Machine Learning with Applications 5 (2021) 100038

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

FogNet: A multiscale 3D CNN with double-branch dense block and attention mechanism for fog prediction

Hamid Kamangir [a,d,*], Waylon Collins [b,d], Philippe Tissot [a,d], Scott A. King [a,c,d], Hue Thi Hong Dinh [a,c,d], Niall Durham [a,d], James Rizzo [a]

[a] Conrad Blucher Institute for Surveying and Science, Texas A&M University - Corpus Christi, Corpus Christi, TX, USA
[b] National Weather Service, Corpus Christi, TX, USA
[c] Department of Computing Sciences, Texas A&M University - Corpus Christi, Corpus Christi, TX, USA
[d] NSF AI Institute for Research on Trustworthy AI in Weather, Climate and Coastal Oceanography, USA

ARTICLE INFO

Keywords:
Fog prediction
Deep learning
3D Convolutional Neural Network
Dense block
Attention mechanism
Dilated convolutions

ABSTRACT

The reduction of visibility adversely affects land, marine, and air transportation. Thus, the ability to skillfully predict fog would provide utility. We predict fog visibility categories below 1600 m, 3200 m and 6400 m by post-processing numerical weather prediction model output and satellite-based sea surface temperature (SST) using a 3D-Convolutional Neural Network (3D-CNN). The target is an airport located on a barrier island adjacent to a major US port; measured visibility from this airport serves as a proxy for fog that develops over the port. The features chosen to calibrate and test the model originate from the North American Mesoscale Forecast System, with values of each feature organized on a 32×32 horizontal grid; the SSTs were obtained from the NASA Multiscale Ultra Resolution dataset. The input to the model is organized as a high dimensional cube containing 288 to 384 layers of 2D horizontal fields of meteorological variables (predictor maps). In this 3D-CNN (hereafter, FogNet), two parallel branches of feature extraction have been designed, one for spatially auto-correlated features (spatial-wise dense block and attention module), and the other for correlation between input variables (variable-wise dense block and attention mechanism.) To extract features representing processes occurring at different scales, a 3D multiscale dilated convolution is used. Data from 2009 to 2017 (2018 to 2020) are used to calibrate (test) the model. FogNet performance results for 6, 12– and $24-h$ lead times are compared to results from the High-Resolution Ensemble Forecast (HREF) system. FogNet outperformed HREF using 8 standard evaluation metrics.

1. Introduction

Fog is defined as a suspension of very small water droplets in proximity with the earth's surface that reduces visibility (Glickman, 2000). A distinction is made between fog, which reduces surface visibility to less than 1 km, and mist, which does not reduce surface visibility to the 1 km threshold (WMO, 2020). However, in this study, we will use the word fog to refer to both mist and fog. The reduction of visibility adversely affects land, marine, and air transportation owing to the associated human and economic costs. Thus, the ability to accurately and skillfully predict fog would provide tremendous utility. In this study, we desire to predict fog by post-processing numerical weather prediction (NWP) model output using a subset of machine learning (ML) known as deep learning (DL) (Goodfellow, Bengio, & Courville, 2016). It has been shown that the future state of a chaotic dynamical system can be skillfully predicted by combining a dynamical model that describes the system with ML (Pathak et al., 2018). In addition, given the nonlinearity of atmospheric processes, the accuracy of atmospheric prediction will deteriorate rapidly without the incorporation of atmospheric dynamics and physics (Kalnay, 2003; Warner, 2011), further adding credence to the use of NWP model output. Further, the post-processing of NWP output can reduce systematic errors, which is also relevant to the present study since systematic errors can represent a significant portion of total error, especially near the surface (Hacker & Rife, 2007; Kleiber et al., 2011; Warner, 2011).

Fog is the result of microphysical processes that govern the formation, growth, and dissipation of cloud particles. In addition, fog development requires a favorable atmospheric boundary layer (PBL) structure. The PBL is the atmospheric layer in contact with the earth's surface where turbulence is the dominant mechanism responsible for the vertical transport of heat, momentum, and moisture (Stensrud, 2009; Stull, 1988). However, the accurate and skillful prediction of

* Corresponding author at: Conrad Blucher Institute for Surveying and Science, Texas A&M University - Corpus Christi, Corpus Christi, TX, USA.
E-mail addresses: hamid.kamangir@tamucc.edu (H. Kamangir), waylon.collins@noaa.gov (W. Collins), Philippe.Tissot@tamucc.edu (P. Tissot), Scott.King@tamucc.edu (S.A. King), tdinh8@islander.tamucc.edu (H.T.H. Dinh), Niall.Durham@tamucc.edu (N. Durham), James.Rizzo@tamucc.edu (J. Rizzo).

https://doi.org/10.1016/j.mlwa.2021.100038
Received 5 January 2021; Received in revised form 15 April 2021; Accepted 2 May 2021
Available online 8 May 2021
2666-8270/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

the occurrence, location, and timing of fog within the PBL using operational NWP models is very difficult. Microphysical processes occur on the scale of individual aerosols and cloud particles (Glickman, 2000) which cannot be resolved by operational NWP models. Instead, NWP models parameterize (use resolved variables to account for the implicit effects of) these unresolved processes which introduces error (Gultepe, Milbrandt, & Zhou, 2017; Koračin et al., 2014). In particular, microphysical parameters critical for the accurate prediction of fog in NWP models include liquid water content (LWC), particle size (r), and fog droplet number concentration (Nd). Yet, the parameterization of these microphysical variables combined with the inadequate spatial resolutions of operational NWP models can result in large uncertainties in LWC and Nd (Gultepe et al., 2017). Further, the difficulty to predict these microphysical parameters is compounded by the high variability of such in nature. For example, the relationship between Nd and aerosol number concentration (Na) is highly variable (Gultepe & Isaac, 1999). Finally, aerosols influence r and hence visibility; for a given mixing ratio, a greater Na results in a larger number of smaller cloud–fog particles and a corresponding larger surface-to-volume ratio, which results in greater extinction (lower visibility) (Koračin et al., 2014; Stoelinga & Warner, 1999). With regard to PBL structure, an adequate representation of turbulence in the NWP model equations requires a model grid spacing $\leq$ 50 m (Stensrud, 2009). Since operational NWP models (at least within the United States) have horizontal grid spacings $\geq$ 4 km, turbulence processes are also parameterized, which contributes to errors in PBL structure and also to uncertainties in LWC and Nd (Gultepe et al., 2017).

In addition to errors associated with the microphysical and PBL parameterizations, the chaotic nature of the atmosphere with extreme dependence on the initial condition (Lorenz, 1963, 1969), is also problematic. NWP is an initial value problem and the initial values provided to the microphysical and PBL parameterizations likely contain errors which lowers practical predictability (Melhauser & Zhang, 2012). Lastly, fog development occurs on the microscale time and space scales ($<$ 1 h and $<$ 2 km; Orlanski, 1975) which creates a challenge to accurate prediction. For example, an Ontario Canada fog event adversely affected 82 motorists, yet dense fog occurred over a spatial and temporal scale of only a few kilometers and for less than 1 h, respectively (Pagowski, Gultepe, & King, 2004). Not only are microscale phenomena not resolvable in the current operational NWP models, but also the intrinsic predictability of the atmosphere is positively correlated with spatial scale (Lorenz, 1969).

One state-of-the-art method used by operational forecasters to account for atmospheric chaos when predicting fog is to utilize an ensemble of NWP model runs to quantify the level of uncertainty inherent in single NWP model deterministic runs (Leith, 1974). Based on the performance of individual ensemble members, the probability that atmospheric variables will reach particular thresholds, or whether specific atmospheric phenomena will occur, can be generated. Ensemble prediction systems (EPS) used by operational meteorologists in the United States National Weather Service (NWS) to assist in the forecasting of fog events include the High-Resolution Ensemble Forecast (HREF) and the Short-Range Ensemble Forecast (SREF) (SPC, 2020). These ensemble systems provide probabilities of visibility below particular thresholds based on the performance of the individual ensemble members. For example, output from the HREF includes parameters for the probability of visibility less than 1600, 3200, and 6400 m. Studies have demonstrated the performance enhancement of NWP model ensembles over single deterministic runs in predicting fog (Zhou & Du, 2010; Zhou, Du, Gultepe, & Dimego, 2012).

However, as mentioned earlier, Pathak et al. (2018) has demonstrated that the combination of output from a single dynamical model that describes a chaotic system (e.g. the atmosphere) with ML can skillfully predict the future state of the chaotic system. In addition, post-processing of NWP output via ML can remove systematic errors, which may be a significant portion of total model error. Furthermore, given that a model ensemble involves running multiple NWP model runs simultaneously, the development of a predictive fog model by combining a single dynamical model and ML, that can perform at least similar to that of a corresponding ensemble system, would represent a tremendous computational cost savings and hence high utility.

A literature search reveals that ML has been used to predict fog. Fabbian, De Dear, and Lellyett (2007) used 8 features from meteorological observations to develop a MLP artificial neural network with 2 hidden layers (and with 3 to 20 neurons per hidden layer) to predict fog occurrence at the Canberra International Airport (YSCB) in Australia, with 3, 6, 12, and 18 h lead times. Also, Zazzaro, Pisano, and Mercogliano (2010) proposed a cost sensitive classifier using the BayesNet algorithm in order to develop a binary classifier to predict fog occurrence, based on thirteen years of meteorological observations from 1990 to 2002 at the Trapani Milo airport (Italy). In another project Saurabh and Dimri applied NNs for fog prediction in Delhi, India using meteorological input data from 1996 to 2015 (Saurabh & Dimri, 2016). A novel artificial neural network with multi-objective evolutionary training has been proposed by Durán-Rosal et al. (2018) for binary fog event classification from meteorological input variables including temperature, relative humidity, wind speed, wind direction and runway visual range at Valladolid airport (Spain). Despite these major advances in the application of fog prediction using ML techniques, the availability of observations is still insufficient for ML models developed by post-processing only observations to learn the non-linear interactions between predictors (features) and fog necessary to predict fog 6 to 24 h in advance with sufficient skill and in a consistent manner. In our study, we use numerical weather prediction (NWP) model output as features to better account for the dynamics of the system and the nonlinear relationships between the features and fog in order to develop a skillful fog prediction model.

For decades, conventional machine learning models were limited to processing raw data. In addition, these pattern recognition systems required careful engineering and considerable domain expertise to extract features from the raw data (Goodfellow et al., 2016). But, recently, Deep Learning (DL) models have demonstrated successes with respect to predictive skill, and have emerged as state-of-the-art across a broad range of applications in computer vision, natural language processing, earth science, and medicine. DL models have shown a powerful ability to automatically extract hierarchical and nonlinear features from raw input data. DL models transform the representation from raw input data into representations at a higher level which results in an understanding of very complex functions. Specifically, for classification, multiple layers of representation would amplify patterns of the input data that are important for discrimination.

Convolutional Neural Networks (CNNs) are designed to process data in the form of multiple arrays such as an image. The input data for CNNs are in the form of multiple arrays, for example 1D for signal, 2D for image and 3D for video or volumetric data. Local connection, shared weights, pooling, and the ability to use many non-linear layers are four key ideas behind CNN (Goodfellow et al., 2016). Weight sharing increases model efficiency by reducing the number of parameters to learn in comparison with fully connected neural networks (traditional neural network) (Goodfellow et al., 2016; Schmidhuber, 2015). Traditional ML models, including MLP, lose the spatial pattern of input data, while CNN retains the spatial correlation between pixels and the pattern of variables.

Over the past four years DL models have been increasingly developed for environmental and meteorological applications (Chaabani et al., 2018; Kamangir, Collins, Tissot, & King, 2020; Khaefi, Pramestri, Amin, & Lee, 2018; Li, Fu, & Lo, 2017; Xiao et al., 2019; Zhou, Zheng, Li, Dong, & Zhang, 2019). For fog predictions, Kipfer (2017) applied a 2D CNN model to predict fog with 24 h lead time for the Zurich Airport. The model was built based on 15,000 cases over 1.4 years of observation and uses seven predictors. To address the skewed balance of their dataset towards high visibility observations, the authors used a transfer

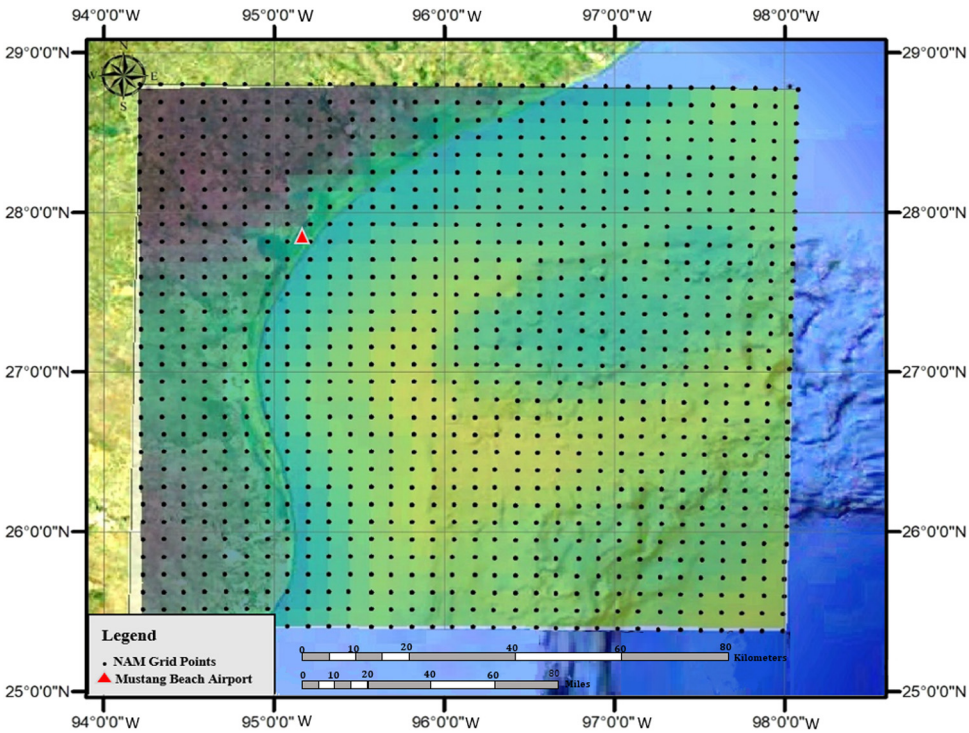

Fig. 1. Study area: Coastal Bend of South Texas, Corpus Christi.

learning approach for tackling this problem showing that this can be accomplished even though the learned features from ImageNet are different than the ones used for their meteorological observations. Miao et al. (2020) tested the application of a LSTM model for short term fog forecasting based on a 3 year time series of meteorological variables including temperature (TMP), air pressure, wind speed, wind direction, rainfall, dew point temperature (DPT), relative humidity (RH), visibility per minute and hourly visibility provided by The Anhui Provincial Meteorological Bureau up to next four hours.

Prior studies using DL models for fog prediction either did not validate on a large time series dataset (Kipfer, 2017) or did not account for vertical structure of the boundary layer (Kipfer, 2017; Miao et al., 2020). Kipfer (2017) had to use transfer learning to take full advantage of DL models to benefit from the correlation between several input variables (represented as images) and fog prediction, which is not appropriate since the feature structure of meteorological data is different than natural images. In this work, between 288–384 (depending on the lead time prediction) different atmospheric variables, including variables that account for the vertical structure of the atmosphere below 700 mb, for 11 years from 2009–2020 were used which enabled us to take full advantage of DL models without any need to use transfer learning or transformation to learn the complexity between input variables and fog. Accounting for the vertical structure in the lower atmosphere is critical when predicting fog. Detailed case studies involving the determination of mechanisms responsible for fog events, illustrate the importance of analyzing the vertical structure of the boundary layer (Koračin et al., 2014; Liu, Yang, Niu, & Li, 2011).

In our work, the dataset of model predictors consists of a high dimensional data cube containing 288–384 different 2D layers of horizontal fields of meteorological (NAM) and oceanic (MUR) variables (predictor maps). The stacking of these fog predictor maps can be expressed as 3D images. The dataset represents conditions below the 700 mb vertical pressure level that influence fog development with variables selected because of their likely direct or indirect relevance to the formation of fog (see Table 1).

In order to extract the relevant latent features from the 3D cube of predictors, a 3D CNN approach was selected. The 3D convolutional network can then simultaneously extract 2D spatial pattern of correlation between individual and also combination of variables based on their pattern. 3D convolution image processing is increasingly tested by researchers particularly in the field of Hyper-spectral Image (HSI) processing, a class of remote sensing images with abundant spectral bands and spatial information (Alipour-Fard et al., 2020; Cao & Guo, 2020; He, Li, & Chen, 2017; Ma, Yang, Wu, Zhao, & Zhang, 2019; Roy, Krishna, Dubey, & Chaudhuri, 2019; Song & Choi, 2020; Wang, Dou, Jiang, & Sun, 2018; Xu, Xiao, Wang, & Luo, 2020). These recent applications of 3D-CNNs for HSI classification indicate that 3D-CNNs perform better than 1D-CNNs and 2D-CNNs. 1D-CNNs are too limited to capture broader spatial characteristics or the spectral structure of the input data and 2D-CNNs will not be able to create latent features describing potential correlations between spectral bands (Ma et al., 2019). In our work, We apply a 3D-CNNs rather than a more commonly used 2D CNNs model to enable the potential identification of multidimensional and multivariate features in the input data, in our case correlations between meteorological and oceanic variables.

Prior research on HSI classification has revealed that only using the conventional 3D-CNNs to extract spectral–spatial features results in too many parameters if the input data has considerable spatial redundancy (Woo, Park, Lee, & So Kweon, 2018). Yu and Koltun (2015) proposed a multiscale convolutional filter approach based on a dilated convolution strategy to extract the contextual features at different scales and resolution and to reduce the spatial redundancy while enlarging the receptive field. To enhance the performance of CNNs, researchers have further investigated three of the models characteristics: depth, width and cardinality (Woo et al., 2018). From the first convolutional deep model LeNet architecture (LeCun, Bottou, Bengio, & Haffner, 1998) through the present day, the networks have become deeper for rich representation. As CNN become increasingly deep, the networks provide a higher level of representation and better performance. However, several problems have emerged, including the gradient vanishing problem, a more time-consuming training process and the requirement to increase training sample size to avoid overfitting. The DenseNet mechanism (Huang, Liu, Van Der Maaten, & Weinberger, 2017) has been proposed to address the aforementioned

problems. DenseNet has shown that if a CNN contains shorter connections between layers close to the input and those close to the output, the network will become more accurate and efficient to train. DenseNet alleviates the vanishing gradient problem, increases the global information and substantially reduces the number of parameters. As a new strategy to further explore the discriminative features and enhance the performance of CNNs, a channel attention module and a spatial attention module are also adopted to optimize the feature maps and improve the classification performance (Woo et al., 2018). As will be further described in the Methodology Section 2.4.4, the attention mechanism forces the network to focus on important features and suppress unnecessary ones. The goal of Attention mechanism is to increase representation power in both channel-wise and spatial-wise architecture by emphasizing meaningful features along those two principal dimensions. Moreover, data augmentation, parametric rectified linear unit, dynamic learning rate, batch normalization, and regularization (including dropout and L2) methods are all used to increase classification accuracy and prevent overfitting.

AI and Deep learning are progressing rapidly and making substantial contributions to the environmental sciences. CNNs in particular have shown a remarkable ability to create implicit representations of 2D images facilitating the extraction of predictors. This work builds on these advances while developing a novel architecture 3D architecture to take advantage of a large dataset combining numerical weather predictions and satellite imagery to better predict a high impact event, coastal fog formation. The novel architecture focuses on capturing air–sea interaction and 3D atmospheric processes with further novelties including:

- the use of a more complete feature set created by combining satellite derived information with NWP model output NW,
- the computation of specific predictors quantifying the air–sea interaction based on NWP model output and high resolution satellite imagery,
- the use of Physics based groupings of the predictors (5 groups in this study), prior to processing by the 3D convolutions of the parallel branches of the feature extraction,
- the inclusion of dense blocks to address the vanishing gradient problem and reduce the number of parameters,
- the inclusion of attention mechanisms to improve classification performance by optimizing feature maps while suppressing weak predictor fields,
- the comparison of the performance of the new DL model with that of an operational NWP model ensemble on a 2 year independent dataset.

In this work, we seek to develop 3D-CNNs to predict fog visibility categories at the Mustang Beach Airport in Port Aransas, Texas (KRAS) (USA) (AirNav, 2020). The airport is located on a barrier island near the Port of Corpus Christi (PCC) ship channel, the fourth largest port in the US by tonnage (Fig. 1). We use visibility measurements at KRAS as a proxy for the presence of coastal fog in the PCC ship channel. It is estimated that fog-related PCC closures total tens of million USD daily. The 3D-CNN DL model is developed by post-processing NWP model output, a method that reduces systematic errors associated with NWP model output (Hacker & Rife, 2007) and has been used to improve the predictive skill of a model that predicts the future state of a chaotic dynamical system such as the atmosphere (Pathak et al., 2018). This study is part of a project with the ultimate goal to skillfully predict visibility associated with fog at least 24 h in advance, which requires predictors originating from a dynamical model (e.g. NWP model) (Warner, 2011). The specific DL model used in this work is a 3D convolution based on dilated convolution, double-branch dense block and attention mechanism (hereafter called **FogNet**). This paper is organized as follows: Section 2 contains the methodology, which includes the FogNet model domain, predictor variables (input features), target, FogNet model structure and development, and the reference model used to assess FogNet performance. Section 3 contains the results and discussion, which includes FogNet model performance assessment and comparison to the reference model. Section 4 contains the conclusion.

2. Methodology

2.1. Model domain

The DLNN model domain used in this study is a subset of the domain used in the operational NAM (Section 2.2. Features), and is represented as a grid of 32 × 32 points, with a grid spacing of 12 km. The domain includes a portion of both the northwest Gulf of Mexico and the adjacent Texas coast (Fig. 1). The latitude–longitude pair of the NE, SE, SW, and NW corners of the domain are 29.00°N–94.20°W/25.20°N–94.20°W/25.20°N–98.01°W/29.00°N–98.01°W, respectively. NAM secondary output variables (and variables derived from NAM output) served as the source for predictor variables (features) used to develop the DLNN models. Based on empirical orthogonal function (EOF) and wavelet analysis of visibility data over South Korea, Lee, Lee, Park, Chang, and Lee (2010) found that radiation fog occurs based on meteorological conditions over a time scale of 1 day, and advection fog is affected by such conditions on a scale of at least 1 week. Meteorological phenomena occurring over daily and weekly time scales correspond to horizontal scales of around 200 km and greater than 2000 km, respectively (Orlanski, 1975). Note that the FogNet domain has a horizontal scale of $12 \times 32 = 384$ km. Thus, the domain in this study would appear sufficient to capture meteorological conditions associated with radiation fog, yet insufficient to account for advection fog development. However, fog predictions in this study extend to a maximum of 24 h, corresponding to the 200 km spatial scale. Thus, the FogNet domain is sufficient.

2.2. Features

The features chosen for this study originated from the North American Mesoscale Forecast System (NAM) (Environmental Modeling Center, 2021) developed by the National Weather Service in the United States of America. NAM model configuration is depicted in Appendix. The training, validation, and testing sets were derived from the 2009–2020 period of the NAM. During this period, the NAM evolved from the WRF-NMM (Weather Research and Forecasting Nonhydrostatic Mesoscale Model) (Janjic, Gerrity, & Nickovic, 2001) (29 June 2006 to 30 September 2011), to the NEMS–NMMB (NOAA Environmental Modeling System Nonhydrostatic Multiscale Model on the Arakawa B-grid) (Janjic & Gall, 2012) (1 October 2011 to present.) Further, the NAM underwent significant changes to model numerics, data assimilation, initialization, and to the land surface, radiation, microphysics, and convective parametrizations (Environmental Modeling Center, 2020b). A change to model configuration warrants an update to the statistical relationship between the NWP model-derived features and the target (e.g. Warner, 2011; Wilks, 2011). However, we do not account for NAM configuration changes in this study. The assumption is made that the training of FogNet using data from the frequently changing NAM will not be detrimental, given the success of Kamangir et al. (2020) in developing a high-performance deep learning model to predict thunderstorm occurrence by post-processing NAM output over the 2004–2006; 2009–2012 period. The NAM uses the following post-processing algorithm by Stoelinga and Warner (1999) to diagnose visibility based on extinction caused by various hydrometeors:

$$vis = -\frac{\ln \epsilon}{\beta} \tag{1}$$

The parameter ϵ is a unitless constant (0.02) provided by Koschmeider (1924). β [km^{-1}] is the extinction coefficient which is determined via an empirical relationship. The following equation is the extinction due to

Table 1
DLNN model predictors (features). The number in parentheses to the right of each feature denotes the following: 1 = NAM initialization and predictor variable, 2 = high resolution sea surface temperature MUR products, 3 = combination of NAM and MUR. NAM refers to the North American Mesoscale Modeling System and MUR refers to Multiscale Ultra High-resolution products.

Acronym	Description (units)	Justification as fog predictor
DPT_{2m} (1)	Dew point temperature at 2 meter height (K)	A low dew point depression (TMP–DPT) near the surface is necessary for fog development. Fog can occur with a higher depression (corresponding to $RH \geq 85\%$) in a polluted environment. Surface ($2m$) DPT > SST (sea surface temperature) is essential for advection fog upstream of the target KRAS (Koračin et al., 2014).
VIS (1)	Surface visibility (m)	The prediction of surface visibility based on the following equation from Stoelinga and Warner (1999): $VIS = -(\frac{\ln \epsilon}{\beta})$, where ϵ (liminal constant) = 0.02, and β [km^{-1}] (extinction coefficient) = $144.7C^{0.88}$; C [gm^{-3}] (mass concentration of liquid water)
VV_{LEVEL} (1)	Vertical velocity [LEVEL = 2 m, 975–700 mb] (ms^{-1})	Fog occurs during low vertical velocity and the activation of cloud condensation nuclei (subsequent higher cloud droplet number concentration) is related to vertical velocity (Gultepe et al., 2017). During radiation and advection–radiation fog events, the vertical structure of surface layer radiative cooling is controlled by horizontal wind, vertical wind shear, and vertical velocities (Dupont, Haeffelin, Stolaki, & Elias, 2016).
Q_{LEVEL} (1)	Specific humidity [LEVEL = 2 m, 975–700 mb; 25 mb interval] (gkg^{-1})	Wind in combination with specific humidity (Q) infer moisture advection, critical for generating the unique TMP and RH profiles conducive to radiation, advection, and frontal fog development (Koračin et al., 2014). Vertical profile of Q a proxy for the determination of PBL height (Stull, 1988); Luan, Guo, Guo, and Zhang (2018) found a positive linear correlation between PBL height and visibility.
U_{LEVEL}, V_{LEVEL} (1)	U,V wind [LEVEL = 2 m, 975–700 mb; 25 mb interval] (ms^{-1})	Wind in combination with specific humidity (Q) infer moisture advection which influences fog development (Koračin et al., 2014). Very light surface wind a requirement for radiation fog. Onshore flow a requirement for advection fog over the target (KRAS). Frontal fog is less dependent on wind velocity than advection or radiation fog. During radiation and advection–radiation fog events, the vertical structure of surface layer radiative cooling is controlled by horizontal wind, vertical wind shear, and vertical velocities (Dupont et al., 2016).
TMP_{LEVEL} (1)	Temperature [LEVEL = 2 m, 975–700 mb; 25 mb interval] (K)	Radiation, advection, advection–radiation, and frontal fog are strongly related to TMP and RH vertical structure in the lower levels of the atmosphere. A low dew point depression (TMP–DPT) near the surface is necessary for fog development. Fog can occur with a higher depression (corresponding to $RH \geq 85\%$) in a polluted environment. Surface ($2m$) TMP > SST (sea surface temperature) is essential for advection fog at the target KRAS (Koračin et al., 2014).
RH_{LEVEL} (1)	Relatively humidity [LEVEL = 2 m, 975–700 mb; 25 mb interval] (%)	Radiation, advection, advection–radiation, and frontal fog strongly related to TMP and RH vertical structure in the lower levels of the atmosphere. Cooling required for condensation (and subsequent fog development) inversely related to RH. A low dew point depression (TMP–DPT) near the surface is necessary for fog development. Fog can occur with a higher depression (corresponding to $RH \geq 85\%$) in a polluted environment. RH structure can identify the stratus cloud layer associated with stratus-lowering fog (Dupont et al., 2016).
TKE_{LEVEL} (1)	Turbulent kinetic energy [LEVEL = 975–700 mb; 25 mb interval] (K^2)	If the $\frac{\partial Q}{\partial z} > 0$ condition is violated in the lower levels, turbulent mixing (TKE) of drier air (lower values of RH or Q) may decrease the chance for fog (Toth et al., 2010). TKE contributes to stratus-lowering fog (Dupont et al., 2016)
T_{LCL} (1)	Temperature at the LCL (K)	With all other factors equal, CNN activation is inversely proportional to T_{LCL} (Gultepe et al., 2017).
SST (2)	Sea surface temperature from the MUR SST (K)	$TMP_{2m} > SST$ and $DPT_{2m} > SST$ are essential for advection fog at the target KRAS (Koračin et al., 2014). Fog occurrence over the sea is most favorable for a particular range of SST values (Li, Wang, Fu, & Lu, 2016)
FRICV (1)	Friction velocity (ms^{-1})	There is a strong relationship between near surface FRICV and radiation fog development (Liu et al., 2011).
DPT_{2m} – SST (3)	Surface dew point temperature minus MUR SST (K)	$TMP_{2m} > SST$ and $DPT_{2m} > SST$ are essential for advection fog at the target KRAS (Koračin et al., 2014).
TMP_{2m} – SST (3)	Surface temperature minus MUR SST (K)	$TMP2m > SST$ and $DPT_{2m} > SST$ are essential for advection fog at the target KRAS (Koračin et al., 2014).
$TMP_{2m} - DPT_{2m}$ (1)	Surface temperature minus surface DPT (K)	A low dew point depression (TMP–DPT) near the surface is necessary for fog development (Koračin et al., 2014).

cloud liquid water and fog based on an observational study by Kunkel (1984):

$$\beta = 144.7C^{0.88} \tag{2}$$

The parameter C [gm^{-3}] is the mass concentration of the hydrometer in question generated by the Ferrier–Algo (Aligo, Ferrier, & Carley, 2018) bulk microphysics parameterization used in the NAM. However, this microphysics scheme and post-processing algorithm combination can result in significant visibility prediction errors (Gultepe et al., 2017). In addition to the mass of liquid water, visibility is also influenced by cloud droplet number concentration, cloud drop-size distribution, and aerosol concentration (Gultepe et al., 2017). The NAM microphysics scheme is single moment with respect to cloud water, thus cloud water mixing ratio is calculated yet cloud droplet number concentration is held constant. A double-moment scheme predicts both cloud water mixing ratio and cloud droplet number concentration (Warner, 2011), which allows for applicability across a wider range of environmental conditions with varying cloud droplet number concentrations that occur in nature (Stensrud, 2009). Aerosol concentration affects cloud droplet number concentration and drop-size distribution, which affects extinction. In particular, aerosols contribute to a larger number of smaller cloud droplets, which results in a larger surface-to-volume ratio, which in turn results in a greater extinction and thus a lower visibility; this is known as the first indirect effect (Twomey, 1974). However, the NAM does not account for aerosol contribution to the cloud microphysics scheme (Eric Rogers, personal communication), further limiting the ability of the NAM to skillfully predict visibility. With respect to fog visibility prediction, the NAM is essentially using an empirical relationship to map cloud liquid water mass to the extinction coefficient, without accounting for the variation in extinction due to changes in cloud particle number concentration and cloud drop-size

distribution caused by aerosols and other factors. Lastly, the NAM has a horizontal grid spacing of 12 km, too coarse to resolve atmospheric turbulence and cloud microphysics processes (Kalnay, 2003; Stensrud, 2009) that influence fog development. Thus, the combination of single moment microphysics with respect to cloud water, the lack of aerosols in the NAM microphysics scheme, and the mesoscale horizontal resolution of the NAM simulation, may result in significant visibility prediction errors.

Thus, the NAM visibility diagnostic variable cannot be used alone if highly skillful visibility predictions are desired. In this study, NAM visibility along with additional NAM predictors (1) are post-processed via FogNet in an attempt to predict fog visibility categories at skills exceeding that of predictions from both deterministic NAM and state-of-the-art NWP model ensembles. As mentioned in the Introduction, post-processing NWP model output can correct for systematic errors (potentially a significant portion of total model error) associated with the NWP model. The additional NAM features were chosen based on expertise and a literature search for variables that influence the development of radiation, advection, advection–radiation, frontal, and stratus-lowering fog over southern Texas and along the adjacent coast. Radiational fog occurs owing to radiative cooling within an environment characterized by moist conditions near the surface, dry conditions aloft, and light or calm wind (Koračin et al., 2014). Along the Texas coast, advection fog generally develops owing to the onshore advection of a moist airmass over the cooler sea surface (cold fog example described in Koračin et al., 2014). Advection–radiation fog is unique to coastal environments and occurs in response to the increase in near surface moisture due to onshore flow during the daylight hours, followed by the radiation fog mechanism at night (Bari, Bergot, & Khlifi, 2016). Frontal fog can be subdivided into the following 3 types, as indicated in Glickman (2000): "Warm-front prefrontal fog", "cold-front post-frontal fog", and "frontal-passage fog". The fog formation mechanism is equivalent for the first 2 types, which involves stratiform rain moistening the shallow cold and stable layer below resulting in fog. Although rain is not a predictor variable (feature) in FogNet, the vertical structure of the atmosphere consistent with stratiform rain is captured by FogNet features. Frontal-passage fog can occur in response to the mixture of the warm and cold airmasses within the frontal zone. Stratus-lowering fog occurs when radiatively-cooled air (in response to radiational cooling at the top of a stratus cloud) mixes vertically-downward and cools the sub-cloud layer towards saturation, resulting in the lowering of the cloud base to the surface (Gultepe et al., 2007). Thus, to predict radiation, advection, advection–radiation, frontal, and stratus-lowering fog, the modeler must account for high resolution vertical profiles of wind, temperature, moisture content, and relative humidity in the lower levels of the atmosphere (Gultepe et al., 2017; Koračin et al., 2014). We include the relevant variables as features/predictors in FogNet. Additional features include surface frictional velocity (strongly related to radiation fog; Liu et al., 2011), vertical velocity (magnitude strongly related to fog; Gultepe et al., 2017), turbulent kinetic energy (can decrease chance for fog under certain conditions; Toth et al., 2010), and lifted condensation level (LCL) temperature (inversely related to the activation of cloud condensation nuclei, which results in a higher cloud droplet number concentration; Gultepe et al., 2017). Several features were also chosen to correct for the limitations of the NAM with regard to advection fog prediction. Advection fog can occur when the surface dew point value of near surface air exceeds the corresponding sea surface temperature (SST) (Koračin et al., 2014). However, the NAM does not utilize SST to predict fog (Eric Rogers, personal communication.) Thus, the Multiscale Ultra High-resolution (MUR) satellite-derived SST product (1 km) (Chin, Vazquez-Cuervo, & Armstrong, 2017), and the NAM surface (2-m height) dew point temperature variable, were combined to generate the feature DPT2meters — SST. Table 1 includes the features, the justification for use of each feature, and corresponding references.

The FogNet predictors (see Table 1) are arranged into 5 separate groups whereby each group contains features that possess a similar physical relationship to fog prior to input into FogNet. Group 1 emphasizes the influence of wind and contains the wind-related features FRICV SURFACE, U10-METERS, V10-METERS, U975-700, and V975-700. Radiation fog requires very light or calm wind, whereas advection fog requires moist onshore flow (Koračin et al., 2014). Further, radiational fog is strongly related to frictional velocity (Liu et al., 2011). Rain-induced frontal fog is less wind dependent than radiation or advection fog. Group 2 focuses on the influence of turbulence kinetic energy and contains features TKE975-700 and Q975-700. If the condition $\frac{\partial Q}{\partial z} > 0$ is violated in the lower atmosphere, turbulent mixing (TKE) of drier air (low Q magnitudes) may prevent fog development (Toth et al., 2010). Group 3 incorporates the thermodynamic profile of the lower atmosphere and contains the features TMP2-METERS, TMP975-700, DPT2-METERS, RH2-METERS, and RH975-700. Radiation and advection–radiation fogs that develop within the domain of this study occur within an environment characterized by moist air near the surface, and much drier air immediately aloft (Koračin et al., 2014). Advection fog generally occurs within an atmospheric profile characterized by both statically stable and nearly saturated conditions throughout the lower atmosphere (Koračin et al., 2014). The rain-induced frontal fog cases at KRAS in the dataset were nearly always characterized by a nearly-saturated strong inversion layer below 500 m (not shown). In addition to vertical mixing, the height of the stratus layer and the moisture content magnitude tend to control the formation of stratus-lowering fog (Dupont et al., 2016). Stratus layer height and moisture content can be inferred from the thermodynamic structure. Group 4 accounts for the influence of atmospheric moisture and microphysics, and contains surface visibility (VIS) (see Section 2.2), QSURFACE, TMP2-METERS-DPT2-METERS (dew point depression), TLCL and VVEL975-700. Low visibility due to fog requires sufficient surface moisture (QSURFACE) and a low dew point depression corresponding to an $RH \geq 85\%$ (Koračin et al., 2014). The activation of cloud condensation nuclei (which results in higher cloud droplet number concentrations and thus lower visibilities per Eq. (1) in Section 2.2) is inversely related to both VVEL and TLCL (Gultepe et al., 2017). Lastly, Group 5 accounts for the influence of the sea surface to advection fog, and contains features SST, DPT2-METERS-SST, and TMP2-METERS-SST. Advection fog tends to occur when SST values are within a certain range (Li et al., 2016) and requires that the conditions (DPT2-METERS-SST) > 0 and (TMP2-METERS-SST) > 0 be satisfied (Koračin et al., 2014).

In this work, FogNet has been applied to predict visibility at 6, 12 and 24 h lead times. The number of predictors is adjusted depending on lead time with 288 meteorological variables for 6-hr predictions including initial NAM, 03- and 06-hour NAM predictions. The predictors for the 12 and 24 h predictions consist both of 384 meteorological variables including initial NAM, 03- , 06- and 12-hour NAM predictions for 12 h lead time and initial NAM, 03- , 12- and 24-hour NAM predictions for 24 h lead time.

2.3. Target

In this study, visibility data was collected from the Mustang Beach Airport in Port Aransas, Texas (KRAS) (United States) during the period 2009–2020. A target vector was generated, which contained visibility measurements at 0000, 0600, 1200, and 1800 UTC each day, corresponding to the times of the 6, 12, and 24 hr predictions of the 0000 and 1200 UTC cycles of the NAM. FogNet is designed as a set of binary classifiers for different ranges and lead times, and thus the visibility measurements were converted to categories. Conversion of all visibility values ≤ 6400 meters to categories occurred only for those cases where the KRAS indicated visibility reduction caused by weather codes FG (fog) or mist (BR). Thus, FogNet was trained to predict visibility

reduction due only to fog or mist. FogNet was trained to predict 3 visibility categories as follows:

$$Target_{d,h} = \begin{cases} 0 & VIS \leq (1600, 3200, 6400) \\ 1 & VIS > (1600, 3200, 6400) \end{cases} \quad (3)$$

Where h and d, are the hour and date (UTC), and VIS is the corresponding visibility (meters).

2.4. Multiscale 3D CNN with double-branch dense block and attention mechanism

FogNet is based on 3D convolution, 3D multiscale dilated convolution, parallel-branch dense block and attention mechanism. In this section, these aforementioned fundamental concepts and their importance to their application to the prediction of coastal fog are discussed followed by a detailed description of the architecture of FogNet.

2.4.1. Features extraction with 3D convolution

Convolutional Networks (ConvNet) are composed of several deep hierarchical convolutional layers in order to extract features quantifying a higher level representation for the input data, such as images. At present, there are three main types of Convolutional kernels: 1D, 2D and 3D (Ma et al., 2019). 1D ConvNets explore along one dimension, such as the correlations among the different layers (spectral) at that pixel location, 2D ConvNets explore spatial features in the image and 3D ConvNets explore combined spatial and spectral features in images. As discussed in the introduction section, due to the structure of the atmospheric processes at work during the formation of fog (3D cube of input data), 3D ConvNet is more suitable for the features extraction than 2D or 1D ConvNet. 3D convolutions apply a 3 dimensional filter to the input dataset and moves along the width (w), height (h) and depth (d) to extract potential patterns within the input data. The 3D convolutional layer is calculated using the following expression (Ma et al., 2019):

$$v_{l,j}^{x,y,z} = \phi\left(\sum_{n} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} \sum_{d=0}^{D-1} w_{ljn}^{hwd} o_{(l-1)n}^{(x+h)(y+w)(z+d)} + b\right) \quad (4)$$

where $v_{l,j}^{x,y,z}$ is the pixel value of position (x, y, z) on the jth feature map in layer l, and H, W and D are the height, width and depth of the kernel, respectively. The w_{ljn}^{hwd} is the weight value at position (h, w, d) connected to the nth feature in the $(l-1)$th layer, $o_{(l-1)n}^{(x+h)(y+w)(z+d)}$ represents the input at position $(x+h)(y+w)(z+d)$ with (h, w, d) denoting its offset to (x, y, z), ϕ is the activation function, and b is a bias parameters (Song & Choi, 2020).

2.4.2. Multiscale 3D dilated convolution block

Multiscale feature extraction using ConvNets has been demonstrated to be effective for classification problems (He et al., 2017; Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). This is in part because multiscale convolutions have the power to extract more complex combined spatial–spectral features. The meteorological data relevant to the formation of fog includes 3D patterns with different spatial resolutions which have the potential to be quantified by using different kernels and receptive fields. Dilated convolution using expansion of receptive field aggregates multiscale contextual information without loss of resolution or coverage (Yu & Koltun, 2015). In fact, dilated convolution modifies the convolution filter in different ways at different ranges using different dilation factors. In our work, a multiscale 3D dilated convolution block, based on a dilated convolution module is proposed, and illustrated in Fig. 2 to learn more complicated meteorological features.

The block consists of three 3D-convolutional layers with different dilation factors. Each layer in the block has the same input map with width (W), Height(H) and Depth (D) dimensions and the size of the 3D convolution filters is the same $(3, 3, 3)$. But, the dilation factor (DF) for each layer is different: 1, 2 and 3 with the receptive field (virtual filter size) size of $(3, 3, 3)$, $(5, 5, 5)$ and $(7, 7, 7)$, respectively. We use the "same padding" technique for all filters in the block in order to have the output with same dimensions as the input. "F" is the number of filters, and None refer to the input batch or image. Then, concatenation is applied to aggregate the multiscale features from the different convolutions.

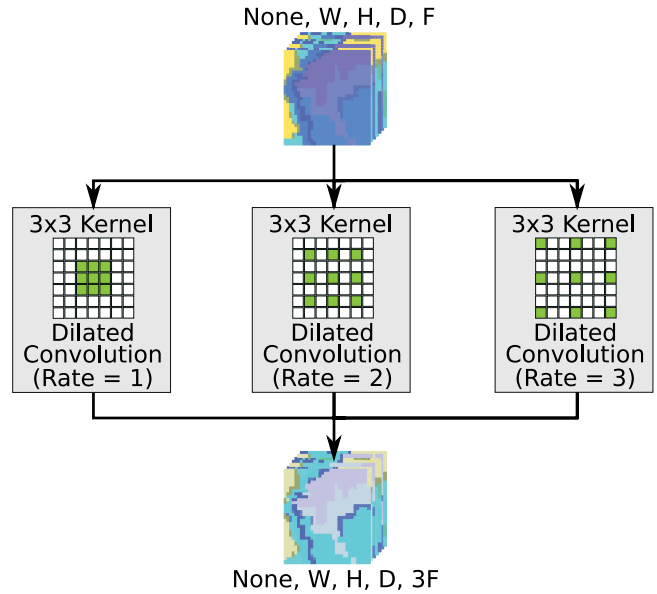

Fig. 2. Multiscale 3D dilated convolution block.

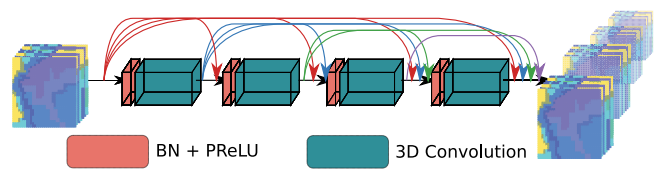

Fig. 3. Dense block with 4 layers, batch normalization and PReLU activation function.

2.4.3. Dense block

Huang et al. proposed DenseNet as a feature extraction method based on the residual mechanism (Huang et al., 2017). The key point in DenseNet is the connection of any hidden layer (H) to all previous layers. All the previous feature maps of the previous $l-1$ layers can be used to compute the output of the lth block:

$$x_l = H_l[f_0, f_1, \ldots, f_{l-1}] \quad (5)$$

where $f_0, f_1, \ldots, f_{l-1}$ are the feature maps of the previous blocks. Each dense block consists of a batch normalization (BN), activation layers and convolution layers. As is shown in Fig. 3, each block has been linked to all previous blocks. As DenseNet combines features by concatenating them, each function H_l produces k feature maps with the $(l+1)$th layer having $k_0 + k \times (l-1)$ input features, while the output of the dense block will still be k feature maps.

2.4.4. Attention module

Depending on the input data structure and complexity, the importance of every input variable and the spatial area of the input map is different while extracting features. Attention mechanism (Xu et al., 2015) is a new technique that can focus on the most informative part and suppress other regions' weights. We apply two attention modules: variable-wise attention module to focus on informative input variable

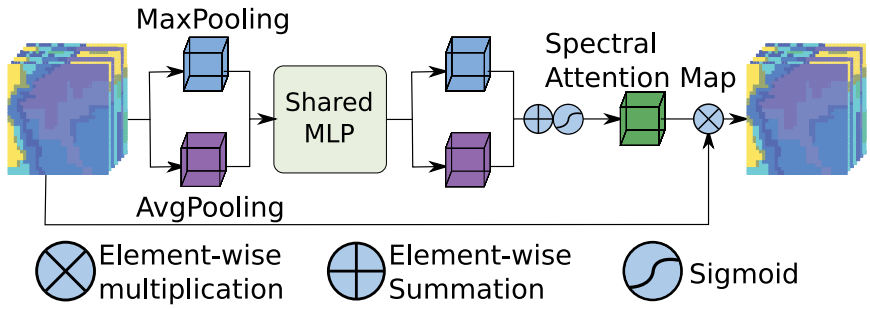

Fig. 4. Structure of variable-wise attention.

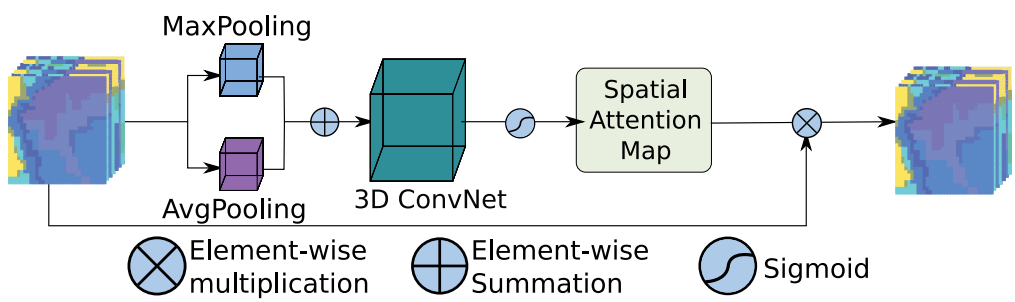

Fig. 5. Structure of spatial-wise attention.

and spatial-wise attention module to extract informative area for each input variable map. We introduce the two modules in detail.

Variable-Wise Attention Module

In a variable-wise attention module each input feature map can be seen as a variable predictor. The goal of variable attention is to focus on the meaningful variables while decreasing the relative contributions of less important variables. As is shown in Fig. 4, a variable-wise attention module has two pooling layers: a MaxPooling layer and an AvgPooling layer to aggregate spatial information with different spatial descriptors: F^c_{avg} and F^c_{max} (Ma et al., 2019). The output features are organized as a one-dimensional vector with the length of the vector equal to the number of input features. Then a shared network composed of a 3-layer perceptron (MLP) with one hidden layer is used to produce the channel attention map. The hidden layer has C/L units, which is used to reduce the training numbers and generate more nonlinear mappings, where L is the reduction ratio and C is the numbers of channel. Note that the output feature vectors are merged using element-wise summation.

The variable attention map is obtained by using sigmoid functions and inputs to the feature maps with values in range of (0, 1). The bigger the value is, the more important the corresponding variable is. Then the variable attention map multiplies the values of the input features to get the variable-refined features. Mapping functions are computing using:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1(W_0(F^c_{avg})) + W_1(W_0(F^c_{max})))$$

where σ is the sigmoid function, $W_0 \in C/L \times C$ and

$W_1 \in C \times C/L$.

Note that the MLP weights are shared for both inputs.

Spatial-Wise Attention Module

The spatial-wise attention focuses on the informative region of the spatial dimension for each variable map. As is shown in Fig. 5, in spatial-wise attention, similarly to variable-wise attention, two types of pooling operations are used to generate different feature descriptors. The pooling operation in the spatial-wise attention module is along the channel axis, so the output feature descriptors are fused by a concatenation operation. Then a convolution layer is applied to the concatenated feature map to generate the spatial attention map. Next, the input feature is multiplied with the spatial attention map to get spatially-refined feature maps which focus on the most informative area. To be summarized, the spatial attention map is computed as:

$$M_s(F) = \sigma(f^{N\times N}([AvgPool(F); MaxPool(F)])) = \sigma(f^{N\times N}([F^c_{avg}; F^c_{max}]))$$

where σ denotes the activation function and we choose the sigmoid function here, $f_{N\times N}$ represents a convolution operation with filter size of $N \times N$.

2.4.5. FogNet

To find an optimal prediction model for coastal fog predictions, we first designed and tested a sequential $3D-2D$ convolution architecture. The results were not satisfying likely because the 3D ConvNet by itself was not able to learn informative features from raw input data and the 2D (spatial) convolution is not capable of accounting for correlations between meteorological variables. To improve the performance of the sequential $3D-2D$ model, we modified the ConvNets architecture and replaced the sequential 3D ConvNets with a dense block to directly take into account the global information from the raw input data through all layers until the last layers and removing the 2D ConvNets. To extract features in different resolutions and scales and to better quantify the relevant interactions within the input data, a 3D dilated convolution block was included. Additionally, an attention mechanism is added to magnify the importance of the most informative meteorological predictors and to magnify the most important spatial area(s) in each feature map. The addition of these three mechanisms to the architecture, dense block, dilated convolution and attention led to significant improvements in the performance of the model and are thought to be key for the performance of FogNet.

Following the work of Wang et al. (2018), who used 3D ConvNet for HSI classification, we tested an architecture that first quantifies correlations between meteorological predictors (equivalent to the different bands for HSIs) and then feeds the extracted features into a spatial dense block to extract the spatial features. However we found that spatial correlations between pixels for each meteorological predictor map were removed through the spectral feature extraction and this

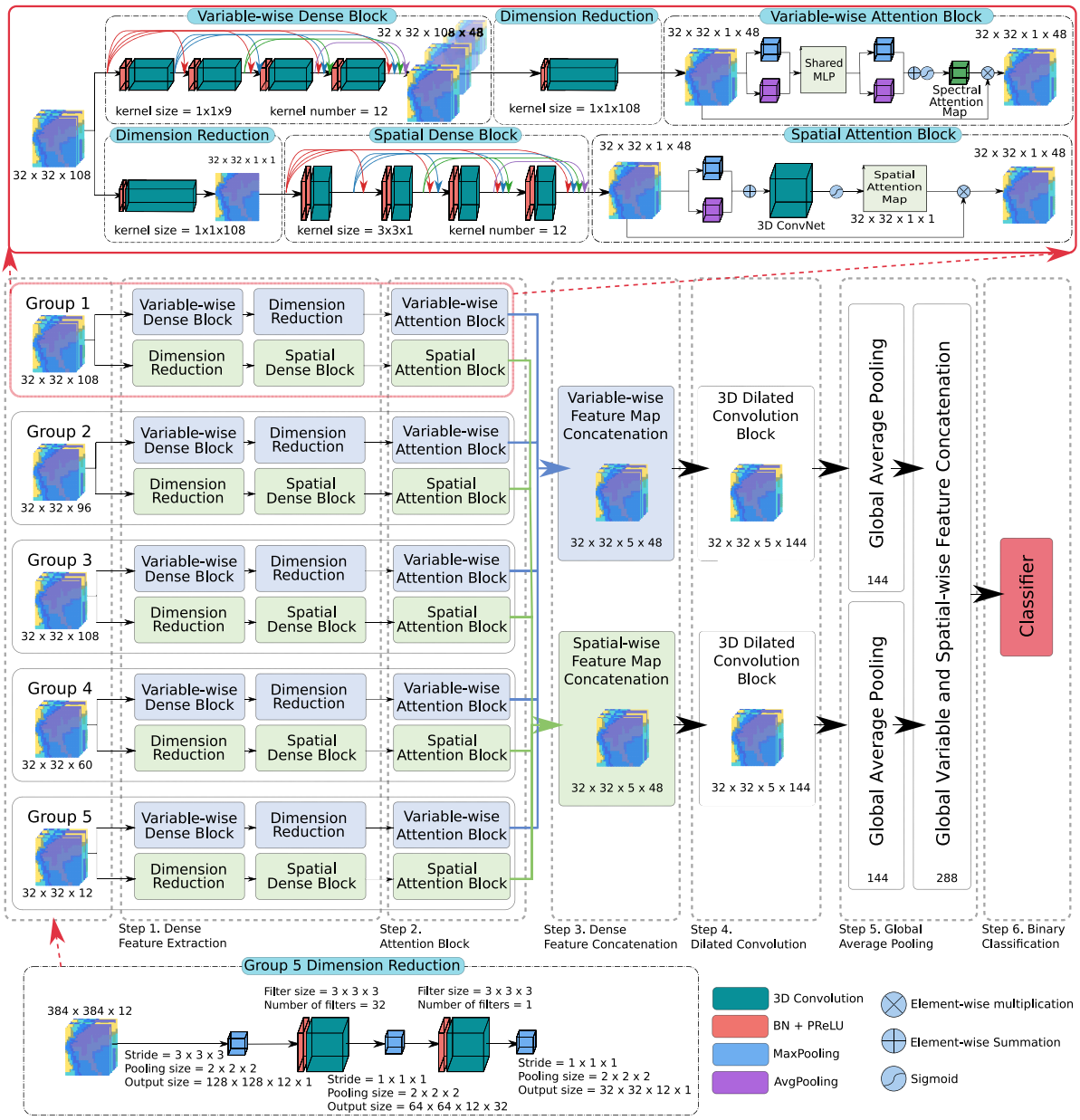

Fig. 6. Structure of the FogNet network. FogNet has 5 groups of predictors each fed into two parallel feature extraction branches (channels): a variable-wise and a spatial-wise both with channel attention. The output features from these two branches are used in multiscale 3D dilated convolutions to extract features at multi resolutions before classification. The depth values for the different groups in this figure is for the 12 hr lead time model.

did not give us performance improvement. Ma et al. (2019) suggested instead to use spectral and spatial feature extraction in sequence, it is better to apply them in parallel and concatenate the output features somewhere before classification. The benefit of this strategy is that the spectral correlations between different maps and spatial correlations between pixels for each map are extracted separately, then all the information for classification is used to label data. Inspired by Ma et al. (2019), we extract correlation between meteorological variables and spatial features for each variable in parallel branches.

The FogNet model (shown in Fig. 6) starts with the separate processing of each of the five input groups. Each subsection consists of a double parallel branch dense block feature extraction (spatial-wise and variable-wise) with attention mechanism. Then, the variable-wise feature outputs from these 5 groups are concatenated to each other. The same process takes place for the spatial-wise feature extraction resulting in two different types of feature maps, spatial-wise and variable-wise. In the next step, 3D dilated convolution blocks are used to further process the information within the two types of feature maps at different resolutions prior to pooling and concatenation and finally binary classification. Generally, FogNet has 6 sequence steps:

1. Spatial-wise and variable-wise auto-correlation feature extraction in two parallel branches.
2. Magnification of the most informative feature map and most informative area in each feature map using double branch attention mechanisms.
3. Concatenation of spatial-wise and variable-wise output feature maps for each group separately.
4. Extraction of multiscale features using a 3D dilated convolution block for each feature type from step 3.
5. Extraction of the global information from these two feature types using global average pooling and concatenating them for the binary classification task.
6. Binary classification into a fog category (visibility greater than or less than a threshold: 1600 m, 3200 m, 6400 m).

The individual processing steps are now described more specifically.

Variable-Wise Branch with Channel Attention

In the variable-wise feature extraction branches, the correlation between the meteorological variables within each of the five groups is considered. The input datasets for each group have the same width and height but different depths depending on the variables of the group, $32\times32\times Depth\times1$. The variable-wise dense block has four 3D convolution layers with batch normalization and a kernel size of $1\times1\times9$ each, with 12 filters used to extract the variable correlation features. In the dense variable-wise block, the stride is set to (1, 1, 1) and the same padding method is used to maintain the size of the feature maps. The variable-wise feature map output size for each group is $32\times32\times Depth\times48$. For example, in Fig. 6 for group 1 including wind-related features for 12 hr lead time predictions the input size is $32\times32\times108\times1$ and the output size for the variable-wise dense block is $32\times32\times108\times48$. In this step, by using a $1\times1\times Depth$ 3D kernel, the dimension of the input maps reduces to 1 for each group but with the same number of features map (48) so the output is $32\times32\times1\times48$. However, the relative importance of the 48 feature maps will vary. To focus on the most important feature maps and obtain more discriminative features, a variable-wise attention block, as illustrated in Section 2.4.4, is applied. After the variable-wise attention block, the most important feature maps will be highlighted while the less important feature maps will be suppressed.

The process in the variable-wise branch with channel attention is the same for all 5 input groups. At the end of this step there are 5 outputs of feature maps with size of $32\times32\times1\times48$. In step 3, all five of the variable-wise feature maps are concatenated to generate new maps with size $32\times32\times5\times48$. In this step, multiscale 3D dilated convolution is used to extract variable-wise and spatial-wise features in parallel at different scales and resolutions. The Multiscale 3D dilated convolution blocks include three 3D convolution kernels with $3\times3\times3$ size and dilation factors 1, 2 and 3, with the same number of filters (48). The output for the multiscale block is feature maps with dimensions $32\times32\times5\times144$ which in the next step, using global average pooling, 144 features are generated.

Spatial-Wise Branch with Channel Attention

In the spatial-wise feature extraction branch, the spatial correlation between pixels for each variable within each group is considered. Before applying the spatial-wise dense block, feature reduction is accomplished using a 3D kernel of $1\times1\times Depth$ for each group to generate a $32\times32\times1$ map which reduces the dimension of the feature maps and the number of model parameters. Similar to the variable-wise dense block, the spatial-wise dense block has 4, 3D convolution layers with batch normalization and a kernel size of $3\times3\times1$ each with 12 filters (48 filters for the block). In the dense spatial-wise block, the stride is set to (1, 1, 1) and uses the same padding method to maintain the size of the feature maps. The spatial-wise feature map output size for each group is the same as the variable-wise map output $32\times32\times1\times48$. To focus on the more important and discriminative areas in each feature map, the spatial-wise attention block, as illustrated in Section 2.4.4, is applied. After the spatial-wise attention block, the important areas and pixels in each map will be emphasized, while the less important pixels of the feature maps will be suppressed. Similar to the variable-wise branch, at the end of this step, there are 5 output feature maps with the same size of $32\times32\times1\times48$. All the spatial-wise feature maps are concatenated to generate new maps with dimensions $32\times32\times5\times48$. A multiscale 3D dilated convolution block is then applied to generate $32\times32\times5\times144$ feature maps, followed by a global average pooling to generate 144 spatial features.

Spatial- and Variable-Wise Fusion for Classification

Through the variable-wise and spatial-wise branches, two different types of features are obtained, including 144 variable-wise features and 144 spatial-wise features. The two sets of features are fused through concatenation resulting in 288 features used for the final binary classification. As the two feature sets where not extracted from the same domain, a concatenation is used instead of an add operation. The final binary classification is obtained by applying a soft-max activation function to the inputs in the globally average pooling layer.

Table 2
Hyperparameter ranges and the tuned values which gave the best performance on validation data for each lead time.

HPs	Range of values	6 h	12 h	24 h
Lr	0.0009–0.01	0.0008	0.001	0.0009
d	0.2–0.6	0.2	0.4	0.4
L2	0.00001–0.01	0.001	0.001	0.001

2.4.6. Fognet hyperparameters tuning

This subsection describes the process used to find the best FogNet hyperparameters based on a grid search (Section 11.4.3 of Goodfellow et al., 2016). The grid search approach was performed using the following 3 steps: First, the range of experimental hyperparameters and associated values were defined (Table 2). Second, FogNet was trained for each combination of hyperparameter values on the training data and evaluated on the validation data. Third, the optimal hyperparameter values were selected based on minimizing the loss on the validation data. All experiments in this work were trained using the Keras Python package (Chollet et al., 2018) and the code is availabel on GitHub.[1]

FogNet training is subdivided into epochs. In each epoch, multiple batches of training examples are presented to the CNN. The Adam optimizer (Kingma & Ba, 2014) was used to update the weights after each epoch in an effort to minimize the loss. In this experiment, FogNet is trained for 50 epochs, with 32 batches per epoch on 5460 training samples and 3328 validation samples. These numbers are chosen to be large enough such that validation loss always converges to a minimum before training is complete. This convergence occurs for FogNet as well, usually well before the 50th epoch, at which point the classifier begins to overfit. Early stopping (Prechelt, 1998) was used, with patience 5, to avoid overfitting.

FogNet has three hyperparameters (HPs) (Table 2) that control overfitting. Without tuning these parameters, the complex interactions between the large number of input meteorological predictor variables (288–384) and fog formation will result in overfitting. The first hyperparameter is the learning rate (lr), which is the main control for model changes after updating model weights during each epoch. Choosing very small values for training rate results in longer training time because it takes longer to reach good weights, whereas too large of a learning rate may result in an unstable training process as weights can jump back and forth between local optima. The second hyperparameter is the dropout rate (d). For each training example and each dense layer, a fraction, d, of the layer's outputs are randomly set to zero (Hinton, Srivastava, Krizhevsky, Sutskever, & Salakhutdinov, 2012). This forces the weights in a given layer to evolve more independently, which reduces overfitting. Dropout is used during training for all dense layers except the last (the last dense layer has only one output, which is the probability that fog will occur). The third hyperparameter is the weight decay, L2, which controls the strength of L2 regularization (Arthur & Robert, 1970). L2 regularization is added to the loss function during training, where L2 is the sum of squared weights in all convolutional layers. This encourages the model to learn smaller weights. Large weights can lead to instability in a model, where small changes in the inputs (predictors) cause sharp changes in the output (prediction). Table 2 shows the hyperparameters that were used to get the best performance on the validation for each lead time.

[1] https://github.com/conrad-blucher-institute/FogNet

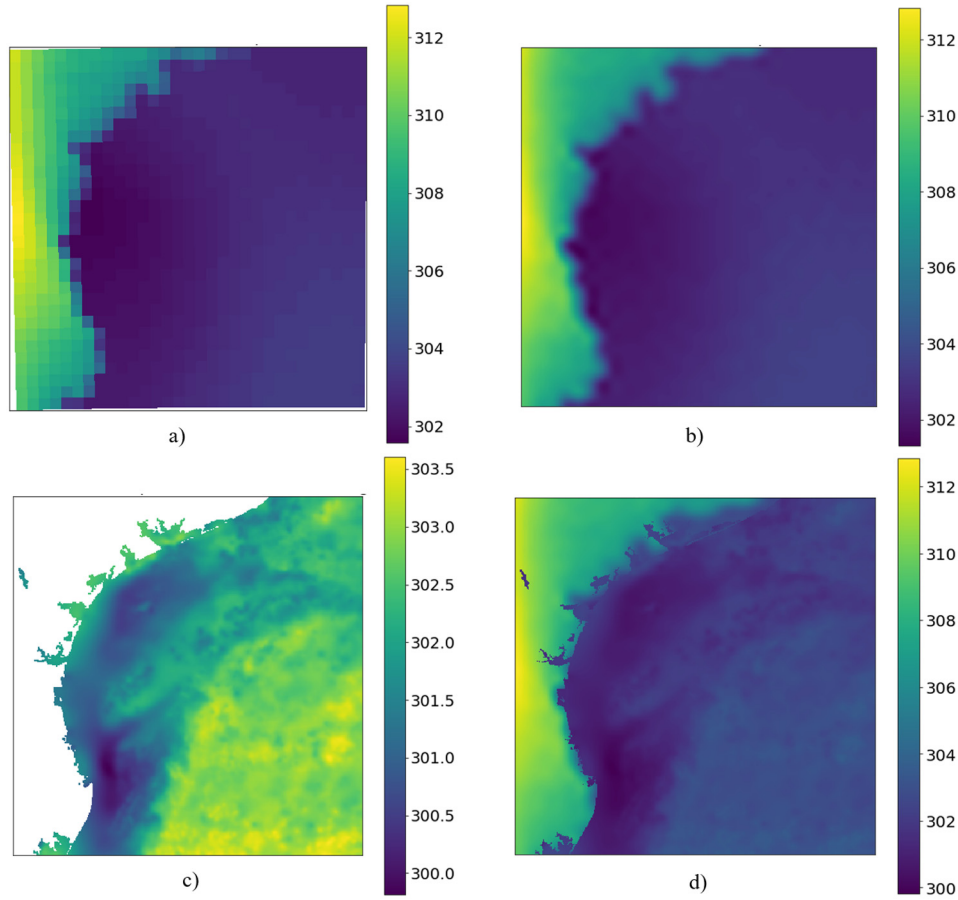

Fig. 7. To generate the variables in group 5, which contains SST, DPT2-METERS-SST, and TMP2-METERS-SST, two main preprocessing steps are needed. First, resampling of DPT2-METERS, TMP2-METERS and Surface TMP from 32 × 32 to 384 × 384 is performed and second, missing values in the land areas of the SST maps are filled with Surface TMP. This figure shows an example of two preprocessing steps on the temperature maps. (a) is the surface temperature at 12 km resolution with 32 × 32 pixels. (b) shows the map resampled at 1 km resolution using bi-cubic interpolation. (c) is the sea surface temperature map (white represents land). (d) is the resulting temperature map after prepossessing is complete.

2.4.7. Resampling TMP-2m and DPT-2m

To generate the variables in group 5, which contains SST, DPT2-METERS-SST, and TMP2-METERS-SST, two further preprocessing steps are needed to account for the different resolutions of the products and the absence of SSTs on land. The resolution of the MUR SST is 1 km with 384 × 384 cells and the resolution of DPT2-METERS and TMP2-METERS is 12 km with 32 × 32 cells. First, we resample DPT2-METERS, TMP2-METERS and Surface TMP from 32 × 32 to 384 × 384 and second, missing values in the land areas of the SST maps are filled with Surface TMP. Surface TMP is the NAM skin temperature, which is appropriate when comparing to SST from satellite retrievals (Youlong Xia, personal communication), such as the MUR SST. Bicubic interpolation is used to resample and artificially increase the resolution of DPT2-METERS, TMP2-METERS and Surface TMP to 1 km to match the MUR SST. For filling the land areas, the corresponding latitude and longitude point in the surface TMP maps is selected. For the last step, DPT2-METERS-SST, and TMP2-METERS-SST have been generated by taking the difference between the two maps. Fig. 7 shows an example of these preprocessing steps and the resulting feature map for that example.

2.5. Baseline method-HREF

The CONUS High-Resolution Ensemble Forecast (HREF) is a multi-model ensemble of 8 separate NWP model deterministic runs taken from the "High-Resolution Window" (HiresW) over the CONUS, and from the NAM CONUS nest (Environmental Modeling Center, 2020a). The HiresW consists of the 3.2 km versions of the NEMS–NMMB and WRF-ARW modeling systems, and the NAM CONUS nest is a higher resolution (3 km) nest within the 12 km NAM. To clarify, the current 12 km NAM runs the NEMS–NMMB. However, the HiresW NEMS–NMMB uses the 13 km RAP as the initial condition (IC). The NEMS–NMMB is the NOAA Environmental Modeling System (architecture model framework) Nonhydrostatic Multiscale Mesoscale Model (dynamic core) on the Arakawa B grid (Janjic & Gall, 2012). The WRF-ARW is the Weather Research and Forecasting (architecture model framework) Advanced Research WRF (dynamic core) (Skamarock et al., 2008). The RAP (Rapid Refresh) is an hourly updated combination assimilation and modeling system which uses the ARW as the dynamic core (Benjamin et al., 2016).

The HREF uses the two most recent HiresW WRF-ARW runs, the two most recent HiresW NEMS–NMMB runs, the two most recent HiresW "NSSL-like" WRF-ARW runs, and the two most recent NAM CONUS nest runs. The "NSSL-like" WRF-ARW runs are similar to the WRF-ARW runs, except a different planetary boundary layer (PBL) parameterization scheme, IC, and lateral boundary condition (LBC) are used, resulting in a WRF-ARW configuration similar to that run at the National Severe Storms Laboratory (University Corporation for Atmospheric Research, 2021). Each ensemble differs with respect to model dynamic core, numerics, IC, LBC, physics (PBL and microphysics parameterizations), and/or cycle. Details of the 8 ensemble members are provided in Table 3. The HREF is considered a state-of-the-art

Table 3
HREF ensemble members (November 2017). MYJ refers to the Mellor–Yamada–Janjic PBL scheme (Janić, 2001; Janjić, 1994). YSU refers to the Yonsei University PBL scheme (Hong, Noh, & Dudhia, 2006). WSM6 refers to the WRF single-moment 6-class microphysics scheme (Grasso et al., 2014). Ferrier–Aligo refers to the Ferrier–Aligo microphysics scheme (Aligo, Ferrier, Carley, Rogers, Pyle, Weiss, & Jirak, 2014).

Member	Grid-spacing	IC/LBC	PBL	Microphysics	Vertical Levels
HiresW "NSSL-like" ARW	3.2 km	NAM/NAM -6 h	MYJ	WSM6	40
HiresW "NSSL-like" ARW -12 h	3.2 km	NAM/NAM -6 h	MYJ	WSM6	40
HiresW ARW	3.2 km	RAP/GFS -6 h	YSU	WSM6	50
HiresW ARW -12 h	3.2 km	RAP/GFS -6 h	YSU	WSM6	50
HiresW NMMB	3.2 km	RAP/GFS -6 h	MYJ	Ferrier–Aligo	50
HiresW NMMB -12 h	3.2 km	RAP/GFS -6 h	MYJ	Ferrier–Aligo	60
NAM CONUS nest	3 km	NAM/NAM	MYJ	Ferrier–Aligo	60
NAM CONUS nest -12 h	3 km	NAM/NAM	MYJ	Ferrier–Aligo	60

Table 4
Confusion matrix for calculating scalar performance metrics.

Forecast	Yes	No	Total
Yes	a(hit)	b(false alarm)	a+b
No	c(miss)	d(correct rejection)	c+d
Total	a+c	b+d	a+b+c+d = n

method to predict fog. Section 3 details the performance comparisons between FogNet and optimized HREF output.

3. Experimental results and discussion

This section presents and discusses the experimental results while comparing FogNet predictions with those of a current state-of-the-art operational model.

3.1. Results for 6, 12 and 24 h lead time prediction

FogNet was calibrated, tested and optimized for overlapping binary classifiers to predict visibility ≤ 1600 m, ≤ 3200 m, ≤ 6400 m and > 6400 m. The predictions are for 6, 12, and 24 hr lead times (9 classifiers total). Calibration and testing of the models are based on 8789 visibility measurements and their related predictors (2009–2017) with 145 cases for ≤ 1600 m, 201 for ≤ 3200 m, and 483 for ≤ 6400 m. In addition to the data used for calibration and testing, an additional independent dataset (2018–2020) with 2229 cases was set aside to further evaluate the model and compare its performance with the HREF operational model predictions. The independent dataset consisted of 67 cases with visibility ≤ 1600 m, 94 ≤ 3200 m, and 172 ≤ 6400 m. Since the output of FogNet is continuous, an optimal threshold was determined for each binary classification to maximize performance based on the HSS skilled-based metric. The HREF predictions consist of probabilities that visibility values will be equivalent to or less than the categories mentioned in the Target section. The HREF probabilistic output was transformed to binary predictions by determining the probability threshold that maximized the HSS skilled-based metric as well. While FogNet and HREF predictions are compared for the independent dataset, FogNet thresholds were determined based on the calibration dataset while the HREF thresholds were determined based on the independent testing set (HREF output was not available for the same period as that of the NAM) giving a potential advantage to the HREF performance. (See Tables 6–8.)

The performance of FogNet is analyzed based on the 9 trained classifiers applied to the independent dataset (2018–2020). The formulation of the confusion matrix and performance metrics for the FogNet and HREF-based predictions are shown in Tables 4 and 5, respectively. The metrics include the pierce skill score (PSS), critical success index (CSI), Heidke skill score (HSS), odds-ratio skill score (ORSS), among others. According to Hogan, Ferro, Jolliffe, and Stephenson (2010), the performance metrics PSS and HSS are truly equitable. Equitability requires that the performance metric equals zero for constant or random forecasts (forecasts devoid of skill.) Hence, reliance on PSS and HSS is warranted when comparing the skill between FogNet and the HREF binary classifiers. The other metrics are shown for completeness. See Jolliffe and Stephenson (2003) and Wilks (2011) for a detailed explanation of all metrics. Tables 6–8 depict the performance results of all 9 FogNet and HREF binary classifiers. Both models are in general quite skillful and a more specific comparison of their respective performance will be discussed at the end of this section.

The performance of FogNet for 6 h predictions is better than the performance for 12 h and 24 h predictions for the skill metrics PSS and HSS and for most of the other metrics Table 6. FogNet PSS and HSS averaged for the three distances are 0.62 and 0.60 respectively for 6 h predictions while the same metrics are in the range of 0.51 to 0.56 for 12 h Table 7 and 24 h Table 8 predictions. Similarly POD is on average 0.65 for 6 predictions as compared to 0.59 for the longer lead times. While the performance of 6 h predictions is significantly higher for all metrics as compared to 12 h and 24 h, the performance for 12 h and 24 h predictions are very similar. The predictive skill of a NWP model (from which a majority of the FogNet features are derived) generally decreases with time, which likely explains the more skillful 6 h FogNet predictions. It is unclear why the 12 h and 24 h FogNet predictions have similar skill. Possibilities include the ability to minimize systematic errors in NWP model output by post-processing with ML (Introduction section), and the ability of FogNet to identify the salient features responsible for fog at longer lead times. For the HREF predictions, performance generally decreases with lead time.

When averaging predictions over lead times, 6 h, 12 h and 24 h and comparing the respective performances for different visibility ranges, FogNet performs better when predicting if visibility will be ≤ 6400 m as compared to ≤ 3200 m and ≤ 1600 m except for the metrics F and for a virtual tie, 0.96–0.97, for ORSS. This improvement is likely in part due to the larger number of cases, 483, for ≤ 6400 m, as compared to 201 for ≤ 3200 m and 145 cases for ≤ 1600 m as compared to the total number of cases, 8789, in the calibration dataset. Meteorological conditions leading to visibility ≤ 6400 m include all forcings leading to reduced visibility including advection and radiation fog which may facilitate the clustering in the two categories. The performance for predicting visibility ≤ 1600 m (145) is higher than for ≤ 3200 m predictions although by small margins, 0.01 or a tie for five out of the eight metrics. A possible explanation is that the increase in the number of below threshold cases when increasing the visibility range from ≤ 1600 m (145) to ≤ 3200 m (201) is relatively small while the meteorological conditions leading to the dense fog conditions quantified by a visibility of ≤ 1600 m may be more straightforward to cluster as compared to a broader range of conditions leading to a visibility of ≤ 3200 m. The average performance for the HREF predictions when averaged over lead times is generally better as the visibility range increases including for the PSS and HSS metrics increasing from 0.30 to 0.47 to 0.48 and 0.33 to 0.34 to 0.42 for ≤ 1600 m, ≤ 3200 m and ≤ 6400 m. The differences are not as clear for some of the other metrics with the highest CSS and ORSS obtained for ≤ 1600 m.

Table 5
Evaluation metrics used in our analysis.

Performance metric	Symbol	Equation
Probability of detection [0, 1]	POD	a/(a+c)
False alarm rate [0, 1]	F	b/(b+d)
False alarm ratio [0, 1]	FAR	b/(a+b)
Critical success index [0, 1]	CSI	a/(a+b+c)
Peirce skill score [−1, 1]	PSS	(ad−bc)/(b+d)(a+c)
Heidke skill score [−1, 1]	HSS	2(ad−bc)/ [(a+c)(c+d) + (a+b)(b+d)]
Odds-ratio skill score [−1, 1]	ORSS	(ad−bc)/(ad+bc)
Clayton skill score[−1, 1]	CSS	(ad−bc)/(a+b)(c+d)

Table 6
Results of FogNet versus HREF for 6 h lead-time predictions.

	≤ 1600 m		≤ 3200 m		≤ 6400 m	
Metrics	FogNet	HREF	FogNet	HREF	FogNet	HREF
POD	0.61	0.36	0.56	0.54	0.77	0.60
F	0.01	0.01	0.01	0.03	0.04	0.05
FAR	0.40	0.64	0.36	0.62	0.42	0.55
CSI	0.38	0.27	0.43	0.29	0.50	0.34
PSS	0.59	0.35	0.55	0.51	0.73	0.54
HSS	0.59	0.42	0.58	0.42	0.63	0.46
ORSS	0.98	0.96	0.98	0.94	0.97	0.92
CSS	0.58	0.51	0.62	0.36	0.56	0.42

Table 7
Results of FogNet versus HREF for 12 h lead-time predictions.

	≤ 1600 m		≤ 3200 m		≤ 6400 m	
Metrics	FogNet	HREF	FogNet	HREF	FogNet	HREF
POD	0.60	0.25	0.48	0.59	0.70	0.52
F	0.02	0.005	0.02	0.06	0.04	0.06
FAR	0.56	0.43	0.56	0.72	0.43	0.61
CSI	0.34	0.21	0.30	0.23	0.46	0.29
PSS	0.57	0.25	0.45	0.53	0.64	0.46
HSS	0.49	0.34	0.43	0.34	0.60	0.40
ORSS	0.96	0.97	0.94	0.91	0.96	0.88
CSS	0.42	0.55	0.41	0.26	0.54	0.35

Table 8
Results of FogNet versus HREF for 24 h lead-time predictions.

	≤ 1600 m		≤ 3200 m		≤ 6400 m	
Metrics	FogNet	HREF	FogNet	HREF	FogNet	HREF
POD	0.54	0.34	0.57	0.40	0.67	0.52
F	0.02	0.04	0.03	0.05	0.04	0.06
FAR	0.50	0.80	0.58	0.76	0.41	0.61
CSI	0.35	0.15	0.32	0.18	0.45	0.28
PSS	0.52	0.30	0.54	0.37	0.63	0.45
HSS	0.50	0.23	0.46	0.27	0.59	0.40
ORSS	0.97	0.85	0.95	0.84	0.96	0.88
CSS	0.48	0.18	0.40	0.26	0.56	0.35

3.2. AUCs and ROC curves

A receiver operating characteristic (ROC) curve is a two-dimensional measure of classification performance and the area under the ROC curve (AUC) is a scalar measure of performance (Marzban, 2004). Specifically, to assess the quality of forecasting problems, ROC curve is a parametric plot of the false alarm rate (F, also known as probability of false detection) on the x-axis, versus the hit-rate (also known as probability of detection, POD) on the y-axis where the decision threshold is varied across the full range of a continuous forecast quantity (Marzban, 2004). The area under the ROC curve (AUC) is a scalar measure (Hanley & McNeil, 1982), where AUC = 0.5 represents zero predictive skill and AUC =1 implies perfect skill (Jolliffe & Stephenson, 2003). To evaluate the performance of FogNet, for 6, 12 and 24 h lead times, three ROC curves are used, shown in Fig. 9a, 9b, and 9c, respectively. The AUC values for 6, 12, and 24 h lead time predictions are all substantially above 0.5, ≈ 0.86, ≈ 0.88, and ≈ 0.89 respectively, indicative of the skillful performance for FogNet.

3.3. Monthly performance of FogNet

The number of example cases varies between 124 for the month of September to 224 for the month of December (Fig. 8). Up to four predictions (cases) per day are possible if corresponding visibility measurements exist for each prediction. If there are missing measurements we remove that example instead of using a data imputation method. For the two years of the independent dataset, fog cases start in the month of October with 1, increase to a peak of 21 in February and end in April with 10. The majority of the cases take place during the period of February through April with hits larger than both misses and false alarms for all months and for all lead times except for a lead time of 24 h for the month of April. This is in contrast with the start of the first part of the fog season, the months of October through January. The October case was not predicted and false alarms and misses are higher than hits for these months and lead times except for the 12 h lead time cases in January for which hits are lower than false alarms but higher than misses. For the months of November and December 6 out of 8 cases were of radiation and advection–radiation fog; categories of fog which FogNet does not predict as well yet (see Section 3.4).

3.4. Performance by fog type

The performance of FogNet can be further analyzed by differentiating between the different types of fog. While advection fog was the dominant fog type affecting the target in the testing set (55/68), cases of radiation (7), advection–radiation (3), and frontal (1) fog were also present in the dataset. Two (2) cases were classified as unknown. Differentiation between the modes of fog formation were determined by matching the atmospheric conditions at the time of the KRAS visibility measurement to the corresponding fog type. The atmospheric conditions were assessed based on the nearest in time 00 UTC and 12 UTC SkewT-logp thermodynamic diagrams at the Corpus Christi International Airport (approximately 41 km west of the target), and from official products issued by NWS operational forecasters. (The detailed analysis necessary to identify stratus-lowering fog was not performed). While there was only one frontal fog case, the ten cases of radiation and advection–radiation fog allow for a differentiate analysis. For lead times of 6, 12, and 24 hours, only 1, 2, and 1 out of 10 radiation and advection–radiation fog events were predicted, respectively. This is considerably below the hit rates of 37, 36, and 37 out of 55 cases of advection fog formation. The ratio of different types of fog formation are similar in the calibration and independent testing sets. It appears that at present, FogNet focuses its calibration on the dominant fog formation mode and does not recognize the processes responsible for radiation and advection–radiation fog formation as well. Methods such as oversampling radiation and advection–radiation fog cases in the calibration set, and further modifying the FogNet architecture, could further enhance FogNet performance and are left for future work.

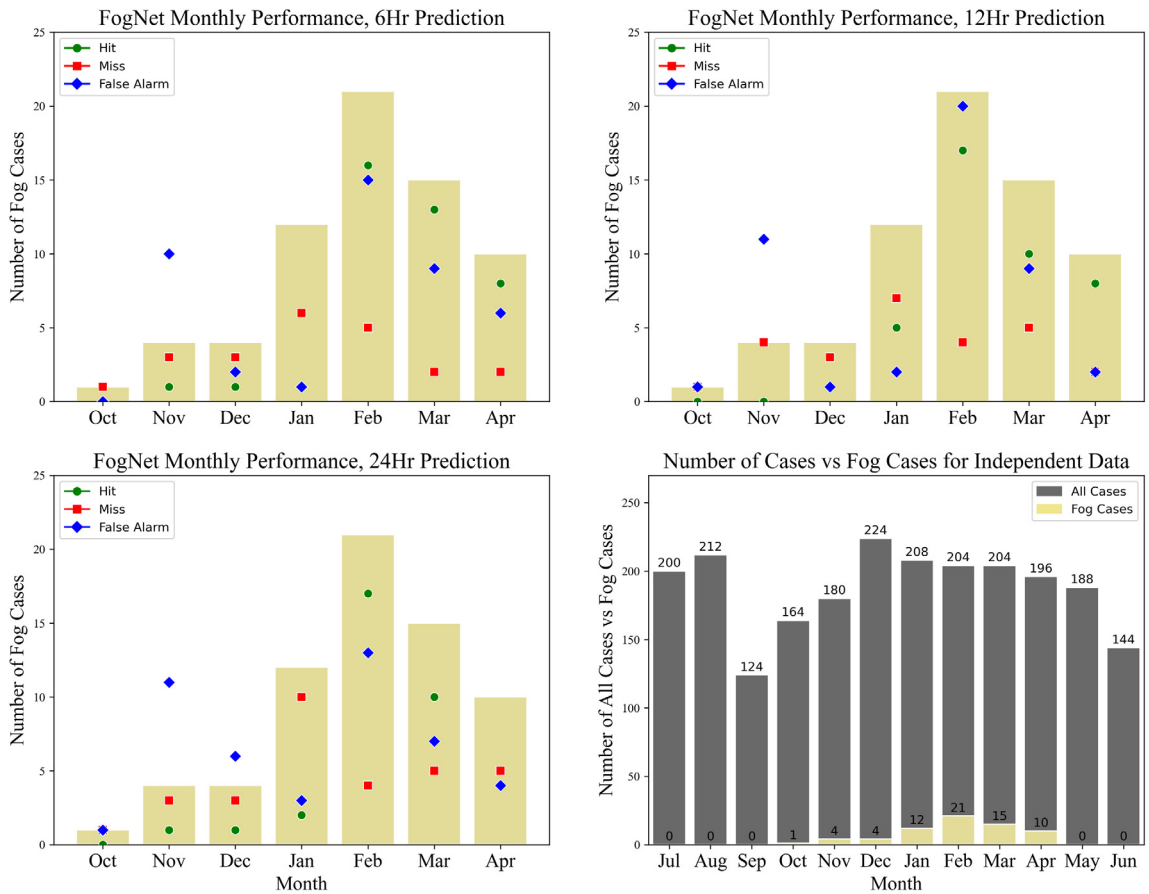

Fig. 8. Monthly performance of FogNet for < 1600 m and 6-, 12- and 24 h lead time prediction and the number of cases by month.

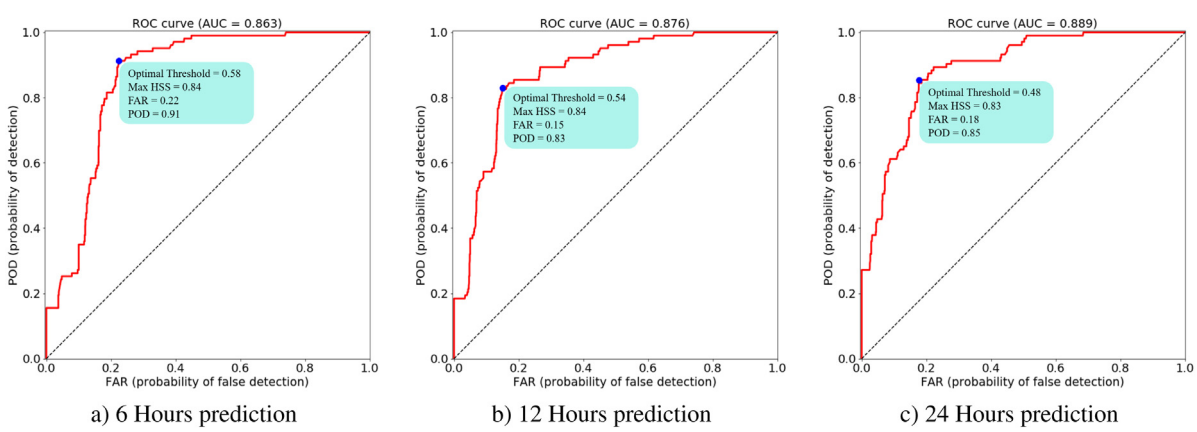

Fig. 9. ROC curve and AUC.

We included features that account for the vertical profile of the atmosphere from the surface (2 or 10-meters) to 700-mb. However, conditions at the 1000-mb level were not included as features in FogNet since the 1000-mb pressure level can sometimes fall below the material surface at KRAS. However, in the radiation and advection–radiation fog cases, the temperature, wind, relative humidity, and mixing ratio profile can change significantly in the lowest 600 meters (Liu et al., 2011). Thus, the inclusion of 1000-mb (sans cases where the 1000-mb touches or falls below the material surface) may provide the additional vertical resolution necessary to improve radiation and advection–radiation fog prediction. Radiation fog prediction may also be improved by adding mean sea level pressure (MSLP) as a feature, since radiation fog tends to occur in the proximity of the surface anticyclone (Gultepe et al., 2007). Further, the effect of aerosols on fog formation is not accounted for in FogNet. However, owing to the first indirect effect (Twomey, 1974) mentioned earlier, the inclusion of aerosol concentration may improve the skill of FogNet with respect to the prediction of all types of fog. Research is ongoing and we will consider adding these additional features in future research.

3.5. Performance comparison of FogNet and HREF

With respect to the skill-based metrics PSS and HSS, all FogNet predictions outperformed the corresponding HREF classifiers (except for the 12 h predictions of category $\leq$ 3200 m with respect to PSS.) When averaging the performance metrics over the three lead times and the three visibility ranges, the differences are substantial. For example averaged PSSs are 0.58 vs 0.42, averaged HSSs are 0.54 vs 0.36 and average PODs are 0.61 vs 0.46 for FogNet vs HREF. Importantly, FogNet outperformed the HREF with respect to the category that most adversely affects marine and aviation transportation, namely, $\leq$ 1600 m. The NWS issues marine dense fog advisories when widespread fog of visibility $\leq$ 1852 m (1 nautical mile) is occurring, imminent, or is deemed to have a high probability of occurrence. Fog in this category precludes maritime travel in and near the Port of Corpus Christi. Further, visibilities below 805 m (0.5 statute mile) halts airline traffic to the Corpus Christi International Airport.

These results illustrate that when developing fog prediction models, the use of ML to post-process single deterministic NWP model output can result in models that outperform NWP model ensemble output. This is remarkable when considering the difference in computing resource requirements. The HREF requires 8 separate and unique deterministic NWP model runs (Table 2) while FogNet utilizes output from only one deterministic NWP model (the NAM in this study.) Further, FogNet performed superior to the HREF even though the horizontal resolution of the HREF members (3 km grid spacing) is higher than that of the NAM (12 km grid spacing and the predominate source of the FogNet features), which allows the HREF members to more accurately resolve moisture, temperature, and wind variations in the meso-γ scale (2–20 km; Orlanski, 1975) that influence fog development. Finally, the superior performance of FogNet over the HREF is remarkable when considering the advantage of the HREF over FogNet during optimization. Recall that HREF probabilistic output was converted to categorical by determining the threshold that optimized performance on the testing set, while FogNet optimization was restricted to the training set, thus giving the HREF an advantage when evaluating performance on the testing set.

4. Conclusions

A 3D-CNN, FogNet, was designed and tested to predict the occurrence of coastal fog, more specifically visibility below 1600 m, 3200 m, 6400 m with lead times of six to twenty four hours. The 3D-CNN combines numerical weather predictions from the North American Mesoscale Forecast System (NAM) and the latest sea surface temperatures measured by the NASA Multiscale Ultra Resolution (MUR) dataset. The selected features are organized into a 288, 384 and 384 layer data cube for 6, 12 and 24 h lead time predictions with a grid spacing of 12 km for the 32 × 32 horizontal grid. The study covers the period 2009–2020. The target data are visibility measurements from a coastal airport located on a barrier island along the ship channel of the fourth largest US port by tonnage. The 3D-CNN is designed to extract information from the spatial patterns of the respective variables but also the correlations within the vertical profiles of the meteorological variables and between the different variables including the interaction between SST and surface atmospheric features such as air temperature and dew point. A dense block and attention mechanism was used to model and reinforce more specifically the interaction between model features and altitude. The performance of FogNet exceeded the performance of the operational HREF ensemble based on 8 standard evaluation metrics, for nearly all HREF visibility categories including for PSS, 0.58 vs 0.42, and HSS 0.54 vs 0.36 when averaging these metrics over all visibility ranges and lead times. While this study did not investigate predictions for lead times longer than 24 h, the superior performance of FogNet over HREF at 24 h makes it likely that FogNet will continue to outperform HREF for longer lead times and warrants a future study to determine the maximum lead time FogNet performance exceeds that of the HREF. The parallel architecture of FogNet facilitates the expansion of the number of predictors as the lead time increases. Further work is needed to better distinguish the relative importance of the inputs and architecture characteristics that are the most important for this significant improvement in performance. It is hoped that the emerging methods of Explainable AI will provide such insights and may also lead to further performance improvements.

There are several implications related to the discovery that FogNet performed superior to a state-of-the-art ensemble prediction system (EPS) used by operational meteorologists to predict fog. First, when using NWP models to predict the future state of the atmosphere and associated phenomena, the post-processing of output from a single deterministic NWP model run using DL represents a valid alternative to EPS development. As mentioned in the Introduction, an EPS was developed to account for the uncertainty in the NWP model initial condition when using NWP models to predict the future atmospheric state (minor changes to the initial condition can result in dramatic changes to the future atmospheric state owing to chaos.) The success of FogNet adds credence to the results of Pathak et al. (2018), who demonstrated that the prediction of a chaotic system can be improved by the combination of a knowledge-based model (a dynamical model that describes the chaotic system) and a ML model. FogNet involves the post-processing of NWP model (knowledge-based model) output by a CNN (ML technique) resulting in superior performance over an operational EPS. This is significant since the EPS strategy has for decades been the preferred state-of-the-art operational forecasting technique to account for atmospheric chaos. The second implication relates to efficiency: A data-driven weather prediction model, more skillful than an EPS, can be obtained by post-processing output from a single NWP deterministic model run, which requires much less computer processing power than that used in generating an EPS. The EPS benchmark used in this study (HREF) provided probabilistic forecasts of 3 visibility categories, by post-processing output from 8 separate NWP model runs, each with a higher spatial resolution (3-km grid spacing) than from the single NWP model used by FogNet (12-km grid spacing.) The improvement in fog prediction skill by post-processing output from one NWP model run over a high-resolution system that post-processed 8 times more NWP model runs is remarkable.

CRediT authorship contribution statement

Hamid Kamangir: Conceptualization, Methodology, Data curation, Visualization, Writing - original draft. **Waylon Collins:** Conceptualization, Methodology, Data curation, Writing - original draft, Writing - review & editing. **Philippe Tissot:** Conceptualization, Supervision, Writing - original draft. **Scott A. King:** Conceptualization, Visualization, Writing - review & editing. **Hue Thi Hong Dinh:** Data curation. **Niall Durham:** Data curation, Software. **James Rizzo:** Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge the assistance of Eric Rogers (U.S. National Weather Service, National Centers for Environmental Prediction, Environmental Modeling Center) (NWS–NCEP–EMC) who provided clarification of both the lack of ocean–atmosphere coupling between the 12 km NAM model atmosphere and the NAM sea surface temperature analysis, and that the NAM SST is not incorporated into the calculation of NAM surface visibility. We thank Eric Aligo (I. M. Systems Group

contractor for the NWS–NCEP–EMC) and Jacob Carley (NWS–NCEP–EMC) for insights on the extent to which aerosols and single-moment microphysics may influence visibility prediction. We thank Youlong Xia (I. M. Systems Group contractor for the NWS–NCEP–EMC) for providing insight that resulted in the decision to use NAM skin temperature as the land surface temperature variable to correspond with the MUR SST. We sincerely thank Matthew Pyle (NWS–NCEP–EMC) for providing HREF visibility probability output for the 1 December 2017 through 31 August 2019 period. And we thank Evan Krell for his contributions to the design of the figures.

This work is part of the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES). This material is based upon work supported in part by the National Science Foundation, USA under Grant No. ICER-2019758.

Appendix. North American Mesoscale (NAM) Modeling System Configuration as of March 2017 (Environmental Modeling Center, 2017)

Infrastructure	NOAA Environmental Modeling System.
Dynamics	Non-Hydrostatic Multiscale Model on Arakawa B-grid.
Horizontal Resolution	12 km grid spacing.
Vertical Levels	60 levels (27 levels in the 0–3 km layer).
Number of grid points	$X = 954$, $Y = 835$.
Map projection	Lambert Conformal.
Horizontal grid	Arakawa b-grid staggering.
Vertical grid	Lorenz staggering.
Vertical coordinate	Terrain-following hybrid sigma-pressure transitioning to pure hydrostatic pressure near the top.
Microphysics	Ferrier–Aligo scheme.
Radiation	Rapid Radiative Transfer Model (RRTM).
Cumulus parameterization	Betts–Miller–Janjic (BMJ) scheme.
Planetary boundary layer parameterization	Mellor–Yamada–Janjic (MYJ) scheme.
Land surface parameterization	Noah Land Surface Model.
Sea surface temperature	NCEP 0.125 deg (RTG-SST-HR).
Lateral boundary condition	Global Forecast System (GFS).
Data assimilation method	Hybrid ensemble-3DVar incorporated into the NCEP Grid point Statistical Interpolation (GSI) Analysis.
Initialization	Diabatic Digital Filter.

References

AirNav (2020). KRAS Mustang Beach Airport in Port Aransas, Texas, USA. URL: https://www.airnav.com/airport/KRAS. (Accessed 16 February 2020).

Aligo, E. A., Ferrier, B., & Carley, J. R. (2018). Modified NAM microphysics for forecasts of deep convective storms. *Monthly Weather Review*, *146*, 4115–4153.

Aligo, E., Ferrier, B., Carley, J., Rogers, E., Pyle, M., Weiss, S., et al. (2014). Modified microphysics for use in high-resolution nam forecasts. In *27th Conf. on Severe Local Storms*.

Alipour-Fard, T., Paoletti, M., Haut, J. M., Arefi, H., Plaza, J., & Plaza, A. (2020). Multibranch selective kernel networks for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*.

Arthur, E. H., & Robert, W. K. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Bari, D., Bergot, T., & Khlifi, M. (2016). Local meteorological and large-scale weather characteristics of fog over the grand casablanca region, morocco. *Journal of Applied Meteorology and Climatology*, *55*, 1731–1745.

Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., et al. (2016). A North American hourly assimilation and model forecast cycle: The rapid refresh. *Monthly Weather Review*, *144*(4), 1669–1694.

Cao, F., & Guo, W. (2020). Cascaded dual-scale crossover network for hyperspectral image classification. *Knowledge-Based Systems*, *189*, Article 105122.

Chaabani, H., Werghi, N., Kamoun, F., Taha, B., Outay, F., et al. (2018). Estimating meteorological visibility range under foggy weather conditions: A deep learning approach. *Procedia Computer Science*, *141*, 478–483.

Chin, T. M., Vazquez-Cuervo, J., & Armstrong, E. M. (2017). A multi-scale high-resolution analysis of global sea surface temperature. *Remote Sensing of Environment*, *200*, 154–169.

Chollet, F., et al. (2018). *Deep learning with Python. volume 361*. Manning New York.

Dupont, J. C., Haeffelin, M., Stolaki, S., & Elias, T. (2016). Analysis of dynamical and thermal processes driving fog and quasi-fog life cycles using the 2010-2013 ParisFog dataset. *Pure and Applied Geophysics*, *173*, 1337–1358.

Durán-Rosal, A. M., Fernández, J. C., Casanova-Mateo, C., Sanz-Justo, J., Salcedo-Sanz, S., & Hervás-Martínez, C. (2018). Efficient fog prediction with multi-objective evolutionary neural networks. *Applied Soft Computing*, *70*, 347–358.

Environmental Modeling Center (2017). North American Mesoscale (NAM) analysis and forecast system characteristics. URL https://www.emc.ncep.noaa.gov/mmb/mmbpll/misc/NAM_2017.pdf (accessed 9 May 2021).

Environmental Modeling Center (2020a). High resolution ensemble forecast. URL https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/href-hiresw.php (Accessed 16 February 2020).

Environmental Modeling Center (2020b). On-line documentation of EMC Mesoscale model/analysis system changes. URL https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/nam/references.php (accessed 9 May 2021).

Environmental Modeling Center (2021). North American Mesoscale. URL https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/nam.php (accessed 9 May 2021).

Fabbian, D., De Dear, R., & Lellyett, S. (2007). Application of artificial neural network forecasts to predict fog at Canberra International Airport. *Weather and Forecasting*, *22*(2), 372–381.

Glickman, T. S. (2000). *Glossary of meteorology*. American Meteorological Society.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Grasso, L., Lindsey, D. T., Sunny Lim, K. -S., Clark, A., Bikos, D., & Dembek, S. R. (2014). Evaluation of and suggested improvements to the WSM6 microphysics in WRF-ARW using synthetic and observed GOES-13 imagery. *Monthly Weather Review*, *142*(10), 3635–3650.

Gultepe, I., & Isaac, G. (1999). Scale effects on averaging of cloud droplet and aerosol number concentrations: Observations and models. *Journal of Climate*, *12*(5), 1268–1279.

Gultepe, I., Milbrandt, J. A., & Zhou, B. (2017). Marine fog: A review on microphysics and visibility prediction. In *Marine fog: challenges and advancements in observations, modeling, and forecasting* (pp. 345–394). Springer.

Gultepe, I., Tardif, R., Michaelides, S., Cermak, J., Bott, A., Bendix, J., et al. (2007). Fog research: A review of past achievements and future perspectives. *Pure and Applied Geophysics*, *164*(6), 1121–1159.

Hacker, J. P., & Rife, D. L. (2007). A practical approach to sequential estimation of systematic error on near-surface mesoscale grids. *Weather and Forecasting*, *22*(6), 1257–1273.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29–36.

He, M., Li, B., & Chen, H. (2017). Multi-scale 3D deep convolutional neural network for hyperspectral image classification. In *2017 IEEE international conference on image processing* (pp. 3904–3908). IEEE.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580.

Hogan, R. J., Ferro, C. A., Jolliffe, I. T., & Stephenson, D. B. (2010). Equitability revisited: Why the "equitable threat score" is not equitable. *Weather and Forecasting*, *25*(2), 710–726.

Hong, S. -Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*(9), 2318–2341.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700–4708).

Janić, Z. I. (2001). Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso model.

Janjić, Z. I. (1994). The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Monthly Weather Review*, *122*(5), 927–945.

Janjic, Z., & Gall, R. (2012). *Scientific documentation of the NCEP nonhydrostatic multiscale model on the B grid (NMMB)*. Part 1 Dynamics (No. NCAR/TN-489+STR), http://dx.doi.org/10.5065/D6WH2MZX.

Janjic, Z. I., Gerrity, J., Jr., & Nickovic, S. (2001). An alternative approach to nonhydrostatic modeling. *Monthly Weather Review*, *129*(5), 1164–1178.

Jolliffe, I., & Stephenson, D. (2003). *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Chichester: John Wiley and Sons.

Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability*. Cambridge university press.

Kamangir, H., Collins, W., Tissot, P., & King, S. A. (2020). Deep-learning model used to predict thunderstorms within 400 km2 of south Texas domains. *Meteorological Applications*, *27*(2).

Khaefi, M. R., Pramestri, Z., Amin, I., & Lee, J. G. (2018). Nowcasting air quality by fusing insights from meteorological data, satellite imagery and social media images using deep learning. In *2018 IEEE/ACMinternational conference on advances in social networks analysis and mining* (pp. 393–396). IEEE.

Kingma, D., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv: 14126980. Retrieved from the arXiv database.

Kipfer, K. (2017). *Fog Prediction with Deep Neural Networks*. ETH Zurich.

Kleiber, W., Raftery, A. E., Baars, J., Gneiting, T., Mass, C. F., & Grimit, E. (2011). Locally calibrated probabilistic temperature forecasting using geostatistical model averaging and local Bayesian model averaging. *Monthly Weather Review*, *139*(8), 2630–2649.

Koračin, D., Dorman, C. E., Lewis, J. M., Hudson, J. G., Wilcox, E. M., & Torregrosa, A. (2014). Marine fog: A review. *Atmospheric Research*, *143*, 142–175.

Koschmeider, H. (1924). Therie der horizontalen sichtweite. *Beitraege zur Physik der Atmosphaere*, *12*, 33–53.

Kunkel, B. A. (1984). Parameterization of droplet terminal velocity and extinction coefficient in fog models. *Journal of Climate and Applied Meteorology*, *23*(1), 34–41.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Lee, Y. H., Lee, J. -S., Park, S. K., Chang, D. -E., & Lee, H. -S. (2010). Temporal and spatial characteristics of fog occurrence over the Korean Peninsula. *Journal of Geophysical Research: Atmospheres*, *115*(D14).

Leith, C. (1974). Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review*, *102*(6), 409–418.

Li, S., Fu, H., & Lo, W. -L. (2017). Meteorological visibility evaluation on webcam weather image using deep learning features. *International Journal of Computer Theory and Engineering*, *20017*(9), 455–461.

Li, P., Wang, G., Fu, G., & Lu, C. (2016). On spatiotemporal characteristics of sea fog occurrence over the Northern Atlantic from 1909 to 2008. *Journal of Ocean University of China*, *15*(6), 958–966.

Liu, D., Yang, J., Niu, S., & Li, Z. (2011). On the evolution and structure of a radiation fog event in Nanjing. *Advances in Atmospheric Sciences*, *28*(1), 223–237.

Lorenz, E. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, *20*, 130–141.

Lorenz, E. N. (1969). The predictability of a flow which possesses many scales of motion. *Tellus*, *21*(3), 289–307.

Luan, T., Guo, X., Guo, L., & Zhang, T. (2018). Quantifying the relationship between PM 2.5 concentration, visibility and planetary boundary layer height for long-lasting haze and fog–haze mixed events in Beijing. *Atmospheric Chemistry and Physics*, *18*(1), 203.

Ma, W., Yang, Q., Wu, Y., Zhao, W., & Zhang, X. (2019). Double-branch multi-attention mechanism network for hyperspectral image classification. *Remote Sensing*, *11*(11), 1307.

Marzban, C. (2004). The ROC curve and the area under it as performance measures. *Weather and Forecasting*, *19*(6), 1106–1114.

Melhauser, C., & Zhang, F. (2012). Practical and intrinsic predictability of severe and convective weather at the mesoscales. *Journal of the Atmospheric Sciences*, *69*(11), 3350–3371.

Miao, K. -c., Han, T. -t., Yao, Y. -q., Lu, H., Chen, P., Wang, B., et al. (2020). Application of LSTM for short term fog forecasting based on meteorological elements. *Neurocomputing*.

Orlanski, I. (1975). A rational subdivision of scales for atmospheric processes. *Bulletin of the American Meteorological Society*, 527–530.

Pagowski, M., Gultepe, I., & King, P. (2004). Analysis and modeling of an extremely dense fog event in southern Ontario. *Journal of Applied Meteorology*, *43*(1), 3–16.

Pathak, J., Wikner, A., Fussell, R., Chandra, S., Hunt, B. R., Girvan, M., et al. (2018). Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model. *Chaos. An Interdisciplinary Journal of Nonlinear Science*, *28*(4), Article 041101.

Prechelt, L. (1998). Early stopping-but when? In *Neural networks: Tricks of the trade* (pp. 55–69). Springer.

Roy, S. K., Krishna, G., Dubey, S. R., & Chaudhuri, B. B. (2019). Hybridsn: Exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*.

Saurabh, K., & Dimri, A. (2016). Non-linearity explanation in artificial neural network application with a case study of fog forecast over Delhi region. *Pure and Applied Geophysics*, *173*(5), 1765–1781.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Wang, W., et al. (2008). *A description of the advanced research WRF version 2 (No. NCAR/TN-468+STR)*. University Corporation for Atmospheric Research.

Song, A., & Choi, J. (2020). Fully convolutional networks with multiscale 3D filters and transfer learning for change detection in high spatial resolution satellite images. *Remote Sensing*, *12*(5), 799.

SPC (2020). Short range ensemble forecast (SREF) products. URL: http://www.spc.noaa.gov/exper/sref. (Accessed 16 February 2020).

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(1), 1929–1958.

Stensrud, D. J. (2009). *Parameterization schemes: Keys to understanding numerical weather prediction models*. Cambridge University Press.

Stoelinga, M. T., & Warner, T. T. (1999). Nonhydrostatic, mesobeta-scale model simulations of cloud ceiling and visibility for an east coast winter precipitation event. *Journal of Applied Meteorology*, *38*(4), 385–404.

Stull, R. (1988). *An introduction to boundary layer meteorology*. Kluwer Academic Publishers, Springer Science & Business Media, 666 pp.

Toth, G., Gultepe, I., Milbrandt, J., Hansen, B., Pearson, G., Fogarty, C., et al. (2010). On fog and fog forecasting.

Twomey, S. (1974). Pollution and the planetary albedo. *Atmospheric Environment*, *8*(12), 1251–1256, (1967).

University Corporation for Atmospheric Research (2021). The COMET® Program. Operational Model Encyclopedia. HREF. URL https://sites.google.com/ucar.edu/operational-models-encyclo/probabilistic-models/href (Accessed 9 May 2021).

Wang, W., Dou, S., Jiang, Z., & Sun, L. (2018). A fast dense spectral–spatial convolution network framework for hyperspectral images classification. *Remote Sensing*, *10*(7), 1068.

Warner, T. T. (2011). Numerical weather and climate prediction. *Meteorological Applications*, 526.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences. volume 100*. Academic press.

WMO (2020). Fog compared with Mis. URL: https://cloudatlas.wmo.int/en/fog-compared-with-mist.html. (Accessed 16 February 2020).

Woo, S., Park, J., Lee, J. -Y., & So Kweon, I. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision* (pp. 3–19).

Xiao, C., Chen, N., Hu, C., Wang, K., Xu, Z., Cai, Y., et al. (2019). A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. *Environmental Modelling & Software*, *120*, Article 104502.

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., et al. (2015). Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning* (pp. 2048–2057).

Xu, Q., Xiao, Y., Wang, D., & Luo, B. (2020). CSA-MSO3DCNN: Multiscale octave 3D CNN with channel and spatial attention for hyperspectral image classification. *Remote Sensing*, *12*(1), 188.

Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122.

Zazzaro, G., Pisano, F. M., & Mercogliano, P. (2010). Data mining to classify fog events by applying cost-sensitive classifier. In *2010 international conference on complex, intelligent and software intensive systems* (pp. 1093–1098). IEEE.

Zhou, B., & Du, J. (2010). Fog prediction from a multimodel mesoscale ensemble prediction system. *Weather and Forecasting*, *25*(1), 303–322.

Zhou, B., Du, J., Gultepe, I., & Dimego, G. (2012). Forecast of low visibility and fog from NCEP: Current status and efforts. *Pure and Applied Geophysics*, *169*(5–6), 895–909.

Zhou, K., Zheng, Y., Li, B., Dong, W., & Zhang, X. (2019). Forecasting different types of convective weather: A deep learning approach. *Journal of Meteorological Research*, *33*(5), 797–809.