Received: 12 March 2024 | Revised: 11 May 2024 | Accepted: 7 June 2024 | Published online: 18 July 2024

REVIEW

Journal of Computational and Cognitive Engineering
2024, Vol. 3(4) 324–347
DOI: 10.47852/bonviewJCCE42022817

BON VIEW PUBLISHING

Transfer Learning in Weather Prediction: Why, How, and What Should

Sankar Kumar Pal[1] and **Debashree Dutta**[1,*]

[1] *Center for Soft Computing Research, Indian Statistical Institute, India*

Abstract: Transfer learning (TL) is a popular phrase in deep learning (DL) domain. It is one of the latest artificial intelligence (AI) technologies that has a significant impact on big data analysis. Methods of traditional machine learning (ML) require the availability of an adequate quantity of training data as well as similarity of characteristics among the feature spaces corresponding to training and test data while performing supervised learning tasks. However, in real-life analytical problems, data scarcity often arises. In such scenarios, the TL approach has shown effectiveness in transferring knowledge from the source tasks that had large training data to a target task that has less training data. Basically, in TL, a model that has been trained on one task is essentially applied to a second related (but not exact) task. In this way, the issue of distribution mismatch can also be addressed. TL is not like conventional machine learning algorithms that try to learn each task starting from the beginning. Meteorological research is such an example of big data analysis which often faces the data scarcity issue. The current study addresses the contemporary challenges in weather forecasting that can be solved (or better dealt with) using TL methods. It presents a brief review of earlier research with the evolution of various technologies used since 1990s, followed by potential applications of TL algorithms to several key challenges in weather prediction, which includes the prediction of air quality, thunderstorms, precipitation, visibility, and cyclones, among others. Special emphasis is given to high-impact weather (HIW) prediction. These high-impact events are extremely difficult to predict, and they can cause enormous property damage and fatalities around the world. TL techniques have shown advantages in predicting HIW. Various challenging issues in implementing TL technology are then discussed. Finally, we address various prospects associated with TL, propose new research directions, and more importantly mention some concerns for beginners in DL-TL research. An extensive list of references is also provided.

Keywords: transfer learning, meteorological applications, AI, weather prediction, high-impact weather, deep learning, climate analytics

1. Introduction

Forecasting of weather and climate has been crucial for humanity. The significance of weather forecasting lies in its ability to predict future climatic expectations. Transitioning from personal planning to extensive commercial development, agricultural activity, construction, and infrastructure development, weather forecasting is an essential tool that facilitates numerous aspects of human life and societal activities. The weather in any given area or location is a major factor as it can significantly impact crop productivity throughout various stages of growth and development. Weather variability throughout the crop season, such as onset of monsoon, early or delayed monsoon, heavy rainfall or flood condition, insufficient rainfall or drought condition, heat waves, and cold waves, can impact on crop production. This in turn enhances agricultural productivity by decreasing the risks and losses and enhancing water usage efficiency. Prediction of weather assists people specially farmers and communities in planning for and responding to aforementioned weather-related occurrences. In addition to minimizing property damage, this can assist save lives. Precise weather forecasts are also necessary for decision-making in sectors including energy, transportation, water, agriculture, and emergency response. Weather predictions are also essential in aviation and marine operations, assisting pilots and ship captains in planning safe routes. In general, sustainable environmental practices, economic stability, and public safety are enhanced by precise weather forecasting. But there is still a significant obstacle for scientists, particularly in tropical areas for accurately predicting meteorological phenomena. Thus, an accurate weather forecast is a crucial adaptation measure as it is necessary to protect lives and livelihoods from occurrences of high-impact weather (HIW).

HIW signifies extreme weather conditions that have the potential to cause significant disruption or damage to property and life. Although we are unable to control these occurrences, we may strengthen our resistance to these HIW by becoming more adept at anticipating and improving the prediction system to forecast such kinds of phenomena. Weather prediction or forecasting involves predicting the state of the land, ocean, and atmosphere, by analyzing the change in meteorological variables for a specific location and time. This process involves collection of quantitative data, a preprocessing framework to address the quality issue of data and analysis of data. For precise forecasts, meteorologists use atmospheric models. These are a series of formulas that accurately depict the condition of the atmosphere. The actual state of the atmosphere is determined by combining the data obtained from the models with the information gathered from various weather stations.

The present article deals with a state-of-the-art review on weather prediction, with emphasis on HIW, in the framework of

***Corresponding author:** Debashree Dutta, Center for Soft Computing Research, Indian Statistical Institute, India. Email: debashree120_t@isical.ac.in

transfer learning (TL), which is an emerging technology in artificial intelligence (AI). Before we mention the scope and novelty of the paper, we explain in brief the evolution of various technologies including AI and TL since 19th century as used in weather prediction and climate analytics problems.

The advent of modern technologies like the electronic computer, remote sensing, and telegraph has completely changed the platform of weather prediction. Plotting weather charts was made possible by the telegraph in the middle of the 19th century, as it allowed for the rapid transmission of weather observation data across great distances. A century or so later, Charney et al. [1] used the first general-purpose electronic computer, the ENIAC, to successfully perform a computer simulation of the weather. Using an IBM 701, one of the first mainframe computers, the U.S. Joint Numerical Weather Prediction Unit began operational numerical weather prediction (NWP) in 1955. Since then, the computer industry has experienced fast growth, and NWP's use of high-performance computers (HPCs) has emerged as a critical technology for weather forecasting. Satellite remote sensing was made possible in the 1970s by advancements in space exploration technology. Since several weather satellites have been launched, it was imperative to update NWP as quickly as possible with this new information. The current weather forecast mostly uses satellite data and NWP with HPCs [2].

In meteorology, statistical techniques are used as the primary method of weather prediction before the development of computational models. Statistical prediction techniques are used in data-driven weather predictions, both short-term and long-term processes. Traditional statistical methods rely on the linear assumption that is inconsistent with the nonlinear characteristics of reality, thereby having some consequences. When it comes to the task of meteorological forecasting, physical models are often found to be more accurate than statistical ones. Physical models are based on a fundamental set of concepts from physics, including Newton's laws of motion, the conservation of mass and energy, and the laws of thermodynamics. In meteorology, such physical models are particularly based on NWP and climate modeling. The forecast of NWP ranges from local to regional and global with a timescale of nowcasting to long-range weather forecasting. Weather forecasting nowadays uses the NWP. While aiming for extremely precise meteorological forecasts over a period of days or weeks, NWP is mostly concerned with short-to-medium-range weather prediction. Though NWP provides greater accuracy and reliability, one of the significant issues is the initial state of the atmosphere, which is not fully understood. It faces challenges such as errors due to the chaotic nature of the atmosphere. Thus, the accuracy of the prediction decreases and the time gap between actual and predicted enhances. Further, there are inherent diversity and uncertainties present in the dataset limiting the performance of NWP. Finally, the theory-based nonlinear equations that are discussed can heavily depend on supercomputer performance. Therefore, conventional methods have some limitations, highlighting the requirement for novel approaches.

Around 2010, "Big data" drew the attention of applied researchers and started revolutionizing the weather prediction once more. Big data is mainly characterized by four Vs, viz., large volume, velocity, variety, and veracity. Since huge weather data are available, meteorologists got engaged in using artificial intelligence (AI) and machine learning (ML) techniques to predict weather events. The predictions of these models are seen to be faster than those of physical models and their outcomes are more precise than those of statistical models. Besides, ML (learning the patterns from previous data or examples) can be utilized for downscaling and correction of errors to improve weather and climate prediction. Thus, AI-ML has great potential for improving prediction results and providing us with a thorough understanding of terrestrial, marine, and atmospheric processes.

AI research deals with developing intelligent systems using a set of algorithms that attempt to mimic human intelligence. ML, a component (subset) of AI, deals with the tasks in which a computer system or a machine can learn patterns from data and generate predictions without the need for human intervention. These predictions can be produced either by unsupervised learning, in which algorithms learn patterns from unlabeled data, or supervised learning, in which algorithms discover general patterns in labeled data. In real-life problems, data (or examples) may be labeled, unlabeled, or both. Although the supervised learning techniques are mainly used for regression and data classification for a range of algorithms including random forest, k-nearest neighbors, decision trees, support vector machines, and linear regression, they can also be effective for data categorization. Unsupervised learning uses unlabeled datasets, where the learning is mostly based on clustering algorithms that categorize the training data depending on their different characteristics. Artificial neural networks (ANNs) are thought to be excellent candidates for ML since they can learn the relationship between input and output from examples. ANNs enjoy characteristics like adaptivity, speed, robustness/ ruggedness, and optimality [3]. ANN models using interconnected layers are designed to mimic the function and structure of the human brain. ANN is trained with training data, and it is expected that the model will behave accurately with the actual dataset in the concerned problem domain. Multilayer perceptron (MLP) is a complex-layered architecture of feedforward ANN that uses a backpropagation algorithm for training the model. It is capable of modeling/generating nonlinear boundaries between classes depending on the number of nodes and layers.

The development of deep learning (DL) was aided by several advances in multilayered neural networks in the early 2000s. DL refers to learning in depth in different stages. This is a specialized and advanced version of ML. DL differs from conventional ML in the sense that it primarily learns the data representation, unlike the task-specific techniques as done in ML. Today, we have an abundant supply of data, so DL becomes a natural and meaningful choice. Accordingly, DL has grown in prominence, as compared to ML, in recent years as a powerful tool for accurately analyzing "Big data." Both ML and DL have proved their efficacy in various domains. These include speech recognition [4], natural language processing (NLP) [5], image classification [6], object detection and segmentation [7], video tracking [8], medical science [9, 10], classification of plant diseases [11], and forecasting stock markets [12] and natural hazards prediction [13], among others.

For accurate representation and prediction of weather phenomena, advanced AI tools based on ML-DL have recently been popular owing to the identification of nonlinear relationships and improved performance. Conventional neural network learning (shallow learning) was widely used in early attempts to simulate the physical processes in meteorology while using less computational power. Such investigations into the problem of weather forecasting include those by Chaudhuri et al. [14, 15], Dutta and Chaudhuri [16], and Bączkiewicz et al. [17]. Some researchers have used advanced DL architectures like recurrent neural network or RNN [18], or long short-term memory (LSTM) [19], convolution neural network or CNN [20], and gated recurrent units or GRU [21] to create more accurate and reliable weather prediction models to predict various weather phenomena.

Two key assumptions underpin the functioning of traditional ML, especially DL-based models are that (1) both the training and testing data are selected from the same distribution and (2) DL models would be given a significant quantity of training data to learn the latent patterns in the dataset. However, these criteria may not always be valid in real-world situations. Here comes the relevance of TL, which allows us to dissolve both the assumptions and becomes a good fit for many real-life applications that have little training data. As the name suggests, TL involves applying knowledge from previously built models, that is, information gained for one task as a foundation for another related task, hence enhancing the predictability and learning processes in the latter [22]. Accordingly, TL technology has advanced the DL models for wider applications. Using deep models to learn high-level abstract features is also seen to make the TL process considerably simpler and more reliable [23].

The basic principle behind TL is to broaden the notion of a domain and a task. The notion of a task and a domain is a fundamental component of TL. In particular, it focuses on a source domain, a target domain, a source task, and a target task. Notably, the source task is only evaluated as a supplementary assessment; the model is evaluated exclusively for the target task. This indicates that the generalization capability of the latter has been enhanced by pattern classification. TL is different from conventional ML (or shallow learning) algorithms which aim to learn a task from the beginning. In TL, the information obtained from previous tasks is transferred to a target task when the latter has less training data, i.e., data scarcity arises.

1.1. Motivation behind applying ML-DL-TL in weather prediction

For accurate prediction of weather phenomena, including HIW, long-term data records are a more reliable source of information for making decisions and studying the trend of different meteorological variables that demonstrate the success of the forecast. Learning the inherent patterns from the long-term data records by a computer is called ML whose performance improves as the volume of the data records used for training the machine (or learning the patterns by machine) increases. This signifies the use of ML-DL for accurate weather prediction including HIW, given that amounts of data are available.

Further, ML-come-DL-based predictions, as mentioned before, assume that the training and the testing data are from the same distribution. However, this assumption may not always be valid in real-world meteorological, or satellite data as used in weather prediction. This is what is called distribution mismatch problem. The other two problems that ML-DL algorithms usually confront are the inadequate training data and incompatible computation power. For the problems of distribution mismatch and inadequate training data, TL has been helpful and has emerged as an effective strategy for improving prediction abilities in weather forecasting. The TL model may also fine-tune its predictions for weather analysis while reducing the computing cost by utilizing the information once gathered from a big dataset. Similarly, in the data scarcity problem that makes the training of ML models difficult, TL has been useful in maintaining the learning ability based on past knowledge acquired in similar activities, even in the absence of a large amount of data. For example, Dutta and Pal [24] recently provided a new transfer learning-based framework for pollution analysis concerning COVID-19. They used a DL model namely stacked-bidirectional LSTM during the normal periods (when the data is abundant) and reused it during the COVID-19 pandemic (when the data is scarce) for the problem of forecasting the concentration of pollutants. Here, the transfer of knowledge acquired from the previous one also increases the accuracy of the forecast. In this work, the distribution mismatch issue of data is also solved as the distribution of data is different during pandemic situation owing to the fact that pollution concentrations drastically changed during pandemic period. Additionally, TL by assisting the bidirectional LSTM model in learning and storing knowledge from samples at smaller temporal resolutions can enhance the prediction performance for samples at bigger temporal resolutions [25].

In case of prediction of real-life HIW events, one may note that the prediction becomes difficult not only by the nonavailability of adequate data but also owing to the possibility of occurrence of hazardous conditions and damage to the infrastructure of the monitoring stations. These factors accordingly make the nowcast and short-range weather forecasts very difficult. Thus, there is a necessity for implementing more pragmatic, consistent real-time forecasting systems that will offer accurate weather conditions by considering a range of weather variables and climatic factors.

There are many reviews available that focus on the usage of ML and DL in predicting weather, such as bad air quality, cyclones, thunderstorms, and tornadoes. For example, reviews emerging on AI and ML techniques for the prediction of air pollution are reported [26–28]. A survey on DL-based weather prediction is provided by Ren et al. [29]. A similar theoretical review can be found in Zhang et al. [30], Abdalla et al. [31], and Liao et al. [32]. Wu et al. [33] presented an overview of DL algorithms in wind forecasting. Yang and Ismail [34] review the application of DL TL methods in air quality prediction. A general survey on TL is reported by Pan and Yang [35]. TL is frequently utilized in classification [36], speech recognition [37], natural language processing [38, 39], building utilization [40], and medical science [41]. Although there have been many significant studies using TL in weather forecasting available in different journals and conference proceedings, very few reviews consolidating this research under one umbrella have yet been reported. Here comes the necessity of the present study that evaluates the applications of TL algorithms, with a special emphasis on high-impact weather forecasting, for the convenience of researchers. In that sense, this evaluation is unique. It explains broadly—what is TL, why and how it can be used in weather forecasting, what are the merits, and what challenging issues need to be addressed to improve the processing and performance further. These are followed by some concerns for the beginners in ML-Dl-TL research from pattern recognition perspective.

The novelty of our study is as follows:

1) Our present research aims to give a systematic review of the characteristics of different TL-based models that are being extensively used in weather forecasting.
2) Applications using the latest models of ML, DL, and TL in predicting HIW such as bad air quality, cyclones, thunderstorms, and tornados, among other HIW systems, are described.
3) The article evaluates the different TL approaches in terms of accuracy and error metrics. The merits achieved are also mentioned.

4) This article finally addresses the research challenges with TL in weather forecasting tasks, as well as some concerns for future researchers in TL.

1.2. Review methodology

This subsection outlines the approach for conducting this review, including identifying sources, keywords for searching, establishing inclusion/exclusion criteria, and selecting papers.

The article provides a systematic review, and the study draws from findings published in journals, conference proceedings, and books. We mainly focus on the research papers during the past ten years (2013–2023) in our survey. We consider different bibliographic databases as primary sources in this review, some of them are listed as follows: (1) Springer, (2) ScienceDirect, (3) IEEE Explore, (4) Wiley Online Library, (5) Google Scholar, (6) ACM Digital Library, (7) Nature, (8) AMS Journals, (9) Scopus, and so on. Preprint works, such as the ArXiv database, are not taken into account when determining our selection criteria. Additionally, we have discarded any duplicate publications that appeared repeatedly in different literature sources.

The strategic keyword selection highlights the central investigation of our study. The keywords for search criteria are provided as follows: ("*Deep Learning*"), ("*Machine Learning*"), ("*Transfer learning*"), ("*Air pollution prediction*"), ("*Weather prediction*"), ("*Data scarcity*"), ("*LCLU classification*"), ("*Rainfall prediction*"), and so on. During the review process, we examined some real-world applications that are properly mentioned. We have considered 62 articles published in different years for showing the application of TL in weather prediction and overall, 114 articles for conducting this review.

The remainder of the article is structured as follows: The details about TL with some definitions and various categorizations are provided in Section 2. Section 3 describes various DL-TL models for weather prediction where different DL architectures with their characteristics for building deep TL are explained. Various evaluation metrics that were used to assess the performance of different models are illustrated in Section 4. Section 5 provides a brief overview of the application of deep transfer learning models (Deep-TL) with their merits in the prediction of different HIW. A comprehensive overview of different TL applications in the field of weather prediction is illustrated in Section 6. The merits of TL techniques in weather forecasting along with some of the challenges that lie ahead are briefly discussed in Section 7. Finally, our work is concluded with Section 8 which delivers concluding opinions on this survey, some challenges and directions for future investigations, and certain concerns for the new researchers interested in applying TL.

2. Transfer Learning: Some Definitions and Categorizations

TL is a strategy that enhances learning in a new task (called target domain) by transferring previously learned knowledge from a related task (referred to as source domain). The gain from such transfer of knowledge is most notable whenever there is data abundance in the source domain, while inadequate in the target domain. The following definitions, which describe the TL, are provided in this section:

Definition 1 [42]**:** (Domain) A domain, which is denoted by $D = \{\chi, P(X)\}$, is composed of two components:

1) Feature space χ; and
2) Marginal probability distribution $P(X)$, where $X = \{x_1, ..., x_n\} \in \chi$.

Definition 2 [42]**:** (Task) A task, which is denoted by $T = \{Y, f(\cdot)\}$, consists of two components:

1) label space $Y = \{y_1, ..., y_m\}$; and
2) An objective predictive function $f(\cdot)$ which is not observed and needs to be learned by pairs $\{x_i, y_i\}$ where $x_i \in X$ and $y_i \in Y$

One can use the function $f(\cdot)$ to predict the corresponding label, $f(x_i)$, of a new instance x_i. From probabilistic point of view, $f(x_i)$ can be represented as $P(y_i|x_i)$.

Definition 3 [42]**:** (TL) Given a source domain D_s and learning task T_s, and a target domain D_t and learning task T_t, the aim of TL is to help improve the learning of the target predictive function $f_t(\cdot)$ in D_t using the knowledge in D_s and T_s where $D_s \neq D_t$ or $T_s \neq T_t$. Here, the condition $D_s \neq D_t$ refers to either $\chi_s \neq \chi_t$ or $P_s(X) \neq P_t(X)$, while the condition $T_s \neq T_t$ means either $Y_s \neq Y_t$ or $P(Y_s|X_s) \neq P(Y_t|X_t)$.

These two conditions mean, the source and target domains have different feature spaces or marginal probability distributions, and the source and target tasks have different label spaces or conditional probability distributions.

If the TL improves the overall performance by using entirely D_t and T_t, then it is termed a positive transfer. If the information learned from a source domain has a detrimental effect on a target learner, it is called negative transfer.

TL can be divided into inductive TL, transductive TL, and unsupervised TL depending on the settings of the domain and task. Such settings of the task and domain corresponding to (traditional) ML and the said three kinds of transfer learning are mentioned below [35].

For Traditional ML: Source and target domains are the same, and the source and target tasks are also the same.
For Inductive TL: Source and target domains are the same, while the source and target tasks are different but related.
For Transductive TL: Source and target domains are different but related, while the source and target tasks are the same.
For Unsupervised TL: Source and target domains are different but related, and the same character holds for the source and target tasks.

2.1. Inductive transfer learning

In this case, the tasks in the source and target domains are not the same, i.e., $T_s \neq T_t$ regardless of the domains are the same or not. In most cases, well-labeled data are available in the target domain regardless well-labeled data are available or not in the source domain as the former receives special attention. For example, one of the most popular inductive TL approaches is zero-shot learning.

2.2. Transductive transfer learning

In this case, the tasks of the source and target domains are the same; however, the source and target domains are different, i.e., $T_s = T_t$ and $D_s \neq D_t$. In this case, there are no labeled data available in the target domain, but there are a lot of labeled data available in the source domain. Due to unique circumstances in the source and target domains, two additional settings arise in transductive TL. These are: (1) The feature spaces between the source and target domains are different, i.e., $\chi_s \neq \chi_t$ and (2) the feature spaces between domains are the same, i.e., $\chi_s = \chi_t$ but the marginal probability distributions of the input data are different, i.e., $P_s(X_s) \neq P_t(X_t)$. For instance, transductive TL is illustrated by the use of a text classification model trained and tested on college reviews to categorize food reviews.

2.3. Unsupervised transfer learning

Unsupervised TL is defined as the absence of labeled data in both the source and target domains and the tasks in the source and target domains being different. Unsupervised TL aims to help improve the learning of the target predictive function $f_t(\cdot)$ in D_t using the knowledge in D_s and T_s, where $T_s \neq T_t$ and Y_s and Y_t are not observable. Unsupervised TL emphasizes on completing unsupervised learning tasks in the target domain, e.g., clustering, dimensionality reduction, and estimation of density [35]. It is the same as inductive TL, except for the absence of labeled data in both the source and target domains.

Figure 1 adapted from Pan and Yang [35] depicts the taxonomy of the aforementioned categories of TL.

According to the survey by Zhuang et al. [43], the TL technique is classified into four categories namely, instance-based, feature-based, parameter-based, and relational-based approaches, depending on the instance weighting strategy. Instance-based TL techniques are based on using the selected parts (or all) of instances in source data and applying different weighting strategies to be used with the target data [44]. Feature-based techniques map instances (or some features) from both source and target data into more homogeneous data [44]. This can be again classified into two subcategories, namely, asymmetric and symmetric feature-based TL. Asymmetric feature-based TL approaches transform the source features to match the target ones. On the other hand, symmetric approaches aim to identify a common latent feature space and thereafter convert both the source and the target features into a new feature representation [43]. The parameter-based (model-based) TL techniques transfer the knowledge at the mode or parameter level. Relational-based (adversarial-based) TL techniques primarily address the problems in relational domains. The logical relationships or rules acquired in the source domain are transferred to the target domain using these kinds of techniques.

Another categorization namely, homogeneous and heterogeneous TL, depends on the consistent nature between the source and the target feature spaces, and label spaces. In the case of homogeneous TL, the semantics and dimensions of the feature space in both the source domain and the target domain are the same, i.e., $\chi_s = \chi_t$, but the corresponding probability distributions (marginal probability distribution) are not the same, i.e., $P_s(X) \neq P_t(X)$.

However, in heterogeneous TL, the semantics and dimensions of the feature set in the source domain and the target domain differ, i.e., $\chi_s \neq \chi_t$ and the corresponding probabilistic distributions are also not the same, i.e., $P_s(X) \neq P_t(X)$. It is more difficult than homogeneous learning because it needs feature or label space adaptation in addition to distribution adaptation.

3. DL-TL Models for Weather Prediction

This section illustrates an overview of various DL models and their attributes, which are used to build various deep TL-based algorithms for weather prediction. These deep-TL algorithms use labeled data in the source domain, while the target domain might be labeled or unlabeled. TL typically involves two main steps including

Figure 1
Subcategories of transfer learning

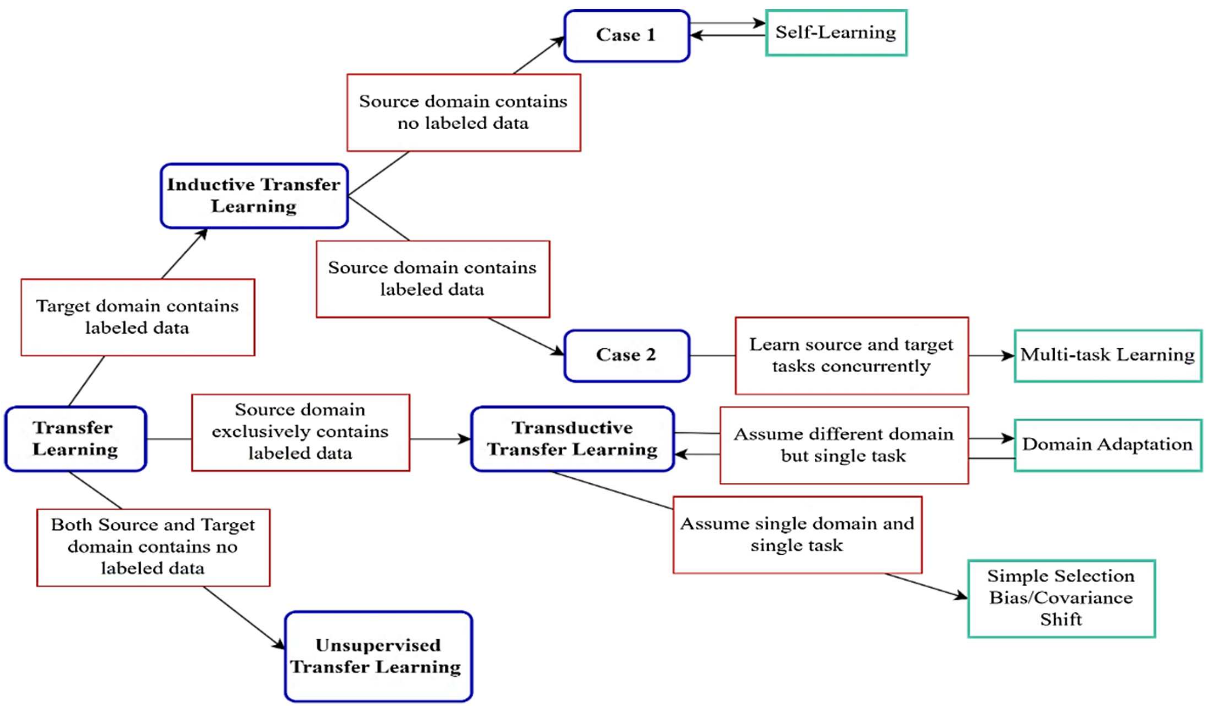
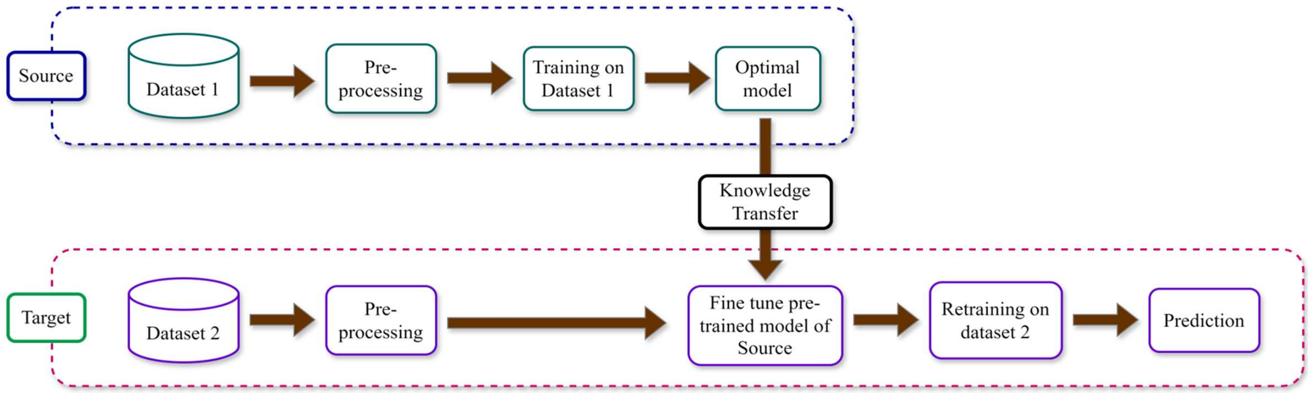


Figure 2
Schematic diagram showing an example of deep transfer learning



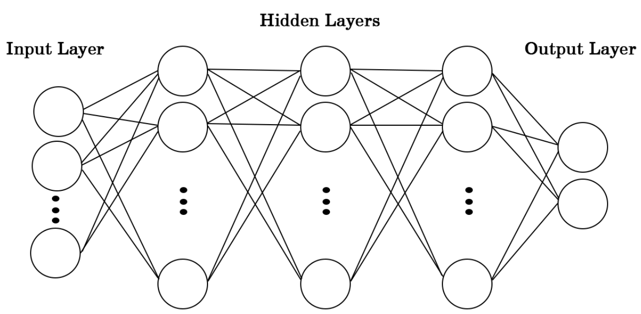
feature extraction and fine-tuning. The fine-tuning requires a pretrained model and utilizing it for a new task by further training it on a task-specific dataset. Fine-tuning involves adjusting the weights of the pretrained model to acquire task-specific patterns while keeping general information from the fresh dataset. In feature extraction, we use the pretrained model as a fixed feature extractor. The final layers responsible for classification are removed and replaced with new layers that are specific to our task. Only the weights of the pretrained model are frozen, while the weights of the newly added layers are trained on the smaller dataset. Thus, with the help of TL, a machine utilizes the knowledge gained from a previous task to improve generalization about another. The TL-based DL network is trained using data from the source domain on a problem that is being solved, and, then reused in a new model to be trained on a related problem to predict the labels of the target domain. Figure 2 depicts deep TL with fine-tuning.

Climatological data are classified under "big data" and can be better analyzed by DL. Yet, most of the ANN-based techniques are unable to enhance the temporal lag of meteorological variables or create long-term dependencies. A number of research have been done using advanced DL techniques such as recurrent neural network (RNN), LSTM, and convolutional neural network (CNN) to model the time series data to deal with this issue [45–47]. Following are some DL models that are employed to design the TL-based framework in predicting HIW.

3.1. Deep neural networks (DNNs)

DNNs are extensively utilized for data-driven modeling. Similar to ANN, DNNs constitute a class of ML paradigms that aims to mimic the information processing of the brain. There are several hidden layers in DNNs situated between the input and output layers (Figure 3). Chai et al. [48] applied a DNN phase picker trained on local seismic data to mesoscale hydraulic fracturing experiments that performed on par with or marginally better than a human expert, thereby resulting in better event locations.

Figure 3
Diagram showing the basic architecture of DNN



3.2. Long short-term memory (LSTM)

LSTM is a type of supervised DNN that is built using ANN and RNN. LSTMs offer good solutions to a variety of sequence learning problems, including prediction of time series. Figure 4(a) shows the structure of an LSTM memory cell. The memory cell of LSTM has three gates: namely, forget gate, input gate, and output gate. The input gate accepts the current input x_t and decides whether the LSTM takes into account its current input. On the other hand, the forget gate permits the LSTM to forget its previous memory C_{t-1}, which is essential for solving the gradient problems. The output gate determines how much of the memory is transferred to the hidden state, h_t, and hence selects what should be the output. Because of its architecture, LSTM is very good at keeping long-term sequences, which are effective for climate modeling and long-term weather forecasting. The outputs of each step in the network are calculated using Equations (1)–(6)

$$f_t = \sigma\left(W_f \bullet [h_{t-1}, x_t] + b_f\right) \tag{1}$$

$$i_t = \sigma\left(W_i \bullet [h_{t-1}, x_t] + b_i\right) \tag{2}$$

$$\widetilde{C}_t = \tanh\left(W_c \bullet [h_{t-1}, x_t] + b_c\right) \tag{3}$$

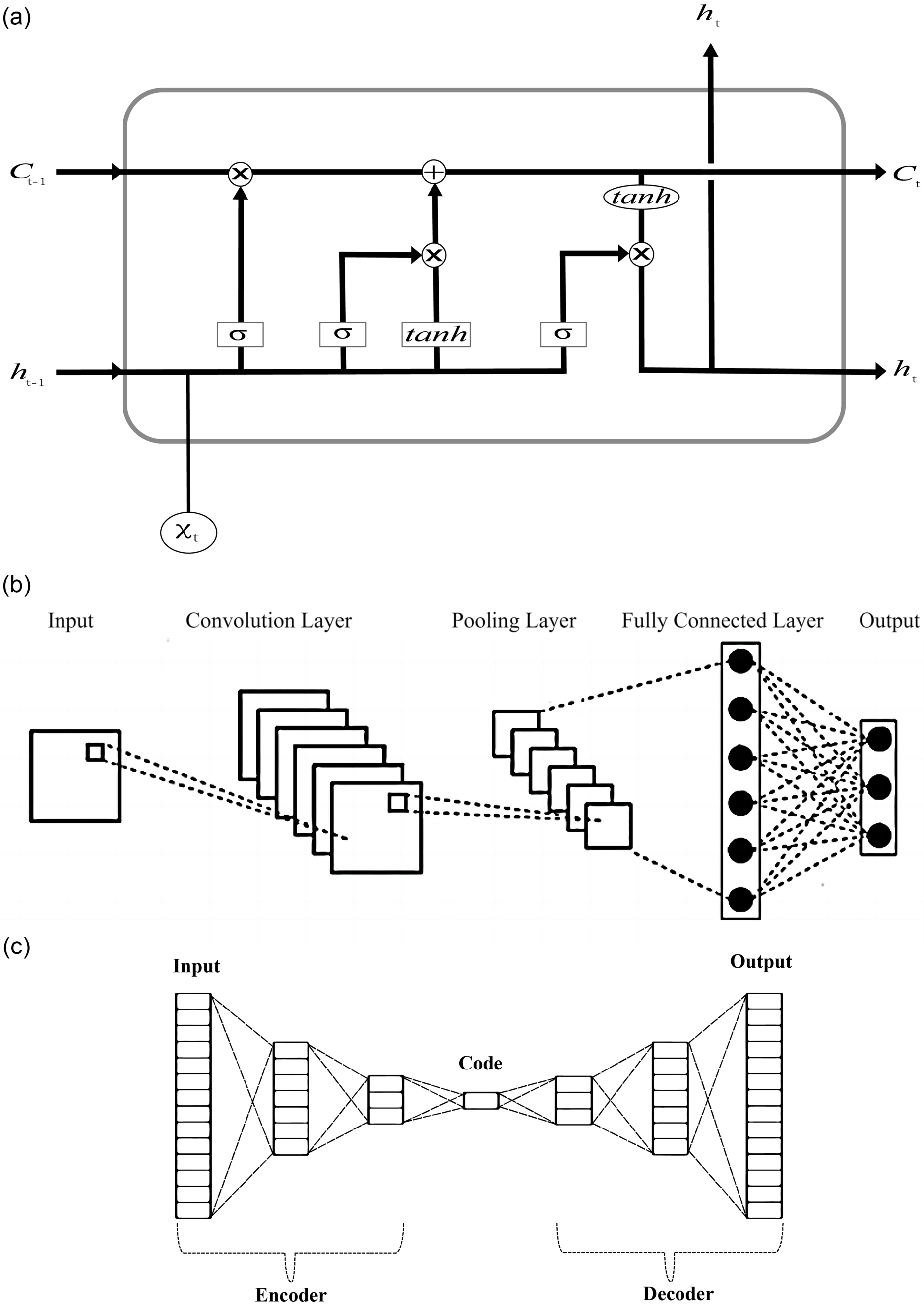
$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{4}$$

$$o_t = \sigma\left(W_o \bullet [h_{t-1}, x_t] + b_o\right) \tag{5}$$

$$h_t = o_t * tanh\left(C_t\right) \tag{6}$$

where h_{t-1} is the hidden state of LSTM, h_t is the new hidden state, C_t and h_t are the output and cell state vectors at time t, i_t is the input gate, o_t is the output gate, f_t is forget gate, and $\sigma(.)$ is the sigmoid activation function.

Figure 4
Diagram showing basic architectures of (a) LSTM block, (b) basic structure of CNN, and (c) fully connected autoencoder architecture



Ma et al. [49] suggested a transfer learned stacked bidirectional long short-term memory (TLS-BLSTM) network to increase the forecast accuracy having transferred knowledge from the current air quality stations to new stations with insufficient data. TL network was useful during the time of data scarcity in the COVID-19 pandemic to predict the concentration of pollutants [24]. Here, a pretrained DL model, viz, Stacked-BDLSTM, incorporating TL was used. The resulting network formed after transfer learning consists of 5 layers, each with two BDLSTM, two dropout layers, and a dense layer. The TL-based model outperformed both single-step and multistep forecasting of pollutant concentrations as compared to other approaches.

3.3. Convolution neural network (CNN)

As mentioned before, DL techniques based on ANNs have grown in prominence in recent years in different ML-driven application domains. Even though computer hardware is becoming more capable nowadays, ANN models are still unable to perform tasks involving images that demand high accuracy due to the fact that each pixel in an image function as a separate input and has a weight assigned to it. ANN-based models for these tasks demand very high processing power, which is beyond the capabilities of existing computer technology. CNN has been useful in such cases. The fundamental distinction between CNN-based ML approaches from conventional ML approaches is that CNN has convolution layers that automatically extract features. CNN models consist fundamentally of convolution, pooling, and fully connected layers (Figure 4(b)). An essential part of the CNN architecture is the convolution layer, which carries out feature extraction. Typically, this involves combining linear and nonlinear processes, such as the activation function and the convolution operation. Typical down sampling is provided by a pooling layer. It calculates the final classification or regression task based on the input from the previous layer. Fully connected layers handle the final classification or regression task using the input from the preceding layer. The output of each class is then converted into a probability score using a logistic function, such as the sigmoid or softmax, using the fully connected layers' output.

CNN-TL was used in time-series flood predictions by Kimura et al. [50]. CNN with TL was used in this work as a technique for converting time-series data to image data. Boonyuen et al. [51] presented a TL-based rainfall forecast using CNN with an Inception-v3 model that could estimate rainfall for the next three days without requiring the experts to analyze the satellite images.

3.4. Encoder–decoder model

The encoder–decoder model is a technique of employing RNNs to predict sequences. The three components of the generally used sequence-to-sequence model are the encoder, the intermediate vector, and the decoder (Figure 4(c)). At each stage, the encoder selects a single data element from the input sequence, analyses it, gathers information about it, and forward it. The intermediate vector is the final state generated from the encoder part and comprises information about the complete input sequence to assist the decoder in providing correct predictions. The decoder provides the entire sentence and forecasts an outcome at each stage. For the prediction of personal air quality, Zhao and Zettsu [52] proposed a TL approach based on an encoder–decoder structure that incorporated the idea of generative adversarial networks (GANs) providing higher accuracy.

3.5. Generative adversarial network (GAN)

These are a class of DL models having two neural networks namely a Generator and a Discriminator. The Generator's job is to produce data that closely resembles a genuine data distribution. On the other hand, the discriminator tries to distinguish between real and generated data. These play a sort of "cat-and-mouse" game while on training, always adapting and getting better in order to produce generated data so convincing that the Discriminator can no longer tell it apart from real data. Zhao and Zettsu [52] introduced a unique decoder TL (DTL) system that includes the concept of GANs which increases the accuracy of the model. They employed an encoder layer similar to GANs to match feature distributions from both the source and target domains simultaneously.

3.6. Hybrid forecasting models with TL

So far, we have discussed the basic models for weather forecasting, hybrid forecasting systems use data-driven methods such as statistical, ML or DL to integrate a wide range of predictions into a final prediction product. They are considered as a potential method for improving the predictability of a meteorological prediction. The use of hybrid forecasting techniques is becoming more popular as a result of improvements in subseasonal to decadal weather and climate prediction systems, a greater understanding of artificial intelligence's advantages, and easier access to computing tools and resources. Gilik et al. [53] provided air quality prediction using CNN+LSTM-based hybrid DL architecture. By using TL, network weights have been transferred from the source city to the target city, and thus, prediction accuracy improves. One may note that utilizing complex hybrid models may increase both computational time and cost. Finding an optimal balance between resource capacity and model complexity requires further effort and investigation. In this context, investigation is required to demonstrate the usefulness of hybrid techniques in diverse areas of weather forecasting to enhance model performance and optimize for specific tasks.

3.7. Hybrid forecasting models based on ensemble learning with TL

An ensemble weather forecast is a collection of forecasts that provide a range of future weather possibilities. Ensemble learning involves training many base learners and combining their predictions using a hybrid technique to get the final outcome. Given that ensemble learning involves combining the skills of several distinct learners, it can outperform single models in most situations in terms of accuracy, robustness, and generalization. In conjunction with other methods, ensemble learning has been widely used in HIW prediction. The method's distinctiveness comes from pretraining models with the same architecture on several source datasets before utilizing the normalized weighted algorithm to ensemble and fine-tune them on the target dataset. For example, Kong et al. [54] suggested an ensemble-based real-time prediction technique for multivariate time-series data to predict air pollution. For both short-term and long-term prediction tasks, they found that the suggested approach for air pollution prediction performed consistently well. A significant number of research have attempted to develop ensemble frameworks that include different methods of TL such as domain adaptation that run a number of times from slightly different starting atmospheric conditions. Developing these ensemble models aims primarily at successfully addressing the shortcomings of pre-trained methods. This strategy reduces the domain mismatch and negative transfer risk by improving the pretrained model's robustness and efficacy, adaptability, and generalizability across domains to various target domains.

3.8. Multi-source transfer learning

The primary goal of the multi-source domain adaptive method is to enable the target domain to acquire rich feature information. The strategy involves learning from several source domains and effectively predicting the target domain data. Multi-source isomorphic transfer learning is the type where the feature space of the various source and target domains is the same. Furthermore, the source domain data are derived from various fields and has the same feature space as the target domain data. Dhole et al. [55] proposed multiSource spatial transfer learning to address the issue of data inadequacy that DL systems face and enhance prediction accuracy. Traditional time series prediction techniques rely on a consistent distribution of training and testing data in a big dataset. Time series data are challenging to analyze due to its time-varying nature, resulting in inconsistencies between new and old data. Gu et al. [56] introduce a new multi-source active-metric transfer learning (MS-AMTL) approach to tackle this issue.

4. Evaluation Metrics of the Models

Weather forecasting algorithms anticipate future occurrences, and there are several methodologies available. Each of these methods has advantages and disadvantages. There are various models accessible in real time, making it difficult to choose the best one for our needs. In order to provide accurate prediction, the model must precisely fit our dataset. The performance the model is assessed using the evaluation metrics. Some of the well-known evaluation metrics are listed below. The predicted, observed, and mean values of the parameter are denoted by $\hat{y}$, y, and $\overline{y}$, respectively, and n is the number of cases (observations).

4.1. R-squared (R^2)/ coefficient of determination

This evaluates the proportion of variance in the response variable that is assessed by the forecast and determines the prediction performance of a model. This represents the accuracy of a model. When the value is higher, this gives better forecast.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{7}$$

4.2. Root mean square error (RMSE)

The standard deviation between the actual and the predicted values of the forecast is assessed using root mean square error (RMSE). Lower RMSE indicates high model efficiency. It can be computed as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{8}$$

4.3. Mean absolute error (MAE)

It represents the deviation between predicted and the measured actual output. It is used to check errors in time series forecasting. It is less sensitive to outliers. A lower mean absolute error (MAE) indicates a greater model effectiveness. It can be computed as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{9}$$

4.4. Mean absolute percentage error (MAPE)

It is the percentage equivalent of MAE as stated earlier. It can be computed as follows:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{10}$$

4.5. Mean square error (MSE)

It represents the square error average that is used as the loss function for the regression of the least squares. It is the sum of the variance between the actual and the predicted variables. RMSE, as discussed earlier, is the square root of MSE. It is calculated as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \tag{11}$$

4.6. Accuracy

The percentage of correct predictions made out of all the observations is known as accuracy. A model that produces more accurate predictions has a higher accuracy score. It is represented as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\% \tag{12}$$

Here, true positives (TP) and true negatives (TN) are the correct predictions, while false negatives (FN) and false positives (FP) are the incorrect predictions.

4.7. Precision

It is a performance statistic that measures a model's ability to accurately forecast positive classes. A model with a higher precision score produces fewer errors while making positive predictions. It is represented as follows:

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

4.8. Recall

It measures how often a model correctly identifies positive instances (true positives) from all the actual positive samples in the dataset. It can be represented as follows:

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

4.9. F1 score

It is a measure of the harmonic mean of precision and recall. F1 score integrates precision and recall into a single metric to gain a better understanding of model performance. It is represented as follows:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{15}$$

4.10. ROC curve

A receiver operating characteristic (ROC) curve is a graphical depiction of a classification model's performance at different threshold values. This plots the true positive rate (TPR) against the false positive rate (FPR) at various classification thresholds. The formula of TPR and FPR are as follows:

$$TPR = \frac{TP}{TP + FN} \tag{16}$$

$$FPR = \frac{FP}{FP + TN} \tag{17}$$

The area under the curve (AUC) is a measurement of the two-dimensional area beneath the entire ROC curve. A model's prediction improves as its AUC increases. It ranges from 0 to 1, with 0 indicating perfectly inaccurate prediction, and a value of 1 indicates a perfectly accurate prediction.

5. Applications of Transfer Learning in Predicting HIW

Here we describe the various attempts made in applying TL in deep networks for different HIW prediction problems, such as air pollution, tropical cyclones, rainfall, flood and drought earthquakes, and LCLU classification. The associated merits exhibited in terms of accuracy and learning time are also mentioned.

5.1. Air pollution

Prediction of air quality has become an important way of managing and preventing pollution, especially in developing countries. Predicting air pollution is a common use of TL, particularly when data are scarce. Incorporating TL into the pretrained model enhances the skill of forecasting. Fong et al. [57] employed LSTM-RNNs for the prediction of future conc. of air pollutants in Macau where certain air quality monitoring stations with less observed data in terms of quantity and type. The results showed greater prediction accuracy with reduced training time by LSTM RNNs having TL methods. To predict air quality for the new stations having data scarcity, Ma et al. [25] developed TLS-BLSTM network that helped to enhance the prediction skill by information transfer from existing air quality stations to new stations (resulting in 35.21% reduction in RMSE). To address the problem of data inadequacy, Dhole et al. [55] described an ensemble strategy for Multi-Source TL. The result provided a cumulative prediction by transferring the knowledge learned from multiple source stations to a specific target station, thereby providing greater use of readily accessible data from neighboring stations to enhance the forecast skill. In predicting pollutant conc., Gilik et al. [53] used a hybrid CNN+LSTM model. For the purpose of weight transfer between cities, the TL approach was used. The said model with the TL approach enhanced the accuracy in terms of RMSE by 11–53% for PM, 20–31% for O_3, 9–47% for NO_2, and 18–46% for SO_2. Zhao and Zettsu [52] used a TL framework based on an encoder–decoder structure for the prediction of personal air quality. For the purpose of predicting air quality in China, DL-based stacking bidirectional long short-term memory was utilized to transfer the information acquired from lower temporal resolutions to higher temporal resolutions, where the model's performance increases in increasing time-based resolutions [25]. Ma et al. [58] developed a unique model for predicting tropospheric ozone (O_3) pollution concentrations in China by combining a long short-term memory neural network with TL (TL-LSTM). Sonawani and Patil [59] used the TL approach to solve the data insufficiency problem in the new air quality monitoring (AQM) system. Forecasting of ambient air pollutants conc. generated during the COVID-19 using TL [60]. Deng et al. [61] proposed transferred CNN long short-term memory (TL-CNN-LSTM) model for the forecast of O_3 conc. In this study, prediction over a large time scale required more data; however, the scarcity of data prevented the CNN-LSTM model from making an accurate prediction. TL-based models reduced the forecast errors (RMSE is reduced by 21%, and the correlation coefficient is enhanced by 13%) and shortened the computational time, thereby improving model prediction accuracy. Njaime et al. [62] presented a data cleaning technique with satellite images and applied TL to estimate NO_2 conc. at Luxembourg with high spatial resolutions based on a pretrained Residual Network 50.

Dutta and Pal [24] used a TL-based stacked-deep architecture to demonstrate the effect of COVID-19 on the concentration of pollutants over Kolkata. Here both the issues of distribution mismatch and scarcity of data are addressed. They first used a bidirectional long short-term memory to predict the concentrations of particulate matter or PM over Kolkata during normal periods. Then using TL approach, this model was re-trained on the data of complete lockdown and partial lockdown periods. Here the distribution of the data during pandemic period is different from that in normal period because of the drastic change in concentration of pollutants due to pandemic situation. This TL-based model is validated during the complete lockdown due to COVID second wave when data are scarce.

Tariq et al. [63] described a TL-based residual neural network architecture for sequence-based prediction of health risk levels measured using subway platform PM2.5 levels. Utilizing the TL framework, there was an improvement of 42.84% of R^2, and the reduction of RMSE was up to 40%. Using TL approaches, Yuan et al. [64] used large-scale stationary and local mobile observations to predict hyperlocal long-term air pollution. Based on a TL approach, Honarvar and Sami [65] proposed a model comprised of numerous components that integrate heterogeneous diverse sources of urban data and forecast PM. Yadav et al. [66] trained a DL model that can map satellite imagery to air quality in high-income countries with adequate ground data, and the model was then modified using TL to learn meaningful air quality estimates in low- and middle-income countries cities. Yang et al. [67] proposed a modified hybrid DL model under the architecture of TL that performed with higher precision and reliability, particularly the forecast of the sites with the scarcity of data in the precise forecast of PM2.5 concentration as compared to classical DL models, hybrid DL models, and the most recent TL approaches.

5.2. Tropical cyclone (TC)

In the precise forecast of tropical cyclones (TC), which is crucial to preventing and minimizing the effect of natural disasters, Pang et al. [68] proposed novel detection techniques from pictures of meteorological satellites integrating the deep convolutional generative adversarial networks (DCGAN) and You Only Look Once (YOLO) v3 model. The novel detection technique comprised three primary components: data augmentation, a pretraining phase, and TL. Experimentally, this technique outperformed the YOLOv3 having an accuracy and average precision of 97.78%, 81.39%, and 93.96%, 80.64%, respectively. In Combinido et al. [69], TL experiments were performed using a Visual Geometry Group 19-1ayer CNN (VGG19) model for the detection of the intensity of TC. Here, VGG19 was pretrained on ImageNet using grayscale IR images of TCs which were collected from different geostationary satellites in the Western North Pacific region (1996–2016). Pan et al. [70] described an architecture of TL-based unbalanced severe typhoon formation prediction (USFP) that used prior knowledge as learned from a constructed balanced dataset. To compensate for the scarcity of seasonal TC observation records, Fu et al. [71] used a TL technique for training an ensemble of CNNs.

5.3. Rainfall, temperature, and wind speed

Prediction of rainfall is crucial due to irregular and heavy rainfall can have a variety of consequences, including widespread damage to farms and crops, damage to infrastructure due to floods, and destruction of property. A better forecasting model is required for early warnings that mitigate flood risk, reduce threats to life and property, and manage agricultural operations. Notarangelo et al. [72] used a TL-based CNN for the detection of rainfall. Taking category-wise ground-based cloud images as input, Ambildhuke and Banik [73] predicted the estimated rainfall. The TL technique provided the maximum prediction accuracy with a small dataset. Boonyuen et al. [51] described a method to forecast the daily rainfall by using CNNs taking satellite images of the areas in Asia as input. Two training methods were used: the former was TL, and the second was training from the beginning. They found that the training from the beginning produced greater accuracy than the TL strategy. This occurred because when training was executed from the start, the model trained all of the layers, but in the TL approach, just the last fully connected layer was trained. Furthermore, while utilizing TL, the batch size could be increased to more than 10 images, when working with the scratch model, the batch size was set at 10 images due to the hardware capability. The TL approach took around 0.3 s to complete one training step, whereas the scratch model took 0.8 s. Liu et al. [74] proposed a TL-based approach for improving precipitation estimation in limited-data areas that would help regional water-related catastrophe prevention and water resource management.

The inhabitants' productivity, well-being, and general health are all impacted by the quality of their indoor environment. When occupants lack sufficient interior thermal comfort, their decision-making and/or professional task execution skills will likely decline, leading to a decline in their performance. To solve the inadequacy in thermal comfort parameters and modeling data, Hu et al. [75] presented a heterogeneous transfer learning (HTL) based intelligent thermal comfort neural network (HTL-ITCNN). In an attempt to lessen the laborious effort of gathering large labeled datasets for every new user, Natarajan and Laftchiev [76] used a transfer active learning framework.

For the purpose of allocating, scheduling, maintaining, and planning wind energy conversion systems, an accurate wind speed forecast is essential. Hu et al. [77] employed deep TL network from data-rich farms to extract the patterns of wind speed and then reused it for newly built farms.

5.4. Flood and drought

Owing to changing climatic scenarios, flood prediction is a critical factor to address. Kimura et al. [50] adopted TL-based CNN to forecast time-series water levels in flood occurrences across domains by applying knowledge of a certain domain. The model with TL in the target domain effectively lowered the training time by one-fifth and the mean error difference by 15% compared to the model without transfer learning. TL approach was applied by Zhao et al. [78] for flood susceptibility mapping highlighting that a pretrained CNN model enhanced the mapping's accuracy for data-scarce areas. TL helped increase the model performance by 10%-25%. Likewise, Muñoz et al. [79] adopted the same approach and illustrated that an inundation model trained on a local scale could be applied to broader scales. Tian et al. [80] used feature-based TL for the prediction of drought in Haihe River Basin.

5.5. Atmospheric visibility

Atmospheric visibility is a crucial factor in determining transparency of the atmosphere, and it is affected by weather and climatic conditions. The scarcity of observational data and unpredictable weather conditions make visibility forecasts challenging. TL approach is utilized to solve the problem while simultaneously improving the quality of the model. In Li et al. [81], a TL algorithm was proposed for predicting atmospheric visibility based on image feature fusion. Here visibility was measured based on data processing and feature extraction in selected subregions of the entire image, resulting in lower computing load and greater efficiency (more than 90%). The article by Li et al. [82] proposed a visibility estimation framework focusing on TL-based DCNN with the scarcity of enough visibility data. The results suggested the detection accuracy surpasses 90% demonstrating that it meets the requirements of daily observation applications. Lo et al. [83] presented a modified approach of the particle swarm optimization (PSO) based TL approach to provide visibility prediction. Gray averaging and the adaptive threshold segmentation method were applied to find the effective subregions. PSO optimized the selection of feature values at the ANN's output layer. Finally, the feature vectors were incorporated into SVR models and overall visibility was evaluated through the fusion method. Lu et al. [84] provided an intelligent offshore visibility prediction approach using a temporal convolutional network and TL (TCN_TL). By using TL, forecast error decreased and with 24 h forecast, the forecast score improved by 0.11 within the 0–1 km level.

5.6. Ocean parameters

Forecasting of ocean parameters is crucial to identify changes in weather patterns. Obara and Nakamura [85] investigated significant wave height (SWH) prediction using LSTM. The findings of the study demonstrated that TL can predict SWH more accurately even with a small amount of training data. In this study, a hybrid model was developed that combines a CNN model with TL to predict SSTA and SSHA on a monthly scale, making full use of the data resources generated by delayed gridding reanalysis products and real-time satellite remote sensing observations. A hybrid model was developed that combines a CNN model with TL to predict SSTA and SSHA by Miao et al. [86]. This could capture the changes in the spatial characteristics of SSTAs and SSHAs over a 30-day period having minimum prediction errors. In the paper of Kumar et al. [87], deep belief networks (DBN) were utilized to transfer learn wave characteristic representations, enabling predictions in new areas of interest.

5.7. ENSO

El Niño and Southern Oscillation (ENSO) is linked to a number of regional climate variability and natural disasters, so, accurate long-range forecasting is crucial for minimizing the economic losses from natural disasters and mitigating more dangerous aspects of climate variability. Hu et al. [88] demonstrated how to predict ENSO using a deep Residual CNN (Res-CNN) model. They noticed that the efficiency of the prediction might be increased by utilizing TL and dropout techniques. TL is utilized for the prediction of ENSO events using historical simulations from CMIP5 and reanalysis data in Ham et al. [89]. Mu et al. [90] employed DL techniques to handle spatiotemporal information while also utilizing TL to transfer knowledge from the dynamical model namely Zebiak–Cane model data to the forecast of realistic El Niño. By applying TL techniques to historical simulations from CMIP5 (Coupled Model Intercomparison Project phase 5, Bellenger et al. [91] and reanalysis data with a CNN model, Ham et al. [89] were able to forecast ENSO events more precisely than current numerical predictions. The forecast was robust and long- term, with a maximum duration of 1.5 years.

5.8. Earthquake

Earthquake prediction is crucial to give early warning of potentially destructive earthquakes to provide an effective response to the disaster, enabling people to reduce loss of life and property. Jozinović et al. [92] demonstrated that CNNs applied to network seismic traces for forecasting earthquake peak ground motion intensity measures (IMs) at distant stations. It was observed that using TL helped to enhance the results in terms of outliers, bias, and variability of the residuals between predicted and true IM values. The application of inductive TL (ITL) was described by Titos et al. [93] as a knowledge basis from which to create dependable and effective volcano-seismic categorization systems. TL methods demonstrate high generalization ability (properly classifying roughly 94% of occurrences) even though using less computation time. Chai et al. [48] applied a DNN phase picker trained on local seismic data to mesoscale hydraulic fracturing experiments. The phase choices generated by the TL model performed similarly to or slightly better than a human expert, resulting in improved event locations. In their study, Wang et al. [94] described a TL technique that used numerical simulations to train a convolutional encoder–decoder for predicting fault-slip behavior in laboratory testing. The model was generalized to yield precise predictions of laboratory fault friction. Maya and Yu [95] used meta-learning and TL to enhance the prediction of earthquakes.

5.9. LCLU classification

Land use and land cover (LCLU) characterize both human activity and the aspects of the plane. The local, regional, and global land use and cover have a significant influence on natural disasters, such as forest fires, calamities, seismic hazards, erosion, floods, and so on. In Alem and Kumar [96], LCLU classification was performed in data-poor cases using two TL frameworks, namely, the Visual Geometry Group (VGG16) and Wide Residual Networks-50 (ResNet-50), on the red–green–blue (RGB) version of the EuroSAT dataset. The model effectiveness and operational efficacy were boosted with the help of model enhancement techniques. A novel approach, namely, Transfer-Ensemble Learning, was used for mapping the urban land use/cover of the Indian metropolitans by Barman et al. [97]. Using TL, the LCLU classification problems were addressed by many researchers [98–105]. Qian et al. [106] proposed an innovative method that combines backdating and TL into an object-based framework. TL is used to choose training samples for classification while backdating optimizes the target region to be categorized. An approach for the automated process of extracting very-high-resolution (VHR) multiclass LC maps from historical orthophotos in the lack of target-specific ground truth annotations was described by van den Broeck et al. [107]. By utilizing domain adaptation and TL, the approach was developed on top of the most recent developments in deep learning. By utilizing transfer learning, Siddamsetty et al. [108] reused the labeled datasets for different regions and thereby minimized the manual annotation costs. The deep TL method of Huang et al. [109] could transfer information from a similarly annotated remote sensing dataset with excellent efficacy and could perform consistently on highly imbalanced classes. It also alleviated the overfitting issue brought on by label noise.

6. Summary of Different TL Approaches in Weather Prediction

This section provides a comprehensive view of different applications of TL in the field of weather prediction with DL models, hybrid DL models, traditional statistical methods, and ML algorithms. These applications have shown the significance of TL in weather prediction. Tables 1–6 depict a list of investigations that utilized TL approaches for the prediction of different HIW tasks. Different features/factors/ parameters such as field of application, year, place of study, method and model involved, assessment metrics, evaluation metrics using TL, and time granularity of the prediction, are also mentioned. In the table, the “Place” column denotes the area on which the research has been done. The column “Methods and Models” indicates the training model and method used in the concerned research. “Evaluation Metrics Using TL” describes the respective evaluation metrices used for predicting weather events. “Target of the prediction or detection” indicates the primary component for which different techniques were applied in order to monitor and measure, and in the final step, to predict the HIW. It varies with respect to the nature and task of HIW. “Time Granularity” means the time resolution which is considered as the prediction interval. The time granularity of the prediction is categorized as -nowcast (up to 6 hours), short range forecast (up to 3 days), medium range forecast (4 to 7 days), extended range forecast (a period extending 10 days to 30 days), or seasonal forecast (from 30 days up to one season) in the future. Some predictions involve multistep predicting a sequence of values in a time series [23, 65, 80, 85, 87]. One may note that in predicting HIW in regional or global scale, different kinds of data are used for analysis. These are, for example, air quality data specially concentration of air pollutants, meteorological data of temperature, pressure and wind speed data, satellite-based image data, and model-based reanalysis data. It may be interesting to investigate how the technologies used in the algorithms varied over time given the distribution of publications, as well as the growing tendency in using ensemble models and hybrid models in recent years.

Figure 5
Number of research papers published on air pollution using transfer learning in prediction high-impact weather

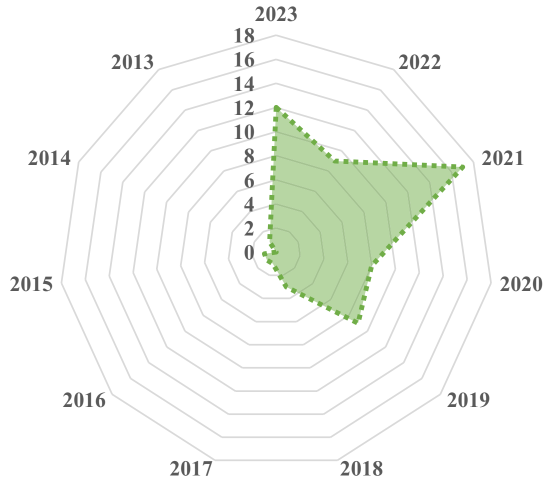


Figure 5 depicts the evolution of research publications with TL for predicting HIW over the years from 2013 to 2023. This is drawn based on the publications collected in our Reference list. One can see a noticeable increase in publication in the year 2021 and an increasing overall trend.

7. Challenges and Opportunities

TL research has delivered advanced results in a number of domains in meteorology thus far. However, there are still several unresolved TL issues that need to be addressed. The following provides a brief overview of such challenges that TL techniques may need to address for better performance in the field of weather prediction.

Table 1
Details of features of different investigations for the prediction of air quality using transfer learning

Authors	Region	Methods and Models	Evaluation Metrics Using TL	Target of the Prediction	Time Granularity
Fong et al. [57]	Macau	LSTM-RNNs with TL	MSE	PM 2.5, PM10, NO_2	1 day
Ma et al. [49]	Anhui, China	TLS-BLSTM	RMSE	PM2.5, NO_2, and O_3	N/S
Dhole et al. [55]	Beijing, China	CNN-GRU and CNN-LSTM with TL	RMSE, MAPE, MAE	PM2.5 conc.	hourly
Gilik et al. [53]	Barcelona, Kocaeli, and İstanbul	Hybrid CNN+LSTM with TL	RMSE	PM, O_3, NO_2, SO_2	N/S
Zhao and Zettsu [52]	Kyushu, Japan and 33 coastal cities in eastern Asia	Convolutional recurrent neural networks (D-CRNN)	SMAPE and accuracy	Air quality index of PM2.5	6 hours
Ma et al. [25]	Guangdong, China	Transferred bi-directional long short-term memory (TL-BLSTM)	RMSE, MAE, MAPE	PM2.5	N/S
Ma et al. [58]	China	TL-LSTM	R^2 and MSE	Tropospheric ozone (O_3) pollution conc.	1 h to 40 h
Sonawani and Patil [59]	Pune in India,	CNN-GRU with TL	RMSE, MSE, MAE	O_3 conc.	Hourly
Chen et al. [60]	Wuhan, China	Gaussian mixture model (GMM) with TL	MSE, MAE, EVS, and R^2_score	Ambient air pollutants conc.	7 days
Deng et al. [61]	Eisenhuttenstadt, Germany	TL-CNN-LSTM	RMSE and Pearson correlation coefficient, Time	O_3 conc.	Daily
Njaime et al. [62]	Luxembourg	Residual Network 50	RMSE, MAE, R^2	NO_2 conc.	N/S
Dutta and Pal [24]	Kolkata	stacked-BDLSTM with TL	RMSE, MAE, R^2	PM10 and PM2.5	24 h,48 h,72 h,96–120 h
Tariq et al. [63]	Korea	TL-based residual neural network	R^2, RMSE	PM2.5 conc.	
Yuan et al. [64]	Amsterdam, Netherlands, Europe	Global2Local- TrAdaBoost	R^2, MAE, RMSE	NO_2	24 h
Honarvar and Sami [65]	Aarhus, in Denmark	Predictive model with TL	RMSE, MAE	PM10	3, 6, 12, and 24 hours
Yadav et al. [66]	Accra in Ghana, Africa	CNN-based Generative Adversarial (GAN)	*nrmse*, R^2	NO_2	N/S
Yang et al. [67]	Beijing and Hengshui	TL-GRU, TL-LSTM, TL-CNN-LSTM, TL-Modified Model	RMSE, MAE, MSE	PM2.5 conc.	N/S

1) At the outset, the challenges lie in the computational complexity of exhaustively investigating TL algorithms to choose the optimal one. The majority of TL algorithms in use today mostly depend on human guidance. In real-world problems in meteorology, we expect (desire) that models will be able to learn an unknown task on their own. Further, in existing TL models, the training often takes a significant amount of time involving computation power. Incorporating human expertise (pre-experience) and encoding domain knowledge into TL models may reduce that time drastically. Such a human-guided TL is seen to enhance the effectiveness of TL algorithms [110]. In this context, one may refer to the hybrid multilayer perceptron model [111] where extracting the domain knowledge as rough rules (reducts) and encoding them as link weights during the formation of the neural architecture was found to enhance the network performance with reduced learning time greatly.
2) In a variety of real-world problems in meteorology, transfer learning has improved prediction efficiency despite having a

Table 2
Details of features of different investigations for the prediction of tropical cyclones and rainfall using transfer learning

Authors	Region	Methods and Models	Evaluation Metrics Using TL	Target of the Prediction	Time Granularity
Pang et al. [68]	Southwest Pacific area (China)	(DCGAN) and You Only Look Once (YOLO) v3 model	Accuracy, precision, recall	Accurate detection of tropical cyclones	N/S
Combinido et al. [69]	Western North Pacific basin	CNN with TL	RMSE	Tropical cyclone (TC) intensity	6 h
Pan et al. [70]	typhoons in the Western Pacific (WP), Eastern Pacific (EP), and North Atlantic (NA) regions	ConvLSTM with TL	Accuracy, ROC_AUC, PR_AUC, F1	Unbalanced severe typhoon formation prediction (USFP)	24 h
Fu et al. [71]	North Atlantic (NAT), eastern North Pacific (ENP), western North Pacific (WNP), North Indian Ocean (NIO), South Indian Ocean (SIO), South Pacific Ocean (SPO), and South Atlantic (SAT)	Ensemble of CNNs with TL	RMSE	Seasonal tropical cyclone activity	N/S
Notarangelo et al. [72]	Japan	TL-based convolutional neural network	Receiver operating characteristic (ROC) curve, F1 score, accuracy	Rainfall	N/S
Boonyuen et al. [51]	Thailand	Convolutional neural networks (CNN)	Accuracy	Rainfall	1–3 days
Liu et al. [74]	China	Domain-adversarial neural network (DANN) with TL	RMSE and MAE	Precipitation	daily

shortage of training data. However, negative transfer occurs when information gained in the source domain has a negative impact on the task in the target domain due to dissimilarity of data, or the transfer method is unable to identify the transferable components. To mitigate the effect of negative transfer, the transferability of the source task to the target task, as well as the similarity of domains or tasks, should be thoroughly assessed prior to the construction of successful models. Improving the data quality from coarse to fine-grained at the domain, instance, or feature level may enhance the transferability of the source domain [112]. For this purpose, one can choose a subset of source domains or weight them when there are multiple source domains present at the domain level. The source instances can be selected or weighted at the instance level. We can boost the transferability at the feature level by transforming the original features into a common latent space.

3) Another drawback of Tl approach is the risk of domain mismatch. This issue arises due to differences in domain data or divergence in domain data distributions. This issue affects knowledge transfer efficiency and negatively impacts the performance of target domain models. It is a major factor that contributed to the negative transfer that occurred in the pretraining model. Therefore, when transferring pretrained models to new environment, it is necessary to design and customize appropriate models, structures, and approaches that can handle these issues efficiently.
4) Similar to “Black Box,” ML or DL models are created by algorithms that are incomprehensible to the users and developers. Due to the intrinsic intricacy of the two-stage training process, pretraining approaches in the models having TL worsen [113]. As a result, there is uncertainty about how target models use acquired knowledge and its influence on the decision-making process leading to the failure of the transfer learning process and the addition of time costs for the users. This creates difficulties in acquiring and implementing TL technologies. In order to generate predictions from a set of input variables, users or even those who develop the model have no knowledge of how variables are combined or interact. Because of the lack of interpretability, the TL technique’s trustworthiness is hampered and its practical applications are questioned, necessitating more research in this area.
5) Another shortcoming of the TL method is label space shifts. This refers to the differences in available labels between the source and target domains. Researchers often avoid utilizing heterogeneous data for knowledge transfer, preferring to use labels that are comparable across domains. Though in some cases this restricts the efficacy of the model resulting in negative transfer.
6) Another issue that arises is the problem of overfitting which is a major shortcoming of practically all prediction methods. Moreover, it is one of the common biases in big data. However, in case a new model learns details and noises from the training data

Table 3
Details of features of different investigations for the prediction of thermal comfort, drought, and flood using transfer learning

Authors	Region	Methods and Models	Evaluation Metrics Using TL	Target of the Prediction	Time Granularity
Hu et al. [75]	N/S	Heterogeneous transfer learning (HTL) based intelligent thermal comfort neural network (HTL-ITCNN)	Accuracy, Macro-F1, MCC	Thermal comfort	N/S
Natarajan and Laftchiev [76]	N/S	Regression with TL	RMSE	Thermal comfort	N/S
Hu et al. [77]	N/S	N/S	N/S	Flood Predictions	N/S
Kimura et al. [50]	Japan	Transfer learning-based CNN	RMSE	Time-series water levels in flood occurrences	hourly
Zhao et al. [78]	Dahongmen, Qinghe, and Bahe in Beijing in China	CNN with TL	N/S	Flood susceptibility assessment	N/S
Muñoz et al. [79]	Southeast Atlantic coast of the U.S	Convolutional neural networks (CNNs) and data fusion (DF) with TL	High water marks (HWMs) and the advanced fitness index (AFI), overall accuracy, f1-scores, Cohen's kappa	Flood mapping	up to 7 days
Tian et al. [80]	China	CNN, RF, LSTM, WNN, SVR With TL	MAPE, SMAPE, MAE, MSE, R^2	Prediction of drought	3, 6, 9, and 12 months

Table 4
Details of features of different investigations for the prediction of atmospheric visibility, ocean parameters, and ENSO using transfer learning

Authors	Region	Methods and Models	Evaluation Metrics Using TL	Target of the Prediction	Time Granularity
Li et al. [81]	China	Support vector regression (SVR) with TL	Accuracy	Atmospheric visibility	Hourly
Li et al. [82]	China	TL-based DCNN	Test time, detection accuracy	Visibility	2 and 16 h
Lo et al. [83]	China	Particle swarm optimization (PSO) based TL approach	Accuracy	Visibility	N/S
Lu et al. [84]	Qiongzhou Strait,Chaina	Temporal convolutional network and TL (TCN_TL)	RMSE, MAE, Threat score (TS)	Offshore visibility	24 h
Obara and Nakamura [85]	Japan	LSTM with TL	RMSE and R^2	Significant wave height (SWH)	6-, 12-, and 24-h

(*Continued*)

Table 4
(*Continued*)

Authors	Region	Methods and Models	Evaluation Metrics Using TL	Target of the Prediction	Time Granularity
Miao et al. [86]	South China Sea	CNN model with TL	Correlation coefficient (CC) and root mean squared error (RMSE)	SSTA and SSHA	30 days
Kumar et al. [87]	Bohai Sea, Yellow Sea, and East China Sea	Deep belief networks (DBN) with TL	Bias, scatter index (SI), correlation coefficient (CORR), and root mean square error (RMSE).	Wave characteristic representations	Hourly
Hu et al. [88]	China	Deep Residual CNN (Res-CNN) with TL	Accuracy	ENSO	3, 6, 9, 12, 18, 23 months
Ham et al. [89]	N/S	CMIP5 and CNN with TL	Hit rate	ENSO	Up to one and a half years
Mu et al. [90]	N/S	ConvLSTM (Convolutional Long Short-Term Memory Network) with TL	Correlation r, MSE, RMSE	Realistic El Niño	3 to 12 months

Table 5
Details of features of different investigations for the prediction of earthquake using transfer learning

Authors	Region	Methods and Models	Evaluation Metrics Using TL	Target of the Prediction	Time Granularity
Wang et al. [94]	Cascadia and the San Andreas Fault	Convolutional encoder–decoder (CED) with TL	MAPE	Fault-slip behavior	N/S
Jozinović et al. [92]	Central western Italy	CNNs	Number of outliers, mean, median, standard deviation	Earthquake peak ground motion intensity measures (IMs)	N/S
Titos et al. [93]	Colima (Mexico)	CNN with TL	Accuracy, speed up	Isolated volcano-seismic events	N/S
Chai et al. [48]	N/S	DNN phase picker with TL	Accuracy	Mesoscale hydraulic fracturing experiments	N/S
Maya and Yu [95]	N/S	MLP with meta-learning and transfer learning	MSE	Earthquakes	N/S
Wang et al. [94]	Cascadia and the San Andreas Fault	Convolutional encoder–decoder (CED)	MAPE	Fault-slip behavior	N/S
Wang et al. [94]	N/S	Convolutional encoder–decoder with TL	Accuracy	Fault-slip behavior	N/S

Table 6
Details of features of different investigations for LULC classification using transfer learning

Authors	Region	Methods and Models	Evaluation Metrics Using TL	Target of the Prediction	Time Granularity
Alem and Kumar [96]		Two TL frameworks, namely, the Visual Geometry Group (VGG16) and Wide Residual Networks-50 (ResNet-50)		LULC classification	N/S
Barman et al. [97]	Kolkata Metropolitan Area (KMA), India	SVM, RF with TL	Precision, recall, F-score, overall accuracy, and Kappa coefficient	Urban land use/cover	N/S
Naushad et al. [98]	Many	CNN with TL	Accuracy	LCLU classification	N/S
Dastour and Hassan [100]	Canada	CNN with TL	Precision, Recall, F1	LCLU classification	N/S
Wu et al. [101]	China	With TL	Accuracy	LCLU classification	N/S
Li et al. [102]	China	Iterative reweighting heterogeneous transfer learning (IRHTL)	Accuracy	LCLU classification	N/S
Demir et al. [103]	China	Change-detection-driven TL	Accuracy	LCLU classification	N/S
Lin et al. [104]	China	TL, Clustering	Accuracy	LCLU classification	N/S
Yifter et al. [105]	Moscow region, Russia	CNN with TL	Accuracy	LCLU classification	N/S
Qian et al. [106]	China	Classification and change analysis with TL	Accuracy	LCLU classification	N/S
van den Broeck et al. [107]	Turkey	Fully convolutional networks (FCN) with TL	Accuracy	Multiclass land cover mapping	N/S
Siddamsetty et al. [108]	Different	N/S	Precision, Recall, F2 score	LCLU classification	N/S
Huang et al. [109]	N/S	CNN with TL	Overall accuracy, F1 score	LCLU classification	N/S

that adversely affect its outputs, overfitting arises in the context of TL.

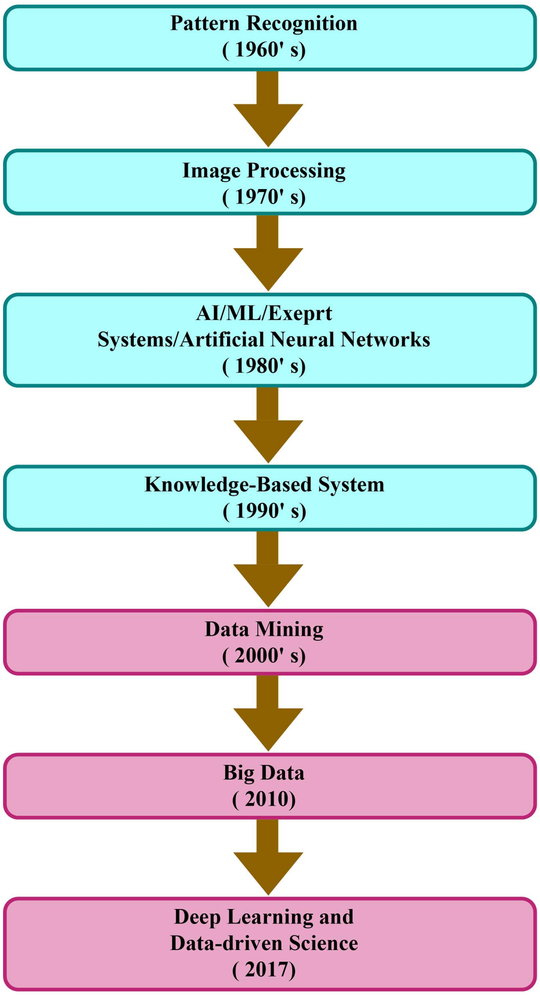
One may note that every algorithm or technology has its own pros and cons according to the task at hand and the environment. The advantages of TL techniques in predicting HIW are listed below.

1) The fundamental benefit of TL approach is its capacity to deal with data scarcity issue. Data scarcity is a significant difficulty in ML-DL models, which can lead to overfitting, performance deterioration, and thus hinder the advancement of DL applications. Our systematic review acknowledges many TL solutions specifically designed to mitigate this problem. TL approaches reduce the need for large-scale data, making them essential for applying DL to real-world scenarios.
2) The TL approaches have the benefit of improving the performance of target model. Positive knowledge transfer enhances the target model's ability to build networks with deeper layers and more parameters without requiring considerable training data. Experimental findings of different applications reveal that TL approaches significantly improve model performance not only single step but also in multistep weather prediction.
3) The other benefit of TL approaches is their high level of generalizability. These techniques extract generalized features from one task to another. Building reusable models may significantly reduce the time and computing resources required for DL training, resulting in considerable cost savings. Researchers aspire for their trained models to be usable in a wider number of disciplines and scenarios, maximizing their usefulness and relevance.
4) The fourth advantage of TL techniques is their ability to speed up the convergence speed of the target model. This approach also reduces the processing expenses, making it useful for real-world applications.
5) TL approaches help models learn from simulations. This helps the model to gather valuable experience and learn optimum actions without the hazards of real-world testing. When a model performs well in simulation, it may be confidently applied in real-world applications.

8. Conclusions, Future Scope, and Some Concerns

In this review, we have described various methods and tactics of TL from both the data and model perspectives, and its application in HIW prediction. The objective is to make the climate analytics researchers aware of the emerging TL technology in AI, viz, what it is, why and how it can be applied in prediction problem, and what are the future scope. The article includes precise definitions

Figure 6
Evolution of disciplines from the mother subject "Pattern recognition" over six decades since 1960s



of TL, relevance of TL to weather prediction, and provides several examples of TL approaches and related literature. A comparative study of various applications in HIW, during the past ten years (2013–2023), assessing the methodologies adopted, performance metrics used for the prediction, study area considered, target of the prediction, etc. is provided. The primary conclusions of the current study are as follows:

1) TL algorithms are widely used in predicting HIWs such as air pollution, tropical cyclones, rainfall, flood and drought earthquakes, and LCLU classification.
2) TL mechanism when embedded into DL makes the deep model possess some merits, such as lesser dependency on data labels and size, saving training time and storage space, and improving the effectiveness of the model. TL approaches enable the adaption of advanced ML-DL models for specific applications and settings.
3) The TL approach has a number of benefits including eliminating the scarcity of data and distribution mismatch issues, minimizing overfitting, and saving training time. By using TL, lower training time has been required to get the output of the model, and the generalization capability of the model is increased. All these characteristics make the TL a significant component in ML-DL-based real-life applications. For example, consider the pollution estimation during COVID-19 pandemic period in Kolkata, India. Here the DL model [24] designed during normal period having abundant data was retrained using TL during the complete lockdown and partial lockdown period. This was achieved by transferring the knowledge acquired during the normal period to the pandemic situation. Here the distribution of data was different from the normal period as a drastic change in concentration of pollutants obtained due to pandemic situation. This new TL-based model validated during the complete lockdown period of COVID second wave where the data was scarce. In this problem, both the data scarcity and distribution mismatch issue are addressed.
4) Drawbacks of TL methods include domain mismatch, negative knowledge transfer, and label shift. Interpretability of the model is necessary for understanding decision making of the model in handling extreme weather events.

One may note that we only considered HIW related to atmospheric phenomena including tropical cyclone, doughts, flood, visibility that are natural or indirectly influenced by man. We have not considered here the HIW directly induced by man. Further, the study is based on the results reported in journals, conferences proceedings, and books, published only over the past decade, 2013–2023. Some important papers might have been overlooked/ omitted inadvertently from inclusion in the study.

Future studies using TL may proceed in several distinct fields as follows:

In the area of weather forecasting, a greater range of applications for TL approaches can be investigated and implemented. To tackle knowledge transfer issues in more intricate applications, novel methodologies are required for development. For instance, it is crucial to determine the best way to quantify the transferability between domains while preventing the negative transfer. Despite a few research, one still needs a more systematic analysis to fully understand the principle of negative transfer. Determining the interpretability of TL is another crucial issue that constitutes a scope of future research. It means, making TL models more interpretable, transparent, and explainable so that the users and readers may comprehend the decision-making process of the model in generating the output.

We have discussed a comprehensive analysis of TL designed for and used in predicting weather phenomena. There are several areas in meteorology where the method of TL is awaited to generalize. These are forecasting of tornado, lightning, thunderstorm, straight-line wind, hail, heatwave, and multi hazard prediction, among others.

More theoretical research needs to be done to offer theoretical background to understand the functioning of TL and for justification of its enhanced efficacy and applicability. In this context, there is a need to research in insecurity aspects of transfer learning, especially vulnerabilities and attacks specific to these methodologies. Further, instead of generating a single forecast, an ensemble forest utilizing TL approaches may provide an indication of the probable future states of the atmosphere.

Moving future, heterogeneous TL systems will become increasingly significant due to the diversity in data acquisition. The availability of "Big data" demonstrates the possibility of implementing deep learning alongside the existing TL techniques. In meteorology or atmospheric sciences, there is a chance of having both labeled and unlabeled source, and unlabeled target data. Unfortunately, not many TL techniques are there that can deal with the situation of unlabeled source and unlabeled target data. This is undoubtedly a topic that needs more investigation.

In conclusion, we believe the present review will enable the readers, particularly in climate analytics and data science, to have a more thorough grasp of the concepts and current status of research concerning the theory of TL and its applicability in weather prediction.

Finally, we mention some concerns for beginners in DL and data science research. While trying to develop AI- and DL-based approaches for different applications in data analytics, one may note the observation made by Pal et al. [22] concerning the evolution of the discipline from the mother subject "Pattern recognition" over six decades since the 1960s approximately as shown in Figure 6 [22].

At every evolution from the mother subject [114], new approaches, theories, and technologies were developed, and new terms were coined with big expectations to deal with the varying nature of data, as well as the decision-making tasks. Still, a beginner should avoid diving headfirst into new technologies before knowing adequately the basic theories. As an instance, to understand the functioning of deep learning including TL, one should know the same for shallow learning (viz, ML and ANNs). For knowing the ML and ANNs, one should have full knowledge of pattern recognition. If not, it can quickly result in dissatisfaction by criticizing the DL-TL approaches and its existing models. This is what we had witnessed in ANN research which experienced a revitalization in the 1980s with a big expectation, but the field nearly lost interest within a period of about 12 to 15 years. One of the reasons for this was a paucity of scientific research into the operation of the "black box" (neural network) systems and designing new application-specific architectures [22].

"Knowing & learning ancestors makes one's knowledge mind development better."

Funding Support

The work of Sankar K. Pal was supported by the National Science Chair, Science and Engineering Research Board-Department of Science and Technology (SERB-DST), Government of India, (National Science Chair Mode-1) under Grant SB/N-SC/2019/01.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Sankar Kumar Pal: Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Debashree Dutta:** Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] Charney, J. G., Fjörtoft, R., & von Neumann, J. (1950). Numerical integration of the barotropic vorticity equation. *Tellus A: Dynamic Meteorology and Oceanography, 2*(4), 237–254. https://doi.org/10.3402/tellusa.v2i4.8607

[2] Miyoshi, T., Lien, G. Y., Satoh, S., Ushio, T., Bessho, K., Tomita, H., ..., & Seko, H. (2016). "Big data assimilation" toward post-petascale severe weather prediction: An overview and progress. *Proceedings of the IEEE, 104*(11), 2155–2179. https://doi.org/10.1109/jproc.2016.2602560

[3] Pal, S. K. (1998). Soft computing tools and pattern recognition. *IETE Journal of Research, 44*(1–2), 61–87. https://doi.org/10.1080/03772063.1998.11416030

[4] Hema, C., & Marquez, F. P. G. (2023). Emotional speech recognition using CNN and deep learning techniques. *Applied Acoustics, 211,* 109492. https://doi.org/10.1016/j.apacoust.2023.109492

[5] Rodzin, S., Bova, V., Kravchenko, Y., & Rodzina, L. (2022). Deep learning techniques for natural language processing. In *Artificial Intelligence Trends in Systems: Proceedings of 11th Computer Science On-Line Conference,* 121–130. https://doi.org/10.1007/978-3-031-09076-9_11

[6] Pal, S. K., & Kumar, D. A. (2023). Adaptive granulation-based convolutional neural networks with single pass learning for remote sensing image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16,* 57–70. https://doi.org/10.1109/jstars.2022.3223180

[7] Almahasneh, M., Paiement, A., Xie, X., & Aboudarham, J. (2022). MLMT-CNN for object detection and segmentation in multi-layer and multi-spectral images. *Machine Vision and Applications, 33*(1), 9. https://doi.org/10.1007/s00138-021-01261-y

[8] Pramanik, A., Pal, S. K., Maiti, J., & Mitra, P. (2022). Traffic anomaly detection and video summarization using spatio-temporal rough fuzzy granulation with Z-numbers. *IEEE Transactions on Intelligent Transportation Systems, 23*(12), 24116–24125. https://doi.org/10.1109/tits.2022.3198595

[9] Wang, C., Pan, R., Wan, X., Tan, Y., Xu, L., McIntyre, R. S., ..., & Ho, C. (2020). A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. *Brain, Behavior, and Immunity, 87,* 40–48. https://doi.org/10.1016/j.bbi.2020.04.028

[10] Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., ..., & Albekairy, A. M. (2023). Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC Medical Education, 23*(1), 689. https://doi.org/10.1186/s12909-023-04698-z

[11] Albattah, W., Nawaz, M., Javed, A., Masood, M., & Albahli, S. (2022). A novel deep learning method for detection and classification of plant diseases. *Complex & Intelligent Systems, 8*(1), 507–524. https://doi.org/10.1007/s40747-021-00536-1

[12] Chandola, D., Mehta, A., Singh, S., Tikkiwal, V. A., & Agrawal, H. (2023). Forecasting directional movement of stock prices using deep learning. *Annals of Data Science, 10*(5), 1361–1378. https://doi.org/10.1007/s40745-022-00432-6

[13] Mondini, A. C., Guzzetti, F., & Melillo, M. (2023). Deep learning forecast of rainfall-induced shallow landslides. *Nature Communications, 14*(1), 2466. https://doi.org/10.1038/s41467-023-38135-y

[14] Chaudhuri, S., Dutta, D., Goswami, S., & Middey, A. (2015). Track and intensity forecast of tropical cyclones over the North Indian Ocean with multilayer feed forward neural nets. *Meteorological Applications, 22*(3), 563–575. https://doi.org/10.1002/met.1488
[15] Chaudhuri, S., Dutta, D., Goswami, S., & Middey, A. (2013). Intensity forecast of tropical cyclones over North Indian Ocean using multilayer perceptron model: Skill and performance verification. *Natural Hazards, 65,* 97–113. https://doi.org/10.1007/s11069-012-0346-7
[16] Dutta, D., & Chaudhuri, S. (2015). Nowcasting visibility during wintertime fog over the airport of a metropolis of India: Decision tree algorithm and artificial neural network approach. *Natural Hazards, 75,* 1349–1368. https://doi.org/10.1007/s11069-014-1388-9
[17] Bączkiewicz, A., Wątróbski, J., Sałabun, W., & Kołodziejczyk, J. (2021). An ANN model trained on regional data in the prediction of particular weather conditions. *Applied Sciences, 11*(11), 4757. https://doi.org/10.3390/app11114757
[18] Akbari Asanjan, A., Yang, T., Hsu, K., Sorooshian, S., Lin, J., & Peng, Q. (2018). Short-term precipitation forecast based on the PERSIANN system and LSTM recurrent neural networks. *Journal of Geophysical Research: Atmospheres, 123*(22), 12543–12563. https://doi.org/10.1029/2018JD028375
[19] Zhao, F., Liang, Z., Zhang, Q., Seng, D., & Chen, X. (2021). Research on PM2.5 spatiotemporal forecasting model based on LSTM neural network. *Computational Intelligence and Neuroscience, 2021*(1), 1616806. https://doi.org/10.1155/2021/1616806
[20] Pan, B., Hsu, K., AghaKouchak, A., & Sorooshian, S. (2019). Improving precipitation estimation using convolutional neural network. *Water Resources Research, 55*(3), 2301–2321. https://doi.org/10.1029/2018wr024090
[21] Dairi, A., Harrou, F., Khadraoui, S., & Sun, Y. (2021). Integrated multiple directed attention-based deep learning for improved air pollution forecasting. *IEEE Transactions on Instrumentation and Measurement, 70,* 1–15. https://doi.org/10.1109/tim.2021.3091511
[22] Pal, S. K., Pramanik, A., Maiti, J., & Mitra, P. (2021). Deep learning in multi-object detection and tracking: State of the art. *Applied Intelligence, 51*(9), 6400–6429. https://doi.org/10.1007/s10489-021-02293-7
[23] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ..., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data, 8,* 53. https://doi.org/10.1186/s40537-021-00444-8
[24] Dutta, D., & Pal, S. K. (2023). Prediction and assessment of the impact of COVID-19 lockdown on air quality over Kolkata: A deep transfer learning approach. *Environmental Monitoring and Assessment, 195*(1), 223. https://doi.org/10.1007/s10661-022-10761-x
[25] Ma, J., Cheng, J. C. P., Lin, C., Tan, Y., & Zhang, J. (2019). Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. *Atmospheric Environment, 214,* 116885. https://doi.org/10.1016/j.atmosenv.2019.116885
[26] Masood, A., & Ahmad, K. (2021). A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. *Journal of Cleaner Production, 322,* 129072. https://doi.org/10.1016/j.jclepro.2021.129072
[27] Chantry, M., Christensen, H., Dueben, P., & Palmer, T. (2021). Opportunities and challenges for machine learning in weather and climate modelling: Hard, medium and soft AI. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 379*(2194), 20200083. https://doi.org/10.1098/rsta.2020.0083
[28] Essamlali, I., Nhaila, H., & El Khaili, M. (2024). Supervised machine learning approaches for predicting key pollutants and for the sustainable enhancement of urban air quality: A systematic review. *Sustainability, 16*(3), 976. https://doi.org/10.3390/su16030976
[29] Ren, X., Li, X., Ren, K., Song, J., Xu, Z., Deng, K., & Wang, X. (2021). Deep learning-based weather prediction: A survey. *Big Data Research, 23,* 100178. https://doi.org/10.1016/j.bdr.2020.100178
[30] Zhang, B., Rong, Y., Yong, R., Qin, D., Li, M., Zou, G., & Pan, J. (2022). Deep learning for air pollutant concentration prediction: A review. *Atmospheric Environment, 290,* 119347. https://doi.org/10.1016/j.atmosenv.2022.119347
[31] Abdalla, A. M., Ghaith, I. H., & Tamimi, A. A. (2021). Deep learning weather forecasting techniques: Literature survey. In *2021 International Conference on Information Technology,* 622–626. https://doi.org/10.1109/ICIT52682.2021.9491774
[32] Liao, Q., Zhu, M., Wu, L., Pan, X., Tang, X., & Wang, Z. (2020). Deep learning for air quality forecasts: A review. *Current Pollution Reports, 6*(4), 399–409. https://doi.org/10.1007/s40726-020-00159-z
[33] Wu, Z., Luo, G., Yang, Z., Guo, Y., Li, K., & Xue, Y. (2022). A comprehensive review on deep learning approaches in wind forecasting applications. *CAAI Transactions on Intelligence Technology, 7*(2), 129–143. https://doi.org/10.1049/cit2.12076
[34] Yang, J., & Ismail, A. W. (2022). Air quality forecasting using deep learning and transfer learning: A survey. In *2022 IEEE Global Conference on Computing, Power and Communication Technologies,* 1–6. https://doi.org/10.1109/GlobConPT57482.2022.9938230
[35] Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering, 22*(10), 1345–1359. https://doi.org/10.1109/tkde.2009.191
[36] Bansal, M., Kumar, M., Sachdeva, M., & Mittal, A. (2023). Transfer learning for image classification using VGG19: Caltech-101 image data set. *Journal of Ambient Intelligence & Humanized Computing, 14,* 3609–3620. https://doi.org/10.1007/s12652-021-03488-z
[37] Shivakumar, P. G., & Georgiou, P. (2020). Transfer learning from adult to children for speech recognition: Evaluation, analysis and recommendations. *Computer Speech & Language, 63,* 101077. https://doi.org/10.1016/j.csl.2020.101077
[38] Azunre, P. (2021). *Transfer learning for natural language processing.* USA: Manning Publications.
[39] Liu, N., Luo, K., Yuan, Z., & Chen, Y. (2022). A transfer learning method for detecting Alzheimer's disease based on speech and natural language processing. *Frontiers in Public Health, 10,* 772592. https://doi.org/10.3389/fpubh.2022.772592
[40] Fan, C., Lei, Y., Sun, Y., Piscitelli, M. S., Chiosa, R., & Capozzoli, A. (2022). Data-centric or algorithm-centric: Exploiting the performance of transfer learning for improving building energy predictions in data-scarce context. *Energy, 240,* 122775. https://doi.org/10.1016/j.energy.2021.122775
[41] Marín, R., & Chang, V. (2021). Impact of transfer learning for human sperm segmentation using deep learning. *Computers in*

Biology and Medicine, 136, 104687. https://doi.org/10.1016/j.compbiomed.2021.104687

[42] Lu, J., Behbood, V., Hao, P., Zuo, H., Xue, S., & Zhang, G. (2015). Transfer learning using computational intelligence: A survey. *Knowledge-Based Systems, 80,* 14–23. https://doi.org/10.1016/j.knosys.2015.01.010

[43] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., ..., & He, Q. (2021). A comprehensive survey on transfer learning. *Proceedings of the IEEE, 109*(1), 43–76. https://doi.org/10.1109/jproc.2020.3004555

[44] Iman, M., Arabnia, H. R., & Rasheed, K. (2023). A review of deep transfer learning and recent advancements. *Technologies, 11*(2), 40. https://doi.org/10.3390/technologies11020040

[45] 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. https://doi.org/10.1007/s13351-019-8162-6

[46] Hewage, P., Behera, A., Trovati, M., & Pereira, E. (2019). Long-short term memory for an effective short-term weather forecasting model using surface weather data. In *IFIP International Conference on Artificial Intelligence Applications and Innovations,* 382–390. https://doi.org/10.1007/978-3-030-19823-7_32

[47] Hewage, P., Trovati, M., Pereira, E., & Behera, A. (2021). Deep learning-based effective fine-grained weather forecasting model. *Pattern Analysis and Applications, 24*(1), 343–366. https://doi.org/10.1007/s10044-020-00898-1

[48] Chai, C., Maceira, M., Santos-Villalobos, H. J., Venkatakrishnan, S. V., Schoenball, M., Zhu, W., ..., & EGS Collab Team. (2020). Using a deep neural network and transfer learning to bridge scales for seismic phase picking. *Geophysical Research Letters, 47*(16), e2020GL088651. https://doi.org/10.1029/2020gl088651

[49] Ma, J., Li, Z., Cheng, J. C. P., Ding, Y., Lin, C., & Xu, Z. (2020). Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. *Science of the Total Environment, 705,* 135771. https://doi.org/10.1016/j.scitotenv.2019.135771

[50] Kimura, N., Yoshinaga, I., Sekijima, K., Azechi, I., & Baba, D. (2020). Convolutional neural network coupled with a transfer-learning approach for time-series flood predictions. *Water, 12*(1), 96. https://doi.org/10.3390/w12010096

[51] Boonyuen, K., Kaewprapha, P., & Srivihok, P. (2018). Daily rainfall forecast model from satellite image using convolution neural network. In *2018 International Conference on Information Technology,* 1–7. https://doi.org/10.23919/INCIT.2018.8584886

[52] Zhao, P., & Zettsu, K. (2019). Convolution recurrent neural networks based dynamic transboundary air pollution predictiona. In *2019 IEEE 4th International Conference on Big Data Analytics,* 410–413. https://doi.org/10.1109/ICBDA.2019.8712835

[53] Gilik, A., Ogrenci, A. S., & Ozmen, A. (2022). Air quality prediction using CNN+LSTM-based hybrid deep learning architecture. *Environmental Science and Pollution Research, 29*(8), 11920–11938. https://doi.org/10.1007/s11356-021-16227-w

[54] Kong, T., Choi, D., Lee, G., & Lee, K. (2021). Air pollution prediction using an ensemble of dynamic transfer models for multivariate time series. *Sustainability, 13*(3), 1367. https://doi.org/10.3390/su13031367

[55] Dhole, A., Ambekar, I., Gunjan, G., & Sonawani, S. (2021). An ensemble approach to multi-source transfer learning for air quality prediction. In *2021 International Conference on Computing, Communication, and Intelligent Systems,* 70–77. https://doi.org/10.1109/ICCCIS51004.2021.9397138

[56] Gu, Q., Dai, Q., Yu, H., & Ye, R. (2021). Integrating multi-source transfer learning, active learning and metric learning paradigms for time series prediction. *Applied Soft Computing, 109,* 107583. https://doi.org/10.1016/j.asoc.2021.107583

[57] Fong, I. H., Li, T., Fong, S., Wong, R. K., & Tallón-Ballesteros, A. J. (2020). Predicting concentration levels of air pollutants by transfer learning and recurrent neural network. *Knowledge-Based Systems, 192,* 105622. https://doi.org/10.1016/j.knosys.2020.105622

[58] Ma, W., Yuan, Z., Lau, A. K. H., Wang, L., Liao, C., & Zhang, Y. (2022). Optimized neural network for daily-scale ozone prediction based on transfer learning. *Science of the Total Environment, 827,* 154279. https://doi.org/10.1016/j.scitotenv.2022.154279

[59] Sonawani, S., & Patil, K. (2024). Air quality measurement, prediction and warning using transfer learning based IOT system for ambient assisted living. *International Journal of Pervasive Computing and Communications, 20*(1), 38–55. https://doi.org/10.1108/ijpcc-07-2022-0271

[60] Chen, S., Xu, Z., Wang, X., & Zhang, C. (2022). Ambient air pollutants concentration prediction during the COVID-19: A method based on transfer learning. *Knowledge-Based Systems, 258,* 109996. https://doi.org/10.1016/j.knosys.2022.109996

[61] Deng, T., Manders, A., Segers, A., Bai, Y., & Lin, H. (2021). Temporal transfer learning for ozone prediction based on CNN-LSTM model. In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence, 2,* 1005–1012.

[62] Njaime, M., Olivier, F. A., Snoussi, H., Akl, J., Chahla, C., & Omrani, H. (2022). Data cleaning to fine-tune a transfer learning approach for air quality prediction. In *2022 IEEE International Smart Cities Conference,* 1–5. https://doi.org/10.1109/ISC255366.2022.9921836

[63] Tariq, S., Loy-Benitez, J., Nam, K., Lee, G., Kim, M., Park, D., & Yoo, C. (2021). Transfer learning driven sequential forecasting and ventilation control of PM2.5 associated health risk levels in underground public facilities. *Journal of Hazardous Materials, 406,* 124753. https://doi.org/10.1016/j.jhazmat.2020.124753

[64] Yuan, Z., Kerckhoffs, J., Shen, Y., de Hoogh, K., Hoek, G., & Vermeulen, R. (2023). Integrating large-scale stationary and local mobile measurements to estimate hyperlocal long-term air pollution using transfer learning methods. *Environmental Research, 228,* 115836. https://doi.org/10.1016/j.envres.2023.115836

[65] Honarvar, A. R., & Sami, A. (2019). Towards sustainable smart city by particulate matter prediction using urban big data, excluding expensive air pollution infrastructures. *Big Data Research, 17,* 56–65. https://doi.org/10.1016/j.bdr.2018.05.006

[66] Yadav, N., Sorek-Hamer, M., von Pohle, M., Asanjan, A. A., Sahasrabhojanee, A., Suel, E., ..., & Ganguly, A. R. (2024). Using deep transfer learning and satellite imagery to estimate urban air quality in data-poor regions. *Environmental Pollution, 342,* 122914. https://doi.org/10.1016/j.envpol.2023.122914

[67] Yang, J., Ismail, A. W., Li, Y., Zhang, L., & Fadzli, F. E. (2023). Transfer learning-driven hourly PM2.5 prediction

based on a modified hybrid deep learning. *IEEE Access, 11,* 99614–99627. https://doi.org/10.1109/access.2023.3314490
[68] Pang, S., Xie, P., Xu, D., Meng, F., Tao, X., Li, B., ..., & Song, T. (2021). NDFTC: A new detection framework of tropical cyclones from meteorological satellite images with deep transfer learning. *Remote Sensing, 13*(9), 1860. https://doi.org/10.3390/rs13091860
[69] Combinido, J. S., Mendoza, J. R., & Aborot, J. (2018). A convolutional neural network approach for estimating tropical cyclone intensity using satellite-based infrared images. In *2018 24th International Conference on Pattern Recognition,* 1474–1480. https://doi.org/10.1109/ICPR.2018.8545593
[70] Pan, X., Wang, X., Zhao, C., Wu, J., Wang, H., Wang, S., & Chen, S. (2023). USFP: An unbalanced severe typhoon formation prediction framework based on transfer learning. *Frontiers in Marine Science, 9,* 1046964. https://doi.org/10.3389/fmars.2022.1046964
[71] Fu, D., Chang, P., & Liu, X. (2023). Using convolutional neural network to emulate seasonal tropical cyclone activity. *Journal of Advances in Modeling Earth Systems, 15*(10), e2022MS003596. https://doi.org/10.1029/2022ms003596
[72] Notarangelo, N., Hirano, K., Albano, R., & Sole, A. (2021). Transfer learning with convolutional neural networks for rainfall detection in single images. *Water, 13*(5), 588. https://doi.org/10.3390/w13050588
[73] Ambildhuke, G. M., & Banik, B. G. (2021). Transfer learning approach-An efficient method to predict rainfall based on ground-based cloud images. *Ingénierie des Systèmes d'Information, 26*(4), 345–356. https://doi.org/10.18280/isi.260402
[74] Liu, Z., Yang, Q., Shao, J., Wang, G., Liu, H., Tang, X., ..., & Bai, L. (2022). Improving daily precipitation estimation in the data scarce area by merging rain gauge and TRMM data with a transfer learning framework. *Journal of Hydrology, 613,* 128455. https://doi.org/10.1016/j.jhydrol.2022.128455
[75] Hu, W., Luo, Y., Lu, Z., & Wen, Y. (2019). Heterogeneous transfer learning for thermal comfort modeling. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation,* 61–70. https://doi.org/10.1145/3360322.3360843
[76] Natarajan, A., & Laftchiev, E. (2019). A transfer active learning framework to predict thermal comfort. *International Journal of Prognostics and Health Management, 10*(3), 1–13. https://doi.org/10.36001/ijphm.2019.v10i3.2629
[77] Hu, Q., Zhang, R., & Zhou, Y. (2016). Transfer learning for short-term wind speed prediction with deep neural networks. *Renewable Energy, 85,* 83–95. https://doi.org/10.1016/j.renene.2015.06.034
[78] Zhao, G., Pang, B., Xu, Z., Cui, L., Wang, J., Zuo, D., & Peng, D. (2021). Improving urban flood susceptibility mapping using transfer learning. *Journal of Hydrology, 602,* 126777. https://doi.org/10.1016/j.jhydrol.2021.126777
[79] Muñoz, D. F., Muñoz, P., Moftakhari, H., & Moradkhani, H. (2021). From local to regional compound flood mapping with deep learning and data fusion techniques. *Science of the Total Environment, 782,* 146927. https://doi.org/10.1016/j.scitotenv.2021.146927
[80] Tian, W., Wu, J., Cui, H., & Hu, T. (2021). Drought prediction based on feature-based transfer learning and time series imaging. *IEEE Access, 9,* 101454–101468. https://doi.org/10.1109/access.2021.3097353
[81] Li, J., Lo, W. L., Fu, H., & Chung, H. S. H. (2021). A transfer learning method for meteorological visibility estimation based on feature fusion method. *Applied Sciences, 11*(3), 997. https://doi.org/10.3390/app11030997
[82] Li, Q., Tang, S., Peng, X., & Ma, Q. (2019). A method of visibility detection based on the transfer learning. *Journal of Atmospheric and Oceanic Technology, 36*(10), 1945–1956. https://doi.org/10.1175/jtech-d-19-0025.1
[83] Lo, W. L., Chung, H. S. H., & Fu, H. (2021). Experimental evaluation of PSO based transfer learning method for meteorological visibility estimation. *Atmosphere, 12*(7), 828. https://doi.org/10.3390/atmos12070828
[84] Lu, Z., Zheng, C., & Yang, T. (2020). Application of offshore visibility forecast based on temporal convolutional network and transfer learning. *Computational Intelligence and Neuroscience, 2020*(1), 8882279. https://doi.org/10.1155/2020/8882279
[85] Obara, Y., & Nakamura, R. (2022). Transfer learning of long short-term memory analysis in significant wave height prediction off the coast of western Tohoku, Japan. *Ocean Engineering, 266,* 113048. https://doi.org/10.1016/j.oceaneng.2022.113048
[86] Miao, Y., Zhang, X., Li, Y., Zhang, L., & Zhang, D. (2023). Monthly extended ocean predictions based on a convolutional neural network via the transfer learning method. *Frontiers in Marine Science, 9*, 1073377. https://doi.org/10.3389/fmars.2022.1073377
[87] Kumar, N. K., Savitha, R., & Al Mamun, A. (2018). Ocean wave characteristics prediction and its load estimation on marine structures: A transfer learning approach. *Marine Structures, 61,* 202–219. https://doi.org/10.1016/j.marstruc.2018.05.007
[88] Hu, J., Weng, B., Huang, T., Gao, J., Ye, F., & You, L. (2021). Deep residual convolutional neural network combining dropout and transfer learning for ENSO forecasting. *Geophysical Research Letters, 48*(24), e2021GL093531. https://doi.org/10.1029/2021GL093531
[89] Ham, Y. G., Kim, J. H., & Luo, J. J. (2019). Deep learning for multi-year ENSO forecasts. *Nature, 573*(7775), 568–572. https://doi.org/10.1038/s41586-019-1559-7
[90] Mu, B., Ma, S., Yuan, S., & Xu, H. (2020). Applying convolutional LSTM network to predict El Niño events: Transfer learning from the data of dynamical model and observation. In *2020 IEEE 10th International Conference on Electronics Information and Emergency Communication,* 215–219. https://doi.org/10.1109/ICEIEC49280.2020.9152317
[91] Bellenger, H., Guilyardi, E., Leloup, J., Lengaigne, M., & Vialard, J. (2014). ENSO representation in climate models: From CMIP3 to CMIP5. *Climate Dynamics, 42,* 1999–2018. https://doi.org/10.1007/s00382-013-1783-z
[92] Jozinović, D., Lomax, A., Štajduhar, I., & Michelini, A. (2022). Transfer learning: Improving neural network based prediction of earthquake ground shaking for an area with insufficient training data. *Geophysical Journal International, 229*(1), 704–718. https://doi.org/10.1093/gji/ggab488
[93] Titos, M., Bueno, A., García, L., Benítez, C., & Segura, J. C. (2020). Classification of isolated volcano-seismic events based on inductive transfer learning. *IEEE Geoscience and Remote Sensing Letters, 17*(5), 869–873. https://doi.org/10.1109/lgrs.2019.2931063

[94] Wang, K., Johnson, C. W., Bennett, K. C., & Johnson, P. A. (2021). Predicting fault slip via transfer learning. *Nature Communications, 12*(1), 7319. https://doi.org/10.1038/s41467-021-27553-5
[95] Maya, M., & Yu, W. (2019). Short-term prediction of the earthquake through neural networks and meta-learning. In *2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control,* 1–6. https://doi.org/10.1109/ICEEE.2019.8884562
[96] Alem, A., & Kumar, S. (2022). Transfer learning models for land cover and land use classification in remote sensing image. *Applied Artificial Intelligence, 36*(1), 2014192. https://doi.org/10.1080/08839514.2021.2014192
[97] Barman, P., Mustak, S., Kuffer, M., & Singh, S. K. (2023). Transfer-ensemble learning: A novel approach for mapping urban land use/cover of the Indian metropolitans. *Sustainability, 15*(24), 16593. https://doi.org/10.3390/su152416593
[98] Naushad, R., Kaur, T., & Ghaderpour, E. (2021). Deep transfer learning for land use and land cover classification: A comparative study. *Sensors, 21*(23), 8083. https://doi.org/10.3390/s21238083
[99] Zhang, D., Liu, Z., & Shi, X. (2020). Transfer learning on efficientnet for remote sensing image classification. In *2020 5th International Conference on Mechanical, Control and Computer Engineering,* 2255–2258. https://doi.org/10.1109/ICMCCE51767.2020.00489
[100] Dastour, H., & Hassan, Q. K. (2023). A comparison of deep transfer learning methods for land use and land cover classification. *Sustainability, 15*(10), 7854. https://doi.org/10.3390/su15107854
[101] Wu, T., Luo, J., Xia, L., Shen, Z., & Hu, X. (2015). Prior knowledge-based automatic object-oriented hierarchical classification for updating detailed land cover maps. *Journal of the Indian Society of Remote Sensing, 43*(4), 653–669. https://doi.org/10.1007/s12524-014-0446-9
[102] Li, X., Zhang, L., Du, B., Zhang, L., & Shi, Q. (2017). Iterative reweighting heterogeneous transfer learning framework for supervised remote sensing image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10*(5), 2022–2035. https://doi.org/10.1109/JSTARS.2016.2646138
[103] Demir, B., Bovolo, F., & Bruzzone, L. (2013). Updating land-cover maps by classification of image time series: A novel change-detection-driven transfer learning approach. *IEEE Transactions on Geoscience and Remote Sensing, 51*(1), 300–312. https://doi.org/10.1109/tgrs.2012.2195727
[104] Lin, C., Du, P., Samat, A., Li, E., Wang, X., & Xia, J. (2019). Automatic updating of land cover maps in rapidly urbanizing regions by relational knowledge transferring from GlobeLand30. *Remote Sensing, 11*(12), 1397. https://doi.org/10.3390/rs11121397
[105] Yifter, T., Razoumny, Y., & Lobanov, V. (2022). Deep transfer learning of satellite imagery for land use and land cover classification. *Artificial Intelligence, Knowledge and Data Engineering, 21*(5), 963–982. https://doi.org/10.15622/ia.21.5.5
[106] Qian, Y., Zhou, W., Yu, W., Han, L., Li, W., & Zhao, W. (2020). Integrating backdating and transfer learning in an object-based framework for high resolution image classification and change analysis. *Remote Sensing, 12*(24), 4094. https://doi.org/10.3390/rs12244094
[107] van den Broeck, W. A. J., Goedemé, T., & Loopmans, M. (2022). Multiclass land cover mapping from historical orthophotos using domain adaptation and spatio-temporal transfer learning. *Remote Sensing, 14*(23), 5911. https://doi.org/10.3390/rs14235911
[108] Siddamsetty, J., Stricker, M., Charfuelan, M., Nuske, M., & Dengel, A. (2023). Inter-region transfer learning for land use land cover classification. *ISPRS Annals of the Photogrammetry Remote Sensing and Spatial Information Sciences, 10,* 881–888. https://doi.org/10.5194/isprs-annals-x-1-w1-2023-881-2023
[109] Huang, Z., Dumitru, C. O., Pan, Z., Lei, B., & Datcu, M. (2021). Classification of large-scale high-resolution SAR images with deep transfer learning. *IEEE Geoscience and Remote Sensing Letters, 18*(1), 107–111. https://doi.org/10.1109/lgrs.2020.2965558
[110] Niu, S., Liu, Y., Wang, J., & Song, H. (2021).A decade survey of transfer learning (2010–2020). *IEEE Transactions on Artificial Intelligence, 1*(2), 151–166. https://doi.org/10.1109/tai.2021.3054609
[111] Banerjee, M., Mitra, S., & Pal, S. K. (1998). Rough fuzzy MLP: Knowledge encoding and classification. *IEEE Transactions on Neural Networks, 9*(6), 1203–1216. https://doi.org/10.1109/72.728363
[112] Zhang, W., Deng, L., Zhang, L., & Wu, D. (2023). A survey on negative transfer. *IEEE/CAA Journal of Automatica Sinica, 10*(2), 305–329. https://doi.org/10.1109/jas.2022.106004
[113] Zhao, Z., Alzubaidi, L., Zhang, J., Duan, Y., & Gu, Y. (2024). A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations. *Expert Systems with Applications, 242,* 122807. https://doi.org/10.1016/j.eswa.2023.122807
[114] Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification.* USA: Wiley.

How to Cite: Pal, S. K., & Dutta, D. (2024). Transfer Learning in Weather Prediction: Why, How, and What Should. *Journal of Computational and Cognitive Engineering*, *3*(4), 324–347. https://doi.org/10.47852/bonviewJCCE42022817

Appendix

AI -Artificial intelligence
ANN- Artificial neural network
BLSTM/ BDLSTM - Bi-Directional LSTM
CNN Convolutional Neural Network
CNN-LSTM Convolutional Neural Network-LSTM
Conc. Concentration
D-CRNN-Convolutional recurrent neural networks
DBN - Deep belief networks
DL - Deep learning
DNN - Deep Neural Network
ENSO - El Niño-Southern Oscillation
EVS - Explained variance score
GANs - Generative adversarial networks
GRU - Gated recurrent unit
HIW - High-impact weather
LCLU - land cover land use
LSTM - long short-term memory
MAE - Mean Absolute Error
MAPE- Mean Absolute Percent Error
MSE - Mean square Error
ML- Machine learning
MLP - Multilayer Perceptron
*n*RMSE - normalized root mean squared error
NO_2 - Nitrogen dioxide
NWP - Numerical Weather Prediction
O_3 - Ozone
PM- Particulate matter
PM2.5 - Particulate matter less than 2.5 µm in diameter
PM10 - Particulate matter less than 10 µm in diameter
RF - Random Forest
RMSE - Root Mean Squared Error
RNN - Recurrent Neural Network
RGB - Red–Green–Blue
SMAPE - Symmetric mean absolute percentage error
SO_2 - Sulfur dioxide
SSTA - Sea surface temperature anomalies
SSHA - sea surface height anomalies
SVR - Support vector regression
TC - Tropical cyclone
TL - Transfer Learning
USFP - Unbalanced severe typhoon formation prediction
VHR - Very-high-resolution
R^2 - Coefficient of determination