., Räisänen, J., and Järvinen, H. (2020). The characteristics and structure of extra-tropical cyclones in a warmer climate. Weather Clim. Dynam. *1*, 1–25.
32. Gray, W.M. (1968). Global view of the origin of tropical disturbances and storms. Mon. Weather Rev. *96*, 669–700.
33. DeMaria, M., and Kaplan, J. (1994). Sea surface temperature and the maximum intensity of Atlantic tropical cyclones. J. Clim. *7*, 1324–1334.
34. Elsberry, R.L., and Jeffries, R.A. (1996). Vertical wind shear influences on tropical cyclone formation and intensification during TCM-92 and TCM-93. Mon. Weather Rev. *124*, 1374–1387.
35. Latif, M., Keenlyside, N., and Bader, J. (2007). Tropical sea surface temperature, vertical wind shear, and hurricane development. Geophys. Res. Lett. *34*.
36. DeMaria, M. (1996). The effect of vertical shear on tropical cyclone intensity change. J. Atmos. Sci. *53*, 2076–2088.
37. Wang, Y., and Holland, G.J. (1996). Tropical cyclone motion and evolution in vertical shear. J. Atmos. Sci. *53*, 3313–3332.
38. Rogers, R., Chen, S., Tenerelli, J., and Willoughby, H. (2003). A numerical study of the impact of vertical shear on the distribution of rainfall in Hurricane Bonnie (1998). Mon. Weather Rev. *131*, 1577–1599.
39. Tang, B., and Emanuel, K. (2010). Midlevel ventilation's constraint on tropical cyclone intensity. J. Atmos. Sci. *67*, 1817–1830.
40. Xu, Y., and Wang, Y. (2013). On the initial development of asymmetric vertical motion and horizontal relative flow in a mature tropical cyclone embedded in environmental vertical shear. J. Atmos. Sci. *70*, 3471–3491.
41. Baek, Y.H., Moon, I.J., Im, J., and Lee, J. (2022). A novel tropical cyclone size estimation model based on a convolutional neural network using geostationary satellite imagery. Rem. Sens. *14*, 426.
42. Schmetz, J., Tjemkes, S.A., Gube, M., and Van de Berg, L. (1997). Monitoring deep convection and convective overshooting with METEOSAT. Adv. Space Res. *19*, 433–441.
43. Schmit, T.J., Griffith, P., Gunshor, M.M., Daniels, J.M., Goodman, S.J., and Lebair, W.J. (2017). A closer look at the ABI on the GOES-R series. Bull. Am. Meteorol. Soc. *98*, 681–698.
44. Van Sang, N., Smith, R.K., and Montgomery, M.T. (2008). Tropical-cyclone intensification and predictability in three dimensions. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences. Q. J. R. Meteorol. Soc. *134*, 563–582.
45. Cha, D.H., and Wang, Y. (2013). A dynamical initialization scheme for real-time forecasts of tropical cyclones using the WRF model. Mon. Weather Rev. *141*, 964–986.
46. Saha, S., Moorthi, S., Pan, H.L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., et al. (2010). The NCEP climate forecast system reanalysis. Bull. Am. Meteorol. Soc. *91*, 1015–1058.
47. Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.T., Chuang, H.y., Iredell, M., et al. (2014). The NCEP climate forecast system version 2. J. Clim. *27*, 2185–2208.
48. Weber, N.J., and Mass, C.F. (2017). Evaluating CFSv2 subseasonal forecast skill with an emphasis on tropical convection. Mon. Weather Rev. *145*, 3795–3815.
49. Cha, T.Y., Bell, M.M., Lee, W.C., and DesRosiers, A.J. (2020). Polygonal eyewall asymmetries during the rapid intensification of Hurricane Michael (2018). Geophys. Res. Lett. *47*, e2020GL087919.
50. Lowry, M.R. (2009). Developing a Unified Superset in Quantifying Ambiguities Among Tropical Cyclone Best Track Data for the Western North Pacific [Master's Thesis, Florida State University]. Sementic Scholar. https://diginole.lib.fsu.edu/islandora/object/fsu:175602/datastream/PDF/view.
51. Emanuel, K., DesAutels, C., Holloway, C., and Korty, R. (2004). Environmental control of tropical cyclone intensity. J. Atmos. Sci. *61*, 843–858.
52. Bai, L., Xu, Y., Tang, J., and Guo, R. (2023). Interagency discrepancies in tropical cyclone intensity estimates over the western North Pacific in recent years. Atmos. Sci. Lett. *24*, e1132.
53. Cocks, S.B., and Gray, W.M. (2002). Variability of the outer wind profiles of western North Pacific typhoons: Classifications and techniques for analysis and forecasting. Mon. Weather Rev. *130*, 1989–2005.
54. Kubat, M., Holte, R.C., and Matwin, S. (1998). Machine learning for the detection of oil spills in satellite radar images. Mach. Learn. *30*, 195–215.
55. Rao, V., Lin, L., and Dunson, D.B. (2016). Data augmentation for models based on rejection sampling. Biometrika *103*, 319–335.
56. Longadge, R., and Dongre, S. (2013). Class imbalance problem in data mining review. Preprint at arXiv *1*. https://doi.org/10.48550/arXiv.1305.1707.
57. DeMaria, M., Knaff, J.A., and Sampson, C. (2007). Evaluation of long-term trends in tropical cyclone intensity forecasts. Meteorol. Atmos. Phys. *97*, 19–28.
58. Haghroosta, T., and Ismail, W.R. (2017). Typhoon activity and some important parameters in the South China Sea. Weather Clim. Extrem. *17*, 29–35.
59. Gray, W.M. (1998). The formation of tropical cyclones. Meteorl. Atmos. Phys. *67*, 37–69.
60. Camargo, S.J., Robertson, A.W., Gaffney, S.J., Smyth, P., and Ghil, M. (2007). Cluster analysis of typhoon tracks. Part I: General properties. J. Clim. *20*, 3635–3653.
61. Chan, J.C.L., and Gray, W.M. (1982). Tropical cyclone movement and surrounding flow relationships. Mon. Weather Rev. *110*, 1354–1374.
62. Holland, G.J. (1983). Angular momentum transports in tropical cyclones. Q. J. R. Meteorol. Soc. *109*, 187–209.
63. Chan, J.C.L., and Liu, K.S. (2004). Global warming and western North Pacific typhoon activity from an observational perspective. J. Clim. *17*, 4590–4602.
64. Bruyère, C.L., Done, J.M., Jaye, A.B., Holland, G.J., Buckley, B., Henderson, D.J., Leplastrier, M., and Chan, P. (2019). Physically-based landfalling tropical cyclone scenarios in support of risk assessment. Weather Clim. Extrem. *26*, 100229.
65. Lee, J., Yoo, C., Im, J., Shin, Y., and Cho, D. (2020). Multi-task Learning Based Tropical Cyclone Intensity Monitoring and Forecasting through Fusion of Geostationary Satellite Data and Numerical Forecasting Model Output. Korean J. Remote Sens. *36*, 1037–1051.
66. Fitzpatrick, P.J. (1997). Understanding and forecasting tropical cyclone intensity change with the Typhoon Intensity Prediction Scheme (TIPS). Weather Forecast. *12*, 826–846.
67. LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., and Jackel, L. (1989). Handwritten digit recognition with a

back-propagation network. Adv. Neural Inf. Process. Syst. *2*.
68. Chen, Y., Weng, Q., Tang, L., Liu, Q., Zhang, X., and Bilal, M. (2021). Automatic mapping of urban green spaces using a geospatial neural network. GIScience Remote Sens. *58*, 624–642.
69. Lu, Y., James, T., Schillaci, C., and Lipani, A. (2022). Snow detection in alpine regions with Convolutional Neural Networks: discriminating snow from cold clouds and water body. GIScience Remote Sens. *59*, 1321–1343.
70. Pyo, J., Hong, S.M., Jang, J., Park, S., Park, J., Noh, J.H., and Cho, K.H. (2022). Drone-borne sensing of major and accessory pigments in algae using deep learning modeling. GIScience Remote Sens. *59*, 310–332.
71. Ma, Y., Zhen, Z., Li, F., Feng, F., and Zhao, Y. (2023). An innovative lightweight 1D-CNN model for efficient monitoring of large-scale forest composition: a case study of Heilongjiang Province, China. GIScience Remote Sens. *60*, 2271246.
72. Han, D., Choo, M., Im, J., Shin, Y., Lee, J., and Jung, S. (2023). Precipitation nowcasting using ground radar data and simpler yet better video prediction deep learning. GIScience Remote Sens. *60*, 2203363.
73. Kim, H.H., Seo, D., Jung, J., and Kim, Y. (2022). A Study on Lightweight CNN-based Interpolation Method for Satellite Images. Korean J. Remote Sens. *38*, 167–177.
74. Jung, S., Choo, M., Im, J., and Cho, D. (2022). Generation of Daily High-resolution Sea Surface Temperature for the Seas around the Korean Peninsula Using Multi-satellite Data and Artificial Intelligence. Korean J. Remote Sens. *38*, 707–723.
75. Karasawa, H., Liu, C.L., and Ohwada, H. (2018, March). Deep 3d convolutional neural network architectures for alzheimer's disease diagnosis. In In Asian conference on intelligent information and database systems (Springer), pp. 287–296.
76. Zhu, G., Zhang, L., Shen, P., Song, J., Shah, S.A.A., and Bennamoun, M. (2019). Continuous gesture segmentation and recognition using 3DCNN and convolutional LSTM. IEEE Trans. Multimed. *21*, 1011–1021.
77. Gu, H., Gan, W., Zhang, C., Feng, A., Wang, H., Huang, Y., Chen, H., Shao, Y., Duan, Y., and Xu, Z. (2021). A 2D–3D hybrid convolutional neural network for lung lobe auto-segmentation on standard slice thickness computed tomography of patients receiving radiotherapy. Biomed. Eng. Online *20*, 94.
78. Zhang, N., Chai, X., Li, N., Zhang, J., and Sun, T. (2023). Applicability of UAV-based optical imagery and classification algorithms for detecting pine wilt disease at different infection stages. GIScience Remote Sens. *60*, 2170479.
79. Ham, Y.G., Kim, J.H., and Luo, J.J. (2019). Deep learning for multi-year ENSO forecasts. Nature *573*, 568–572.
80. Sahoo, B., and Bhaskaran, P.K. (2019). Prediction of storm surge and coastal inundation using Artificial Neural Network–A case study for 1999 Odisha Super Cyclone. Weather Clim. Extrem. *23*, 100196.
81. Yan, J., Mu, L., Wang, L., Ranjan, R., and Zomaya, A.Y. (2020). Temporal convolutional networks for the advance prediction of ENSO. Sci. Rep. *10*, 8055–8115.
82. Lee, J., Kim, M., Im, J., Han, H., and Han, D. (2021). Pre-trained feature aggregated deep learning-based monitoring of overshooting tops using multi-spectral channels of GeoKompsat-2A advanced meteorological imagery. GIScience Remote Sens. *58*, 1052–1071.
83. Kwak, G., Park, S., and Park, N. (2022). Combining Conditional Generative Adversarial Network and Regression-based Calibration for Cloud Removal of Optical Imagery. Korean J. Remote Sens. *38*, 1357–1369. https://doi.org/10.7780/kjrs.2022.38.6.1.28.
84. Lee, J., and Nam, J. (2017). Multi-level and multi-scale feature aggregation using pretrained convolutional neural networks for music auto-tagging. IEEE Signal Process. Lett. *24*, 1208–1212.
85. Ruder, S. (2017). An overview of multi-task learning in deep neural networks. Preprint at arXiv *1*. https://doi.org/10.48550/arXiv.1706.05098.
86. Lewis, C.D. (1982). Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting (Butterworth-Heinemann).
87. Kaplan, J., and DeMaria, M. (2003). Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. Weather Forecast. *18*, 1093–1108.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Communicate, Ocean and Meteorological Satellite (COMS) Meteorological Imager (MI) observations	National Meteorological Satellite Administration (NMSC)	https://datasvc.nmsc.kma.go.kr
Climate forecasting system version 2 (CFSv2) operational forecasts	National Centers for Environmental Information (NCEI), National Oceanic and Atmospheric Administration (NOAA)	https://www.ncei.noaa.gov
Joint Typhoon Warning Center (JTWC) Best track	Naval Meteorology and Oceanography Command	https://www.metoc.navy.mil/jtwc
Korea Meteorological Administration (KMA) tropical cyclone forecasting report	KMA	http://afso.kma.go.kr
Software and algorithm		
MATLAB	Mathworks	https://www.mathworks.com/products/matlab.html
Spyder	Anaconda	https://anaconda.org/anaconda/spyder

RESOURCE AVAILABILITY

Lead contact

All requests for additional information and resources should be directed to the lead contact, Jungho Im (ersgis@unist.ac.kr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data can be obtained from the lead contact, provided the request is reasonable.
- The code related to the developed model can be accessed by reaching out to the lead contact.

METHOD DETAILS

Data

Satellite data

Images from the Communication, Ocean, and Meteorological Satellite (COMS) Meteorological Imager (MI) were used as the main input data for TC intensity forecasts (Table S1). The instrument had five channels: one visible and four infrared channels. Since April 2011, the sensor has collected data with a field of view (FOV) of 1–4 km every 15 min. Thermal infrared channels, from shortwave to longwave, have different characteristics depending on atmospheric conditions. Because atmospheric convective systems contain droplets of various sizes, multi-spectral infrared channels can detect them based on particle size. In this study, two infrared channels were used: WV (centered at 6.7 μm) and infrared 1 (IR1 centered at 10.8 μm). WV detects the water vapor content, and IR1 characterizes the cloud-top pattern and has less sensitivity to water vapor absorption than WV does.[42,43] Figure 1 illustrates the trend of the mean brightness temperature at the center of 2019 Typhoon LINGLING based on the pixel distance. Brightness temperatures between the two channels exhibited significant differences as the distance from the TC center increased, depicted by the gray shading in Figure S1. The red boxes in the strong TCs (Figures S1B, S1C, and S1D) indicate deep convective regions with high ice crystal concentrations. For weak TCs (Figure S1A), the brightness temperature difference between the two channels did not change substantially with increasing distance. As the TC intensified, deep convective regions moved closer to the TC center; this result is consistent with the results of dynamic model-based experiments.[44,45] The spatial and spectral characteristics of TCs observed by multi-spectral channels vary considerably with the TC intensity.

Climate forecasting system version 2 operational forecasts

Climate forecasting system version 2 (CFSv2) is a fully coupled forecast model that incorporates interactions among the atmosphere, ocean, sea ice, and land, implemented by the National Center of Environmental Prediction (NCEP) of the National Oceanic and Atmospheric

Administration (NOAA). The first climate forecast system was executed by the NCEP in 2004; it was developed using four predictions: from the NCEP–Department of Energy, Reanalysis 2 (NCEP-DOE R2) for initial atmospheric and land factors, the Global Ocean Data Assimilation System from NCEP for the initial oceanic factor, the Global forecast system from NCEP for atmospheric forecasts, and Modular Ocean Model version 3 produced by the Geophysical Fluid Dynamics Laboratory for ocean forecasts.[46,47] The CFSv2 product contains various data, including two-dimensional energetics, a two-dimensional surface and radiative fluxes, three-dimensional pressure-level data, three-dimensional ocean data, low-resolution output, dumps, and high- and low-resolution initial conditions. These data are provided four times per day (i.e., at 00:00, 06:00, 12:00, and 18:00 UTC) with a horizontal resolution of 1°. Weber and Mass (2017)[48] evaluated the sub-seasonal forecasting skill of CFSv2 with a lead time of 3–5 weeks, focusing on tropical convection. They confirmed that its ocean-atmosphere forecasting skill shows promise for up to 2 weeks. Cha et al. (2021)[49] evaluated the performance of three models—CFSv2, European Center for Medium-Range Weather Forecasting Interim Reanalysis (ERA), and NCEP-R2—focusing on sea surface wind. The CFSv2-based wind vector was more consistent with the buoy-based observations than the others were. When evaluating CFSv2-based TC intensity predictability using the Joint Typhoon Warning Center (JTWC) best track in terms of minimum pressure at mean sea level from 2011 to 2018, the correlations for the 24-, 48-, and 72-h forecasting periods were 0.82, 0.72, and 0.60, respectively.

JTWC best track

The United States publishes best track data for TCs in the western North Pacific (WNP), northern Indian Ocean, and the Southern Hemisphere. These data contain TC information on the location, maximum sustained wind speed, minimum sea surface pressure, and wind intensity by radius. JTWC issues real-time TC warning reports, but the best track provides data processed using observations and numerical reanalysis data. Annual best track data with a temporal resolution of 6 h are officially published once per year.[50] This study focused on the WNP region, where the destructive potential of TCs has increased in recent decades.[51] Maximum sustained wind speed (kts) was used to represent TC intensity with the linear temporal interpolation of the six hourly best track data into hourly data. The best track data from 2011 to 2018 were used to construct a model for predicting TC intensity, and the data from 2019 were used to evaluate the model. The samples from 2011 to 2018 were further randomly divided into training (80%) and validation (20%) sets by track. A validation set was used to optimize the model parameters.

KMA-based TC forecasting report

KMA releases 5-day TC forecast reports. Similar to other TC prediction agencies (e.g., JMA and JTWC), they subjectively determine TC forecasts using TC observations and a numerical weather prediction model called the Global Data Assimilation and Prediction System (GDAPS). Compared with the TC intensity forecasting performance of GDAPS, the intensity forecasts determined through the official guidance of KMA improved by 35%, 29%, and 25% for the 24-, 48-, and 72-h TC forecasting periods, respectively.[15] KMA-based forecasts demonstrated a performance comparable to that of the regional specialized meteorological center in Tokyo, which yielded excellent performance for TCs in the WNP.[7,8,15] In this study, KMA-based 24-, 48-, and 72-h TC forecasts were used to evaluate the proposed model; hereafter, they are referred to as K24, K48, and K72, respectively. Given the difference in intensity definition standards between KMA and JTWC, it becomes necessary to adopt a scaling factor when comparing them. KMA adheres to a sustained wind speed of 10 min, whereas JTWC uses a sustained wind speed of 1 min. Recently, the revised scaling factors were proposed by Bai et al. (2023) in order to facilitate the comparison of various standards-based intensities.[52] Following Bai et al. (2023), the category-wise scaling factors were set as 0.9 for category 1 TCs (Intensity <48 kts), 1.0 for category 2 TCs (48 kts $\leq$Intensity <64 kts), 1.1 for category 3 TCs (64 kts $\leq$Intensity <85 kts), 1.2 for category 4 TCs (85 kts $\leq$Intensity <105 kts), and 1.3 for category 5 TCs (105 kts $\leq$ Intensity).

Overall methodology

Based on multiple types of datasets, 24-, 48-, and 72-h TC intensity forecasting models were proposed (Figure S2). Three types of data sources were used:1) the JTWC best track, 2) satellite observations, and 3) numerical model-based predictions. A deep learning-based hybrid algorithm (i.e., hybrid-CNN) was proposed for the effective fusion of different types of datasets. The hybrid-CNN consists of several CNN-based models that characterize each input data type. Synergistic fusion of input variables is anticipated through parameter sharing among sub-models.

Data preprocessing

Satellite data preprocessing

Based on the JTWC best track, consecutive TC observations from two infrared channels of COMS MI (WV and IR1) from 2011 to 2019 were collected. Because more than 90% of the TCs developed in WNP had a radius up to 450 km,[53] a square region of 1204 km (i.e., 301 pixels) on one side was extracted as a TC observation. The extracted 301 x 301 images were resized to 101 x 101 pixels to reduce computational demand. This procedure was repeated for the IR1 and WV channels, and each image was normalized using the minimum and maximum values from both images. Because satellite observations for the entire TC lifetime were used in this study, 12% of the collected TCs had an intensity greater than 96 kts, and 70% had an intensity less than 63 kts. The sample size was not evenly distributed by intensity, which may have resulted in biased training with deep learning.[54–56] Thus, subsampling of the training data was performed to mitigate this limitation. In the 10 kts interval histogram, over-frequent samples exceeding the upper 25% frequency were randomly removed. Data augmentation through image

rotation was then performed to increase the overall sample size. These two preprocessing steps (i.e., image upscaling and sub/over-sampling processing) allowed the training samples to be balanced and sufficient. The preprocessing referred to Lee et al. (2020),[20] who proposed a deep learning-based TC intensity estimation approach using multi-infrared channels from the same geostationary satellite sensor. The six hourly time series observations from the previous 72 h (i.e., 13 observations) were used to model 24-, 48-, and 72-h forecasts (Figure S3).

Numerical model-based forecasts

In this study, the CFSv2 atmosphere and ocean forecasts were used. They provided atmospheric predictors by altitude and ocean predictors by water depth as 3-D pressure-level data (PGB) and 3-D ocean data (OCN), respectively. PGB provides predictions from the time of initialization, and OCN provides predictions starting with a 6-h forecast. The 10 PGB prediction members were used from the current to the forecasting time, and the seven OCN prediction members from a 6-h forecast to a later time were also used (Table S2). These variables have been widely used in statistical models to predict TC intensity.[11,12,18,57,58] Because the 17 environmental prediction variables had varied value ranges, min-max normalization was conducted for each variable. Because of the possibility of perturbational errors in the numerical model-based predictors,[6] the lower and upper 95% of the whole pixel-wise values of each environmental predictor were used as the minimum and maximum values for normalization. This normalization approach enables each predictor to preserve the relative values of each predictor while allowing it to maintain its spatial pattern. The region including the WNP (i.e., 0–45 °N, 80–163 °E) was extracted from global-scale predictions to focus on the TCs that developed in the WNP; thus, the WNP input dataset was 46 × 83 pixels. Because the model forecasts were initialized and provided four times per day (i.e., 00:00, 06:00, 12:00, and 18:00 UTC), they were linearly interpolated into hourly forecasts (i.e., at 00:00, 01:00, …, 23:00 UTC) to match the hourly interpolated best track.

Although large-scale WNP input data can consider synoptic tele-connective environmental impacts on a target TC, distinguishing the environmental factors affecting the targeted TC when multiple TCs occur simultaneously within the WNP remains difficult. To avoid confusion, this study extracted the area surrounding a TC was and used it as a local environmental input dataset. Because the atmosphere-ocean environmental variables within 800 km of a TC center could directly affect the TC,[59] a 17 × 17 pixel-squared region with a radius of 8° was extracted as the local input dataset. In this study, 24-, 48-, and 72-h forecasting environmental variables with a 6-h interval were used for the 24-, 48-, and 72-h intensity forecasting models, respectively.

Past trajectory and intensity observations from JTWC best track

TC trajectory is closely related to TC intensity change.[60] Large-scale interactions between the atmosphere and ocean determine the trajectory structure and movement pattern of TCs.[61–65] To consider the impact of TC movement on intensity change, this study used trajectory images as input data. TC trajectory images over the previous 72 h were created using three classes: trajectory, continent, and ocean, which were represented as values of 1, 0.5, and 0, respectively (Figure S4). They have the same dimensions as the CFSv2-based WNP input dataset (i.e., the lengths of the x axis and y axis were 83 and 46 pixels, respectively). In addition, initial TC intensity and previous changes in TC intensity have been identified as significant predictors.[33,66] Therefore, the consecutive maximum sustained wind speed over the past 72 h provided by the JTWC best track was used as the past intensity trend. The hourly intensity trend (73 × 1) extracted from the hourly interpolated best track was used in the TC intensity prediction models.

Deep learning-based TC intensity forecasting models

CNN

CNNs are neural networks suitable for image recognition.[67] A CNN consists of convolutional, pooling, and fully connected layers, and the kernels used as weights in the convolutional layers are shared in the network. During network training, the parameters—kernels and biases—are optimized. The well-optimized kernels of each convolutional layer activate significant regions using various activation functions (e.g., rectified linear unit [ReLu], hyperbolic tangent [tanh], or linear), which are then used to extract the output. CNNs are classified into 1-dimensional CNN (1D-CNN), 2-dimensional CNN (2D-CNN), and 3-dimensional CNN (3D-CNN) based on the dimension of the kernels used in the convolutional layers. 2D-CNN has been frequently used to analyze images with multi channels (e.g., urban area classification, snow detection using satellite observations), and 1D-CNN has been widely used to analyze one-dimensional datasets, such as ocean pigments forecasting.[68–74] 3D-CNN has been used to analyze data with more than two dimensions, such as multi-layered magnetic resonance imaging and human gesture video.[75–77] Recently, several studies in the atmospheric and climate fields have used CNNs to analyze various types of variables, ranging from one-dimensional climate indices (e.g., El Nino-Southern Oscillation [ENSO]) to three-dimensional images (e.g., satellite observations and numerical model products).[78–83]

Hybrid convolutional neural networks (hybrid-CNNs)

Because various data types have become available for environmental modeling, several advanced deep learning approaches have been proposed to incorporate them. For example, feature-aggregated multimodal deep learning was proposed to enable the efficient analysis of various types of input data,[84] and multi-task learning was proposed to allow input datasets to be trained for multiple outputs by sharing their parameters.[85] These approaches have been adopted for meteorological applications, such as detecting overshooting tops using multi-dimensional satellite observations and predicting short-term precipitation with multi-output tasks, resulting in significant improvement in performance.[82]

Since various datasets (i.e., satellite-based TC observations, numerical model-based environmental predictions, past TC trajectory, and past TC intensity) have been used for TC intensity prediction, their impact and interactions on TC intensity forecasts should be carefully investigated. Therefore, a hybrid-CNN approach that combines feature-aggregated multimodal deep learning and multi-task learning was proposed (Figure S5; Table S3). The proposed model improved the network's understanding of our diverse input datasets by training each data type using individual CNN-based models (hereafter referred to as sub-CNNs) with a multi-task learning approach. The hybrid-CNN model consists of four modules: three sub-task modules (i.e., sub-models A, B, and C [SM-A, B, and C]) and one main task module (i.e., forecasting model [FM]). Previous satellite TC observations were pre-trained using the current TC intensity via a sub-CNN (SM-A), and local environmental variables from the numerical prediction model and the past trajectory and intensity trend were individually pre-trained with target TC intensity (SM-B and SM-C). In the present study, the rectified linear unit (ReLU) activation function was employed across all the neural networks to efficiently handle the non-linearity of each dataset. To account for each module-wise TC intensity forecasting-related characteristics, the total loss is calculated, which includes not only the loss from FM but also the losses from SMs. The features extracted from each sub-model were combined to train the FM to forecast TC intensity. Parameter sharing in the main model (i.e., FM) rendered all sub-CNN parameters interactive. In this model, the L1 loss function was employed, and the adoptive moment estimation (Adam) optimizer was implemented with a learning rate of 0.001. Three hybrid-CNN models were constructed to forecast TC intensity with lead times of 24, 48, and 72 h (hereafter, H24, H48, and H72, respectively).

Performance evaluation

H24, H48, and H72 were evaluated using data from 27 typhoons in 2019. All test cases were categorized into five stages (i.e., categories 1–5) based on TC intensity (Table S4). Results from hybrid-CNN models were compared with the KMA-based operational TC forecasting results. Four metrics were used to evaluate the performance of the proposed models and KMA-based models: MAE, R, mean absolute percentage error (MAPE), and skill score (SS). MAE, R, and MAPE were used to evaluate individual models, and SS was used to compare the proposed model to the KMA-based operational forecasts.

$$MAE = \frac{1}{n} * \sum abs\left(y_{pred} - y_{observed}\right) \quad \text{(Equation 1)}$$

$$R = \frac{\sum\left(y_{pred} - \overline{y_{pred}}\right) * \left(y_{observed} - \overline{y_{observed}}\right)}{\sqrt{\sum\left(y_{pred} - \overline{y_{pred}}\right)^2 * \sum\left(y_{observed} - \overline{y_{observed}}\right)^2}} \quad \text{(Equation 2)}$$

$$MAPE = \frac{100\%}{n} \sum \left| \frac{y_{observed} - y_{pred}}{y_{observed}} \right| \quad \text{(Equation 3)}$$

$$SS = \frac{MSE_{control} - MSE_{model}}{MSE_{model}} \quad \text{(Equation 4)}$$

where n represents the number of validation samples, and y_{pred} and $y_{observed}$ indicate the predicted and observed intensity (kts), respectively. $MAE_{control}$ and MAE_{model} indicate MAEs of the KMA forecasts and MAEs of the proposed model, respectively. MAPE has been widely used to evaluate forecasting models[86]: good forecasting (MAPE<20), reasonable forecasting ($20 \leq$ MAPE ≤ 50), and inaccurate forecasting (50<MAPE).

Before evaluating the performance of the proposed models with respect to KMA-based forecasting, the effectiveness of hybrid-CNN-based data fusion was compared with that of deep learning-based TC intensity forecasting models without data fusion, using satellite observations (COMS-CNN) or numerical model data (CFSv2-CNN). The forecasting results were also evaluated by focusing on rapid intensification (RI), defined as the intensity gradient (kts) per 24 h exceeding 30 kts[87] All RIs that occurred in 2019 were used to validate the performance of the models for rapid changes in intensification. Here, RI cases were examined independently because the hybrid-CNN-based TC intensification forecasts were simulated on each TC. In addition, the consecutive forecasting performance of the hybrid-CNN-based model was evaluated using Typhoon KAMMURI (2019). A heatmap, an explainable artificial intelligence approach, was used to examine how the model led to the prediction results. This approach enabled the identification of the contribution and interaction of the input data to the model.