Survey paper

Ensemble deep learning: A review

M.A. Ganaie [a], Minghui Hu [b], A.K. Malik [a], M. Tanveer [a,*], P.N. Suganthan [b,c,*]

[a] *Department of Mathematics, Indian Institute of Technology Indore, Simrol, Indore, 453552, India*
[b] *School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore*
[c] *KINDI Center for Computing Research College of Engineering, Qatar University, Qatar*

ARTICLE INFO

ABSTRACT

Ensemble learning combines several individual models to obtain better generalization performance. Currently, deep learning architectures are showing better performance compared to the shallow or traditional models. Deep ensemble learning models combine the advantages of both the deep learning models as well as the ensemble learning such that the final model has better generalization performance. This paper reviews the state-of-art deep ensemble models and hence serves as an extensive summary for the researchers. The ensemble models are broadly categorized into bagging, boosting, stacking, negative correlation based deep ensemble models, explicit/implicit ensembles, homogeneous/heterogeneous ensemble, decision fusion strategies based deep ensemble models. Applications of deep ensemble models in different domains are also briefly discussed. Finally, we conclude this paper with some potential future research directions.

1. Introduction

Deep learning architectures have been successfully employed across a wide range of applications from image/video classification to the health care. The success of these models is attributed to the better feature representation via multi layer processing architectures. The deep learning models have been mainly used for classification, regression and clustering problems. Classification problem is defined as the categorization of the new observations based on the hypothesis h learned from the set of training data. The hypothesis h represents a mapping of input data features to the appropriate target labels/classes. The main objective, while learning the hypothesis h, is that it should approximate the true unknown function as close as possible to reduce the generalization error. There exist several applications of these classification algorithms ranging from medical diagnosis to remote sensing. Mathematically,

$$O_c = h(x, \theta_c), \ O_c \in \mathbb{Z}, \tag{1}$$

where x is the input feature vector, O_c is the category of the sample x, θ_c is the set of learning parameters of the hypothesis h and $\mathbb{Z}$ is the set of class labels.

Regression problems deal with the continuous decisions, instead of discrete categories. Mathematically,

$$O_r = h(x, \theta_r), \ O_r \in \mathbb{R}, \tag{2}$$

where x is the observation vector, O_r is the output, and θ_r is the set of learning parameters of the hypothesis h.

Broadly speaking, there are different approaches of classification like supervised, unsupervised classification, few-shot, one-shot and so on. Here, we only discuss supervised and unsupervised classification problems. In supervised learning, the building of hypothesis h is supervised based on the known output labels provided in the training data samples, while as in unsupervised learning hypothesis h is generated without any supervision as no known output values are available with the training data. This approach, also known as clustering, generates the hypothesis h based on the similarities and dissimilarities present in the training data.

Generally speaking, the goal of generating the hypothesis h in Machine learning area is that it should perform better when applied to unknown data. The performance of the model is measured with respect to the area in which the model is applied. Combining the predictions from several models has proven to be an elegant approach for increasing the performance of the models. Combination of several different predictions from different models to make the final prediction is known as ensemble learning or ensemble model. The ensemble learning involves multiple models combined in some fashion like averaging, voting such that the ensemble model is better than any of the individual models. To prove that average voting in an ensemble is better than individual model, Marquis de Condorcet proposed a theorem wherein he proved that if the probability of each voter being correct is above 0.5 and the voters are independent, then addition of more voters increases the probability of majority vote being correct until it approaches 1 (Condorcet, 1785). Although Marquis de Condorcet

proposed this theorem in the field of political science and had no idea of the field of Machine learning, but it is the similar mechanism that leads to better performance of the ensemble models. Assumptions of Marquis de Condorcet theorem also holds true for ensembles (Hansen and Salamon, 1990). The reasons for the success of ensemble learning include: statistical, computational and representation learning (Dietterich, 2000), bias–variance decomposition (Kohavi and Wolpert, 1996) and strength-correlation (Breiman, 2001).

In this era of machine learning, deep learning automates the extraction of high-level features via hierarchical feature learning mechanism wherein the upper layer of features are generated on the previous set of layer/layers. Deep learning has been successfully applied across different fields since the ImageNet Large Scale Recognition Challenge (ILSVRC) competitions (Russakovsky et al., 2015; Krizhevsky et al., 2012) and has achieved state-of-art performance. It has obtained promising results in object detection, semantic segmentation, edge detection and number of other domains. However, given the computational cost, the training of deep ensemble models is an uphill task. Different views have been provided to understand how the deep learning models learn the features like learning through hierarchy of concepts via many levels of representation (Deng and Yu, 2014; Goodfellow et al., 2016; LeCun et al., 2015). Given the advantages of deep learning models from deep architectures, there are several bottlenecks like vanishing/exploding gradients (Hochreiter, 1991; Glorot and Bengio, 2010) and degradation problem (He et al., 2016) which prevent to reach this goal. Recently, training deep network's has become feasible through the Highway networks (Srivastava et al., 2015) and Residual networks (He et al., 2016). Both these networks enabled to train very deep networks. The ensemble learning has been recently known to be strong reason for enhancing the performance of deep learning models (Veit et al., 2016). Thus, the objective of deep ensemble models is to obtain a model that has best of both the ensemble and deep models.

There exist multiple surveys in the literature which mainly focus on the review of ensemble learning like learning of ensemble models in classification problems (Zhao et al., 2005; Rokach, 2010; Gopika and Azhagusundari, 2014; Yang et al., 2010), regression problems (Mendes-Moreira et al., 2012; Ren et al., 2015) and clustering (Vega-Pons and Ruiz-Shulcloper, 2011). Review of both the classification and regression models was given in Ren et al. (2016). Comprehensive review of the ensemble methods and the challenges were given in Sagi and Rokach (2018). Though Sagi and Rokach (2018) provided some insight about the deep ensemble models but could not give the comprehensive review of the deep ensemble learning while as Cao et al. (2020) reviewed the ensemble deep models in the context of bioinformatics. The past decade has successively evolved different deep learning strategies which have lead to the exploration and innovation of these models in multiple areas like health care, speech, image classification, forecasting and other applications. Broadly speaking, ensemble learning approaches have followed classical methods, general methods and different fusion strategies for improving the performance of the models. Since deep learning models are computation and data extensive, hence, ensemble deep learning models need special attention while exploring the complementary information of multiple algorithms into a uniform framework. Ensemble deep learning models need to handle multiple questions like how to induce diversity among the baseline models, how to keep the training time as well the models complexity lower for the practical applications, how to fuse the predictions of the complementary algorithms. Multiple studies have handled these problems differently. In this review paper, we comprehensively review the different approaches used to handle the aforementioned problems. In this paper, we give a comprehensive review of deep ensemble models. **To the best of our knowledge, this is the first comprehensive review paper on deep ensemble models.**

The rest of this paper is organized as follows: Section 3 discusses the theoretical aspects of deep ensemble learning, Section 4 discusses the different approaches used in deep ensemble strategies, applications of deep ensemble methods are given in Section 5 and finally conclusions and future directions are given in Section 6 (see Fig. 1).

2. Research methodology

The studies in this review are searched from the Google Scholar and Scopus search engines. The papers are the result of ensemble learning, ensemble deep learning, deep ensemble learning, deep ensembles keywords. The articles where screened based on the title and abstract, followed by the screening of full-text version. The articles are elaborated based on the ensemble learning and deep learning approaches.

3. Theory

The various reasons which have been justified for the success of ensemble learning can be discussed under the following subheadings:

3.1. Bias–variance decomposition

Initially, the success of ensemble methods was theoretically investigated for regression problems. Krogh and Vedelsby (1995) and Brown et al. (2005a) proved via ambiguity decomposition that the proper ensemble classifier guarantees a smaller squared error as compared to the individual predictors of the classifier. Ambiguity decomposition was given for single dataset based ensemble methods, later on, multiple dataset bias–variance–covariance decomposition was introduced in Brown et al. (2005a), Geman et al. (1992), Brown et al. (2005b) and Pedro (2000) and is given as:

$$
\begin{aligned}
E[o - t]^2 &= bias^2 + \frac{1}{M}var + (1 - \frac{1}{M})covar, \\
bias &= \frac{1}{M}\sum_i (E[o_i] - t), \\
var &= \frac{1}{M}\sum_i E[o_i - E[o_i]]^2, \\
covar &= \frac{1}{M(M-1)}\sum_i \sum_{j \neq i} E[o_i - E[o_i]][o_j - E[o_j]],
\end{aligned}
\tag{3}
$$

where t is target, o_i is the output of ith model and M is the ensemble size. Here, $bias$ term measures the average difference between the base learner and the model output, var indicates their average variance, and $covar$ is the covariance term measuring the pairwise difference of the base learners.

Ensemble methods have been supported by several theories like bias–variance (Kohavi and Wolpert, 1996; Wolpert, 1997), strength correlation (Breiman, 2001), stochastic discrimination (Kleinberg, 1990), and margin theory (Schapire et al., 1998). These theories provide the equivalent of bias–variance–covariance decomposition (Pisetta, 2012).

The above given equations of decomposition error cannot be directly applied to the datasets with discrete class labels due to their categorical nature. However, alternate ways to decompose the error in classification problems are given in Kohavi and Wolpert (1996), Kong and Dietterich (1995), Friedman (1997), Breiman (1998) and James (2003).

Multiple approaches like bagging, boosting have been proposed for generating the ensemble methods. Bagging reduces the variance among the base classifiers (Breiman, 1996b) while as boosting based ensembles lead to the bias and variance reduction (Breiman, 1996a; Zhang and Zhang, 2008).

3.2. Statistical, computational and representational aspects

Dietterich provided Statistical, Computational and Representational reasons (Dietterich, 2000) for success of ensemble models. The learning model is viewed as the search of the optimal hypothesis h among the several hypothesis in the search space. When the amount of data available for the training is smaller compared to the size of the hypothesis space, the statistical problem arises. Due to this statistical problem, the learning algorithm identifies the different hypothesis which gives same
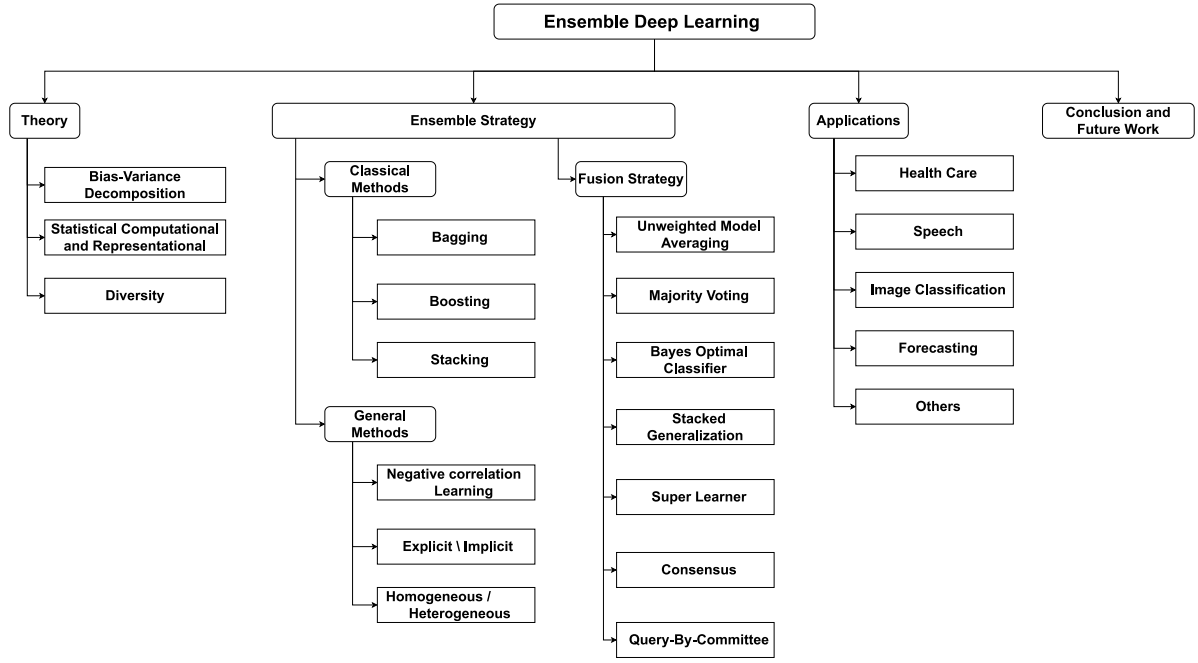
Fig. 1. Layout of the paper.

performance on the training samples. Ensembling of these hypothesis results in an algorithm which reduces the risk of being a wrong classifier. The second reason is computational wherein a learning algorithm stucks in a local optima due to some form of local search. Ensemble model overcomes this issue by performing some form of local search via different starting points which leads to better approximation of the true unknown function. Another reason is representational wherein none of the hypotheses among the set of hypothesis is able to represent the true unknown function. Hence, ensembling of these hypothesis via some weighting technique results into the hypothesis which expands the representable function space.

3.3. Diversity

One of the main reasons behind the success of ensemble methods is increasing the diversity among the base classifiers and the same thing was highlighted in Dietterich (2000). Different approaches have been followed to generate diverse classifiers. Different methods like bootstrap aggregation (bagging) (Breiman, 1996b), Adaptive Boosting (AdaBoost) (Freund and Schapire, 1996), random subspace (Barandiaran, 1998), and random forest (Breiman, 2001) approaches are followed for generating the multiple datasets from the original dataset to train the different predictors such that the outputs of predictors are diverse. Attempts have been made to increase diversity in the output data wherein multiple outputs are created instead of multiple datasets for the supervision of the base learners. 'Output smearing' (Breiman, 2000) is one of this kind which induces random noise to introduce diversity in the output space.

4. Ensemble strategies

The different ensemble strategies have evolved over a period of time which results in better generalization of the learning models. The ensemble strategies are broadly categorized as follows:

4.1. Bagging

Bagging (Breiman, 1996b), also known as bootstrap aggregating, is one of the standard techniques for generating the ensemble-based
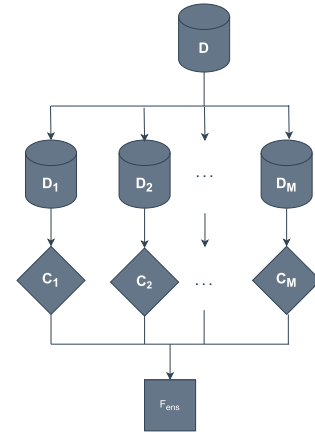

Fig. 2. Bagging.

algorithms. Bagging is applied to enhance the performance of an ensemble classifier. The main idea in bagging is to generate a series of independent observations with the same size, and distribution as that of the original data. Given the series of observations, generate an ensemble predictor which is better than the single predictor generated on the original data. Bagging increases two steps in the original models: First, generating the bagging samples and passing each bag of samples to the base models and second, strategy for combining the predictions of the multiple predictors. Bagging samples may be generated with or without replacement. Combining the output of the base predictors may vary as mostly majority voting is used for classification problems while the averaging strategy is used in regression problems for generating the ensemble output. Fig. 2 shows the diagram of the bagging technique. Here, D_i represents the bagged datasets, C_i represents the algorithms and F_{ens} calculates the final outcome.

Random Forest (Breiman, 2001) is an improved version of the decision trees that uses the bagging strategy for improving the predictions of the base classifier which is a decision tree. The fundamental difference between these two methods is that at each tree split in Random Forest, only a subset of features is randomly selected and
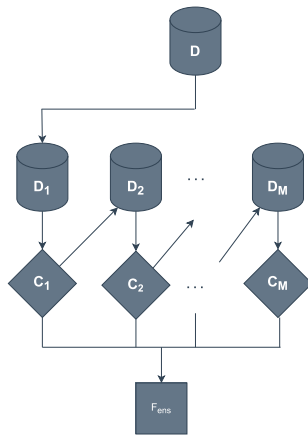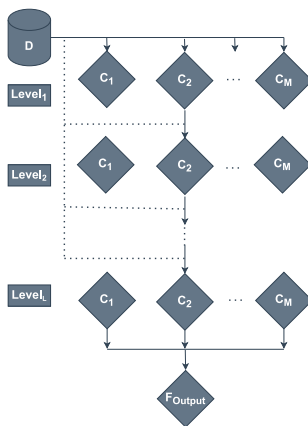
Fig. 3. Boosting.

Fig. 4. Stacking.

considered for splitting. The purpose of this method is to decorrelate the trees and prevent over-fitting. Breiman (2001) showed heuristically that the variance of the bagged predictor is smaller than the original predictor and proposed that bagging is better in higher dimensional data. However, the analysis of the smoothing effect of bagging (Buja and Stuetzle, 2000) revealed that bagging does not depend on the data dimensionality.

Bühlmann and Yu (2002) gave theoretical explanation of how bagging gives smooth hard decisions, small variance, and mean squared error. Since bagging is computationally expensive, hence subbagging and half subbagging (Bühlmann and Yu, 2002) were introduced. Half subbagging, being computationally efficient, is as accurate as the bagging.

Several attempts tried to combine bagging with other machine learning algorithms. Kim et al. (2002) used bagging method to generate multiple bags of the dataset and multiple support vector machines were trained independently with each bag as the input. The output of the models is combined via majority voting, least squares estimation weighting and double layer hierarchical approach. In the double layer hierarchical approach, another support vector machines (SVM) is used to combine the outcomes of the multiple SVM's efficiently. Tao et al. (2006) used asymmetric bagging strategy to generate the ensemble model to handle the class imbalance problems. A case study of bagging, boosting and basic ensembles (Mao, 1998) revealed that at higher rejection rates of samples boosting is better as compared to bagging and basic ensembles. However, as the rejection rate increases the difference disappears among the boosting, bagging and basic ensembles. Bagging based multilayer perceptron (Ha et al., 2005) combined bagging to

train multiple perceptrons with the corresponding bag and showed that bagging based ensemble models perform better as compared to individual multilayer perceptron. In Gençay and Qi (2001), the analysis of the bagging approach and other regularization techniques revealed that bagging regularized the neural networks and hence provide better generalization. In Khwaja et al. (2015), bagged neural networks (BNNs) was proposed wherein each neural network was trained over different dataset sampled randomly with replacement from original dataset and was implemented for the short term load forecasting. Unlike Random forest (Breiman, 2001) which uses majority voting for aggregating the ensemble of decision trees, bagging based survival trees (Hothorn et al., 2004) used Kaplan–Meier curve to predict the ensemble output for breast cancer and lymphoma patients. In Alvear-Sandoval and Figueiras-Vidal (2018), ensembles of stacked denoising autoencoders for classification showed that the bagging and switching technique in a general deep machine results in improved diversity.

Bagging has also been applied to solve the problem of imbalanced data. Roughly Balanced Bagging (Hido et al., 2009) tries to equalize each class's sampling probability in binary class problems wherein the negative class samples are sampled via negative binomial distribution, instead of keeping the sample size of each class the same number. Neighborhood Balanced Bagging (Błaszczyński and Stefanowski, 2015) incorporated the neighborhood information for generating the bagging samples for the class imbalance problems. Błaszczyński and Stefanowski (2015) concluded that applying conventional diversification is more effective when applied at the last classification methods. Both roughly balanced Bagging and Neighborhood Balanced Bagging have not been explored in deep learning architectures. Thus, these approaches can be exploited to handle the class imbalance problems via deep ensemble models.

The theoretical and experimental analysis of online bagging and boosting (Oza, 2005) showed that the online bagging algorithm can achieve similar accuracy as the batch bagging algorithm with only a little more training time. However, online bagging is an option when all training samples cannot be loaded into the memory due to memory issues.

Although ensembling may lead to increase in the computational complexity, but bagging possesses the property that it can be paralleled and can lead to effective reduction in the training time subject to the availability of hardware for running the parallel models. Since deep learning models have high training time, hence optimization of multiple deep models on different training bags is not a feasible option (see Table 1).

4.2. Boosting

Boosting technique is used in ensemble models for converting a weak learning model into a learning model with better generalization. Fig. 3 shows the diagram of the boosting technique. The techniques such as majority voting in case of classification problems or a linear combination of weak learners in the regression problems results in better prediction as compared to the single weak learner. Boosting methods like AdaBoost (Freund and Schapire, 1996) and Gradient Boosting (Friedman, 2001) have been used across different domains. AdaBoost uses a greedy technique for minimizing a convex surrogate function upper bounded by misclassification loss via augmentation, at each iteration, the current model with the appropriately weighted predictor. AdaBoost learns an effective ensemble classifier as it leverages the incorrectly classified sample at each stage of the learning. AdaBoost minimizes the exponential loss function while as the Gradient boosting generalized this framework to the arbitrary differential loss function.

Boosting, also known as forward stagewise additive modeling, was originally proposed to improve the performance of the classification trees. It has been recently incorporated in the deep learning models to further improve their performance.

Boosted deep belief network (DBN) (Liu et al., 2014) for facial expression recognition unified the boosting technique and multiple

Table 1
Bagging based ensemble models.

Years	Authors	Contribution
1996	Breiman (1996b)	Proposed the idea of Bagging
1998	Mao (1998)	Case study of bagging, boosting and basic ensembles
2000	Buja and Stuetzle (2000)	Theoretical analysis of bagging
2001	Breiman (2001)	Bagging with random subspace Decision trees and ensembling outputs via majority voting
2001	Gençay and Qi (2001)	Study of Bayesian regularization, early stopping and Bagging
2002	Kim et al. (2002)	Bagging with SVM's and ensembling outputs via SVM's, majority voting and least squares estimation
2002	Bühlmann and Yu (2002)	Theoretical justification of Bagging, proposed subbagging and half subagging
2004	Hothorn et al. (2004)	Bagging with decision trees and ensembling outputs via Kaplan–Meier curve
2005	Oza (2005)	Theoretical and experimental analysis of online bagging and boosting
2006	Tao et al. (2006)	Proposed asymmetric bagging with SVM's and ensembling outputs SVM's
2009	Hido et al. (2009)	Roughly balanced bagging on decision trees and ensembling outputs via majority voting
2005, 2015	Ha et al. (2005), Khwaja et al. (2015)	Bagging with Neural networks and ensembling outputs via majority voting
2015	Błaszczyński and Stefanowski (2015)	Neighborhood balanced bagging ensembling outputs via majority voting

DBN's via objective function which results in a strong classifier. The model learns complex feature representation to build a strong classifier in an iterative manner. Deep boosting (Cortes et al., 2014) is an ensemble model that uses the deep decision trees. It can also be used in combination with any other rich family classifier to improve the generalization performance. In each stage of the deep boosting, the decisions of which classifier to add and what weights should be chosen depends on the (data-dependent) complexity of the classifier to which it belongs. The interpretation of the deep boosting classifier is given via structural risk minimization principle at each stage of the learning. Multiclass Deep boosting (Kuznetsov et al., 2014) extended the Deep boosting (Cortes et al., 2014) algorithm to theoretical, algorithmic, and empirical results to the multiclass problems. Due to the limitation of the training data in each mini batch, Boosting CNN may overfit the data. To avoid overfitting, incremental Boosting CNN (IBCNN) (Han et al., 2016) accumulated the information of multiple batches of the training data samples. IBCNN uses decision stumps on the top of single neurons as the weak learners and learns weights via AdaBoost method in each mini batch. Unlike DBN (Liu et al., 2014) which uses image patch for learning the weak classifiers, IBCNN trains the weak classifiers from the fully connected layer i.e. the whole image is used for learning the weak classifiers. To make the IBCNN model more efficient, the weak learners loss functions are combined with the global loss function.

Boosted CNN (Moghimi et al., 2016) used boosting for training the deep CNN. Instead of averaging, least squares objective function was used to incorporate the boosting weights into CNN. Moghimi et al. (2016) also showed that CNN can be replaced by network structure within their boosting framework for improving the performance of the base classifier. Boosting increases the complexity of training the networks, hence the concept of dense connections was introduced in a deep boosting framework to overcome the problem of vanishing gradient problem for image denoising (Chen et al., 2018b). Deep boosting framework was extended to image restoration in Chen et al. (2019b) wherein the dilated dense fusion network was used to boost the performance.

The convolutional channel features (Yang et al., 2015) generated the high level features via CNN and then used boosted forest for final classification. Since CNN has high number of hyperparameters than the boosted forest, hence the model proved to be efficient than end-to-end training of CNN models both in terms of performance and time. Yang et al. (2015) showed its application in edge detection, object proposal generation, pedestrian and face detection. A stagewise boosting deep CNN (Walach and Wolf, 2016) trains several models of the CNNs within the offline paradigm boosting framework. To extend the concept of boosting in online scenario's wherein only a chunk of data is available at given time, Boosting Independent Embeddings Robustly (BIER) (Opitz et al., 2017) was proposed to cope up the online scenario's. In BIER, a single CNN model is trained end-to-end with an online boosting technique. The training set in the BIER is reweighed via the negative gradient of the loss function to project the input spaces (images) into a collection of independent output spaces. To make BIER more robust, Hierarchical Boosted deep metric learning (Waltner et al., 2019) incorporated the hierarchical label information into the embedding ensemble which improves the performance of the model on the large scale image retrieval application. Using deep boosting results in higher training time, to reduce the warm-up phase of training which trains the classifier from scratch deep incremental boosting (Mosca and Magoulas, 2017) used transfer learning approach. This approach leveraged the initial warm-up phase of each incremental base model of the ensemble during the training of the network. To reduce the training time of boosting based ensembles, snapshot boosting (Zhang et al., 2020a) combined the merits of snapshot ensembling and boosting to improve the generalization without increasing the cost of training. Snapshot boosting trains each base network and combines the outputs via meta learner to combine the output of base learners more efficiently.

Literature shows that the boosting concept is the backbone behind well-known architectures like Deep Residual networks (He et al., 2016; Siu, 2019), AdaNet (Cortes et al., 2017) . The theoretical background for the success of the Deep Residual networks (DeepResNet) (He et al., 2016) was explained in the context of boosting theory (Huang et al., 2018). The authors proposed multi-channel telescoping sum boosting learning framework, known as BoostResNet, wherein each channel is a scalar value updated during rounds of boosting to minimize the multiclass error rate. The fundamental difference between the AdaNet and BoostResnet is that the former maps the feature vectors to classifier space and boosts weak classifiers while the latter used multi-channel representation boosting. Moreover, BoostResNet is more efficient than DeepResnet in terms of computational time.

The theory of boosting was extended to online boosting in Beygelzimer et al. (2015) and provided theoretical convergence guarantees. Online boosting shows improved convergence guarantees for batch boosting algorithms.

The ensembles of bagging and boosting have been evaluated in González et al. (2020). The study evaluated the different algorithms based on the concept of bagging and boosting along with the availability of software tools. The study highlighted the practical issues and opportunities of their feasibility in ensemble modeling (see Table 2).

4.3. Stacking

Ensembling can be done either by combining outputs of multiple base models in some fashion or using some method to choose the "best" base model. Fig. 4 shows the stacking technique. Stacking is one of the integration techniques wherein the meta-learning model is used to integrate the output of base models. If the final decision part is a linear

Table 2
Boosting based ensemble models.

Years	Authors	Contribution
2014	Liu et al. (2014)	Boosted deep belief network (DBN) as base classifiers for facial expression recognition.
2014	Cortes et al. (2014)	Decision trees as base classifiers for binary class classification problems.
2014	Kuznetsov et al. (2014)	Decision trees as base classifiers for multiclass classification problems.
2015	Yang et al. (2015)	Ensemble of CNN and boosted forest for edge detection, object proposal generation, pedestrian and face detection.
2016	Moghimi et al. (2016)	Boosted CNN
2016	Walach and Wolf (2016)	CNN Boosting applied to bacterila cell images and crowd counting.
2017	Opitz et al. (2017)	Boosted deep independent embedding model for online scenarios.
2017	Mosca and Magoulas (2017)	Transfer learning based deep incremental boosting.
2017	Han et al. (2016)	Boosting based CNN with incremental approach for facial action unit recognition.
2018	Chen et al. (2018b)	Deep boosting for image denoising with dense connections.
2019	Chen et al. (2019b)	Deep boosting for image restoration and image denoising.
2019	Waltner et al. (2019)	Hierarchical boosted deep metric learning with hierarchical label embedding.
2020	Zhang et al. (2020a)	Snapshot boosting.

model, the staking is often referred to as "model blending" or simply "blending". The concept of stacking or stacked regression was initially given by Wolpert (1992). In this technique, the dataset is randomly split into J equal parts. For the jth-fold cross-validation one set is used for testing and the rest are used for training. With these training testing pair subsets, we obtain the predictions of different learning models which are used as the meta-data to build the meta-model. Meta-model makes the final prediction, which is also called the winner-takes-all strategy.

Stacking is a bias reducing technique (Leblanc and Tibshirani, 1996). Following Wolpert (1992), Deep convex net (DCN) (Deng and Yu, 2011) was proposed which is a deep learning architecture composed of a variable number of modules stacked together to form the deep architecture. Each learning module in DCN is convex. DCN is a stack of several modules consisting of linear input units, hidden layer non-linear units, and the second linear layer with the number of units as that of target classification classes. The modules are connected layerwise as the output of the lower module is given as input to the adjacent higher module in addition to the original input data. The deep stacking network (DSN) enabling parallel training on very large scale datasets was proposed in Deng et al. (2012c), the network was named stacking based as it shared the concept of "stacked generalization" (Wolpert, 1992). The kernelized version of DCN, known as kernel deep convex networks (K-DCN), was given in Deng et al. (2012a), here the number of hidden layer approach infinity via kernel trick. Deng et al. (2012a) showed that K-DCN performs better as compared to the DCN. However, due to kernel trick the memory requirements increase and hence may not be scalable to large scale datasets. Also, we need to optimize the hyperparameters like the number of levels in the stacked network, the kernel parameters to get the optimal performance of the network. To leverage the memory requirements, random Fourier feature-based kernel deep convex network (Huang et al., 2013) approximated the Gaussian kernel which reduces the training time and helps in the evaluation of K-DCN over large scale datasets. A framework for parameter estimation and model selection in kernel deep stacking networks (Welchowski and Schmid, 2016) is based on the combination of model-based optimization and hill-climbing approaches. Welchowski and Schmid (2016) used data-driven framework for parameter estimation, hyperparameter tuning and model selection in kernel deep stacking networks. Another improvement over DSN was Tensor Deep Stacking Network (T-DSN) (Hutchinson et al., 2012), here in each block of the stacked network, large single hidden layer was split into two smaller ones and then mapped bilinearly to capture the higher-order interactions among the features. Comprehensive evaluation, the more detailed analysis of the learning algorithm and T-DSN implementation is given in Hutchinson et al. (2013). Sparse coding is another popular method that is used in the deep learning area. The advantage of sparse representation is numerous, including robust to noise, effective for learning useful features, etc. Sparse Deep Stacking Network (S-DSN)

is proposed for image classification and abnormal detection (Li et al., 2015; Sun et al., 2018). Li et al. (2015) and Sun et al. (2018) stacked many sparse simplified neural network modules (SNNM) with mixed-norm regularization, in which weights are solved by using the convex optimization and the gradient descent algorithm. In order to make sparse SNNM learning the local dependencies between hidden units, Li et al. (2017a) split the hidden units or representations into different groups, which is termed as group sparse DSN (GS-DSN). The DSN idea is also utilized in the Deep Reinforcement Learning field. Zhang et al. (2020b) employed DSN method to integrate the observations from the formal network: Grasp network and Stacking network based on Q-learning algorithm to make an integrated robotic arm system do grasp and place actions. Wang et al. (2020) stacked blocks multiple times to increase the performance of the neural architecture search task. Zhang et al. (2019a) presents a deep hierarchical multi-patch network for image deblurring via stacking approach.

Since there is no temporal representation of the data in DSNs, they are less effective to the problems where temporal dependencies exist in the input data. To embed the temporal information in DSNs, Recurrent Deep Stacking Networks (R-DSNs) (Palangi et al., 2014) combined the advantages of DSNs and Recurrent neural networks (RNN). Unlike RNN which uses Back Propagation through time for training the network, R-DSNs use Echo State Network (ESN) to initialize the weights and then fine-tuning them via batch-mode gradient descent. A stacked extreme learning machine was proposed in Zhou et al. (2015). Here, at each level of the network ELM with the reduced number of hidden nodes was used to solve the large scale problems. The number of hidden nodes was reduced via the principal component analysis (PCA) reduction technique. Keeping in view the efficiency of stacked models, the number of stacked models based on support vector machine have been proposed (Wang et al., 2019d,a; Li et al., 2019b). Traditional models like Random Forests have also been extended to deep architecture, known as deep forests (Zhou and Feng, 2017), via stacking concept.

In addition to DSNs, there are some novel network architectures proposed based on the stacking method, Low et al. (2019) contributed a stacking-based deep neural network (S-DNN) which is trained without a backpropagation algorithm. Kang et al. (2020) presented a model by stacking conditionally restricted Boltzmann machine and deep neural network, which achieved significant superior performance with fewer parameters and fewer training samples.

4.4. Negative correlation based deep ensemble methods

Negative correlation learning (NCL) (Liu and Yao, 1999) is an important technique for training the learning algorithms. The main concept behind the NCL is to encourage diversity among the individual models of the ensemble to learn the diverse aspects of the training data. NCL minimizes the empirical risk function of the ensemble model via minimization of error functions of the individual networks. NCL (Liu

and Yao, 1999) was evaluated for regression as well as classification tasks. The evaluation used different measures like simple averaging and winner-takes-all measures on classification tasks and simple average combination methods for regression problems. The authors figured out that winner-takes-all is better as compared to simple averaging in NCL ensemble models.

Shi et al. (2018) proposed deep negative correlation learning architecture for crowd counting known as D-ConvNet i.e. decorrelated convolutional networks. Here, counting is done based on regression-based ensemble learning from a pool of convolutional feature mapped weak regressors. The main idea behind this is to introduce the NCL concept in deep architectures. Robust regression via deep NCL (Zhang et al., 2019b) is an extension of Shi et al. (2018) in which theoretical insights about the Rademacher complexity are given and extended to more regression-based problems.

Buschjäger et al. (2020) formulated a generalized bias–variance decomposition method to control the diversity and smoothly interpolates. They present the Generalized Negative Correlation Learning (GNCL) algorithm, which can encapsulate many existing works in literature and achieve superior performance.

The NCL can also be employed for incremental learning tasks. Muhlbaier and Polikar (2007) employed a dynamically modified weighted majority voting strategy to combine the sub-classifiers. Tang et al. (2009) proposed a negative correlation learning (NCL) based approach for ensemble incremental learning.

4.5. Explicit/implicit ensembles

Ensembling of deep neural networks does not seem to be an easy option as it may lead to increase in computational cost heavily due to the training of multiple neural networks. High performance hardware's with GPU acceleration may take weeks of weeks to train the deep networks. Implicit/Explicit ensembles obtain the contradictory goal wherein a single model is trained in such a manner that it behaves like ensemble of training multiple neural networks without incurring additional cost or to keep the additional cost as minimum as possible. Here, the training time of an ensemble is same as the training time of a single model. In implicit ensembles, the model parameters are shared and the single unthinned network at test times approximates the model averaging of the ensemble models. However, in explicit ensembles model parameters are not shared and the ensemble output is taken as the combination of the predictions of the ensemble models via different approaches like majority voting, averaging and so on.

Dropout (Srivastava et al., 2014) creates an ensemble network by randomly dropping out hidden nodes from the network during the training of the network. During the time of testing, all nodes are active. Dropout provides regularization of the network to avoid overfitting and introduces sparsity in the output vectors. Overfitting is reduced as it trains exponential number of models with shared weights and provides an implicit ensemble of networks during testing. Dropping the units randomly avoids coadaptation of the units by making the presence of a particular unit unreliable. The network with dropout takes $2-3$ times more time for training as compared to a standard neural network. Hence, a balance is to be set appropriately between the training time of the network and the overfitting. Generalization of DropOut was given in DropConnect (Wan et al., 2013). Unlike DropOut which drops each output unit, DropConnect randomly drops each connection and hence, introduces sparsity in the weight parameters of the model. Similar to DropOut, DropConnect creates an implicit ensemble during test time by dropping out the connections (setting weights to zero) during training. Both DropOut and DropConnect suffer from high training time. To alleviate this problem, deep networks with Stochastic depth (Huang et al., 2016b) aimed to reduce the network depth during training while keeping it unchanged during testing of the network. Stochastic depth is an improvement on ResNet (He et al., 2016) wherein residual blocks are randomly dropped during training and bypassing these transformation blocks connections via skip connections. Swapout (Singh et al.,

2016) is a generalization of DropOut and Stochastic depth. Swapout involves dropping of individual units or to skip the blocks randomly. Embarking on a distinctive approach of reducing the test time, distilling the knowledge in a network (Hinton et al., 2015) transferred the "knowledge" from ensembles to a single model. Gradual DropIn or regularized DropIn Smith et al. (2016) of layers starts from a shallow network wherein the layers are added gradually. DropIN trains the exponential number of thinner networks, similar to DropOut, and also shallower networks.

All the aforementioned methods provided an ensemble of networks by sharing the weights. There have been attempts to explore explicit ensembles in which models do not share the weights. Snapshot ensembling (Huang et al., 2017a) develops an explicit ensemble without sharing the weights. The authors exploited good and bad local minima and let the stochastic gradient descent (SGD) converge M-times to local minima along the optimization path and take the snapshots only when the model reaches the minimum. These snapshots are then ensembled by averaging at multiple local minima for object recognition. The training time of the ensemble is the same as that of the single model. The ensemble out is taken as the average of the output of the snapshot outputs at multiple local minimas. Random vector functional link network (Pao et al., 1994; Malik et al., 2022) has also been explored for creating the explicit ensembles (Shi et al., 2021) where different random initialization of the hidden layer weights in a hierarchy diversifies the ensemble predictions.

Explicit/implicit produce ensembles out of a single network at the expense of base model diversity (Cao et al., 2020) as the lower level features across the models are likely to be the same. To alleviate this issue, branching based deep models (Han et al., 2017) branch the network to induce more diversity. Motivated by different initializations of the neural networks leads to different local minima, Xue et al. (2021) proposed deep ensemble model wherein ensemble of fully convolution neural network over multiloss module with coarse fine compensation module resulted in better segmentation of central serous chorioretinopathy lesion. Multiple neural networks with different initializations, multiple loss functions resulted in better diversity in an ensemble (see Table 3).

4.6. Homogeneous & heterogeneous ensembles

Homogeneous ensemble (HOE) and heterogeneous ensemble (HEE) involve training a group of base learners either from the same family or different families, as shown in Fig. 5 and Fig. 6, respectively. Hence, each model of an ensemble must be as diverse as possible, and each base model must perform better than the random guess. The base learner can be a decision tree, neural network, or any other learning model.

In homogeneous ensembles, the same base learner is used multiple times to generate the family of base classifiers. However, the key issue is to train each base model such that the ensemble model is as diverse as possible, i.e. no two models are making the same error on a particular data sample. The two most common ways of inducing randomness in a homogeneous ensemble are either sampling of the training set multiple times, thereby training each model on a different bootstrapped sample of the training data or sampling the feature space of the training data and train each model on a different feature subset of the training data. In some ensemble models like Random forest (Breiman, 2001) used both these techniques for introducing diversity in the ensemble of decision trees. In neural networks, training models independently with different initialization of the models also induces diversity. However, deep learning models have high training costs and hence, training of multiple deep learning models is not a feasible option. Some attempts, like horizontal vertical voting of deep ensembles (Xie et al., 2013) have been made to obtain ensembles of deep models without independent training. Temporal ensemble (Laine and Aila, 2016) trains multiple models with different input augmentation, different regularization and different training epochs. Training of multiple deep neural networks

Table 3
Implicit/Explicit ensembles.

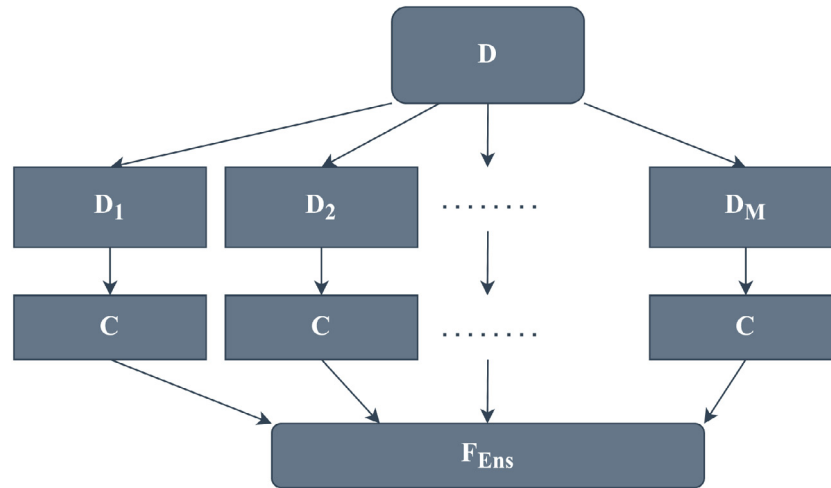| Year	Authors	Contribution
2013	Wan et al. (2013)	Introduced DropConnect (Random skipping of connections)
2014	Srivastava et al. (2014)	Introduced Dropout (Random skipping of units)
2016	Huang et al. (2016b)	Deep networks with Stochastic depth (Random skipping of blocks)
2016	Singh et al. (2016)	Introduced Swapout (Hybrid of Dropout and Stochastic depth approach)

Fig. 5. Homogeneous ensemble (HOE) has models based on the same algorithm, but each individual model are fed with distinct datasets.
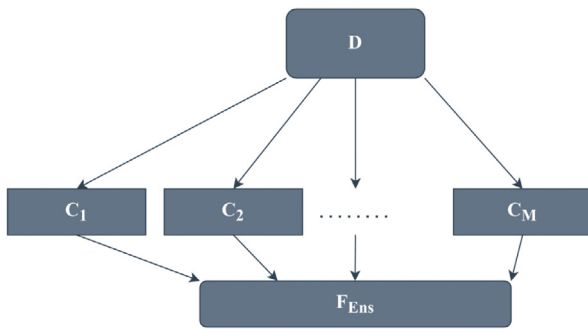

Fig. 6. The components in heterogeneous ensemble (HEE) share the same dataset but consists of various algorithms.

for image classification (Ciregan et al., 2012) and for disease prediction (Grassmann et al., 2018) showed that better performance is achieved via an ensemble of multiple networks and averaging the outputs. Despite these models, training multiple deep learning models for ensemble is an uphill task as millions or billions of parameters need to be optimized. Hence, some studies have used deep learning in combination with traditional models to build heterogeneous ensemble models, enjoying the benefits of lower computation and higher diversity. Heterogeneous ensemble for default prediction (Li et al., 2018) is an ensemble of the extreme gradient boosting, deep neural network and logistic regression. Heterogeneous ensemble for text classification (Kilimci and Akyokus, 2018) is an ensemble of multivariate Bernoulli naïve Bayes (MVNB), multinomial naïve Bayes (MNB), support vector machine (SVM), random forest (RF), and convolutional neural network (CNN) learning algorithms. Using different perspectives of data, model and decision fusion, heterogeneous deep network fusion (Tabik et al., 2020) showed that complex heterogeneous fusion architectures are more diverse and hence, show better generalization performance. Furthermore, Seijo-Pardo et al. (2017) employed both homogeneous and heterogeneous ensembles for feature selection. Zhao et al. (2010) suggested that the heterogeneous bagging based ensemble strategy performs better than boosting based Learn++ algorithms and some other

NCL methods. Other examples that employed homogeneous ensemble methods were used to deal with the presence of incremental tasks, such as concept drift (Minku et al., 2009), power load forecasting (Qiu et al., 2018; Grmanová et al., 2016), myoelectric prosthetic hands surface electromyogram characteristics (Duan and Dai, 2017), etc. Das et al. (2016) proposed an ensemble incremental learning with pseudo-outer-product fuzzy neural network for traffic flow prediction, real-life stock price, and volatility predictions, etc.

4.7. Decision fusion strategies

Ensemble learning trains several base learners and aggregates the outputs of base learners using some rules. The rule used to combine the outputs determines the effective performance of an ensemble. Most of the ensemble models focus on the ensemble architecture followed by their naive averaging to predict the ensemble output. However, naive averaging of the models, followed in most of the ensemble models, is not data adaptive and leads to less optimal performance (Ju et al., 2018) as it is sensitive to the performance of the biased learners. As there are billions of hyperparameters in deep learning architecture, the issue of overfitting may lead to the failure of some base learners. Hence, to overcome these issues, approaches like Bayes optimal classifier and super learner have been followed (Ju et al., 2018).

The different approaches followed in the literature for combining the outputs of the ensemble models are:

4.7.1. Unweighted model averaging

Unweighted averaging of the outputs of the base learners in an ensemble is the most followed approach for fusing the decisions in the literature. Here, the outcomes of the base learners are averaged to get the final prediction of the ensemble model. Deep learning architectures have high variance and low bias, thus, simple averaging of the ensemble models improve the generalization performance due to the reduction of the variance among the models.

The averaging of the base learners is performed either on the outputs of the base learners directly or on the predicted probabilities

of the classes via softmax function:

$$P_i^j = softmax^j(O_i) = \frac{O_i^j}{\sum_{k=1}^{K} exp(O_k^j)} \tag{4}$$

where P_i^j is the probability outcome of the ith unit on the jth base learner, O_i^j is the output of the ith unite of the jth base learner and K is the number of the classes.

Unweighted averaging is a reasonable choice when the performance of the base learners is comparable, as suggested in He et al. (2016), Simonyan and Zisserman (2014) and Szegedy et al. (2015). However, when the ensemble contains heterogeneous base learners naive unweighted averaging may result in suboptimal performance as it is affected by the performance of the weak learners and the overconfident learners (Ju et al., 2018). The adaptive metalearner should be good enough to adaptively combine the strengths of the base learners as some learners may have lower overall performance but maybe good at the classification of certain subclasses and hence, leading to better overall performance.

4.7.2. Majority voting

Similar to unweighted averaging, majority voting combines the outputs of the base learners. However, instead of taking the average of the probability outcomes, majority voting counts the votes of the base learners and predicts the final labels as the label with the majority of votes. In comparison to unweighted averaging, majority voting is less biased towards the outcome of a particular base learner as the effect is mitigated by majority vote count. However, favoring of a particular event by most of the similar base learners or dependent base learners leads to the dominance of the event in the ensemble model. In majority voting, the analysis by Kuncheva et al. (2003) showed that the pairwise dependence among the base learners plays an important role and for the classification of images, the prediction of shallow networks is more diverse as compared to the deeper networks (Choromanska et al., 2015). Hence, Ju et al. (2018) hypothesized that the performance of the majority voting based shallows ensemble models is better as compared to the majority based deep ensemble models.

Voting methods have also started to be integrated with semi-supervised deep learning. Li et al. (2017b) proposed an ensemble semi-supervised deep acoustic models for in automatic speech recognition. Wang et al. (2019c) explored an ensemble self-learning method to enhance semi-supervised performance and extract adverse drug events from social media in Liu et al. (2018). In the semi-supervised classification area, the author proposed a deep coupled ensemble learning method which is combined with complementary consistency regularization and gets the state of the art performance in Li et al. (2019a). Some results have also been achieved with semi-supervised ensemble learning on some datasets where the annotation is costly. Pio et al. (2014) employed an ensemble method to improve the reliability of miRNA:miRNA predicted interactions.

Furthermore, the multi-label classification (Tsoumakas and Katakis, 2007) problem is also a major point addressed by the voting method, a typical application is the RAndom k-labELsets (RAKEL) algorithm (Tsoumakas and Vlahavas, 2007). The author trained several single-label classifiers using small random subsets of actual labels. Then the final output is carried out by a voting scheme based on the predictions of these single classifiers. There are also many variants of RAKEL proposed in recent years (Moyano et al., 2019; Kimura et al., 2016; Wang et al., 2021). Shi et al. (2011) proposed a solution for multi-label ensemble learning problem, which construct several accurate and diverse multi-label based basic classifiers and employ two objective functions to evaluate the accuracy and diversity of multi-label base learners. Another work (Li et al., 2013) proposed an ensemble multi-label classification framework based on variable pairwise constraint projection. Xia et al. (2021) proposed a weighted stacked ensemble scheme that employs the sparsity regularization to facilitate classifier

selection and ensemble construction. Besides, there are many applications of ensemble multi-label methods. Some publications employ multi-label ensemble classifiers to explore the protein, such as protein subcellular localization (Guo et al., 2016), protein function prediction (Yu et al., 2012), etc. The Muli-label classifier is also utilized in predicting the drug side effects (Zhang et al., 2015), predicting the gene prediction (Schietgat et al., 2010), etc. Moreover, there is another critical ensemble multi-label algorithm called ensemble classifier chains (ECC) (Read et al., 2011). This method involves binary classifiers linked along a chain. The first classifier is trained using only the input data, and then each subsequent classifier is trained on the input space and all previous classifiers in the chain. The final prediction is obtained by the integration of the predictions and selection above a manually set threshold. Chen et al. (2017) propose an ensemble application of convolutional and recurrent neural networks to capture both the global and local textual semantics and to model high-order label correlations.

4.7.3. Bayes optimal classifier

In Bayesian method, hypothesis h_j of each base learner with the conditional distribution of target label t given x. Let h_j be the hypothesis generated on the training data D evaluated on test data (x, t), mathematically, $h_j(t|x) = P[y|x, h_j, D]$. With Bayes rule, we have

$$P(t|x, D) \propto \sum_{h_j} P[t|h_j, x, D]P[D|h_j]P[h_j] \tag{5}$$

and the Bayesian Optimal classifier is given as:

$$\underset{t}{argmax} \sum_{h_j} P[t|h_j, x, D]P[D|h_j]P[h_j], \tag{6}$$

where $P[D|h_j] = \Pi_{(t,x) \in D} h_j(t|x)$ is the likelihood of the data under h_j. However, due to overfitting issues this might be not a good measure. Hence, training data is divided into two sets-one for training the model and the other for evaluating the model. Usually validation set is used to tune the hyperparameters of the model.

Choosing prior probabilities in Bayes optimal classifier is difficult and hence, usually set to uniform distribution for simplicity. With a large sample size, one hypothesis tends to give larger posterior probabilities than others and hence the weight vector is dominated by a single base learner and hence Bayes optimal classifier would behave as the discrete super learner with a negative likelihood loss function.

4.7.4. Stacked generalization

Stacked generalization (Wolpert, 1992) works by deducing the biases of the generalizer(s) with respect to a provided learning set. To obtain the good linear combination of the base learners in regression, cross-validation data and least squares under non-negativity constraints was used to get the optimal weights of combination (Breiman, 1996c). Consider the linear combination of the predictions of the base learners $f_1, f_2, \ldots, f_m$ given as:

$$f_{stacking}(x) = \sum_{j=1}^{m} w_j f_j(x) \tag{7}$$

where w is the optimal weight vector learned by the meta learner.

4.7.5. Super learner

Inspired by the cross validation for choosing the optimal classifier, Van der Laan et al. (2007) proposed super learner which is weighted combination of the predictions of the base learner. Unlike the stacking approach, it uses cross validation approach to select the optimal weights for combining the predictions of the base learners.

With smaller datasets, cross validation approach can be used to optimize the weights. However, with the increase in the size of the data and the number of base learners in the model, it may not be a feasible option. Instead of optimizing the V-fold cross validation, single split cross validation can also be used for optimizing the weights for optimal combination (Ju et al., 2019). In deep learning models, usually,
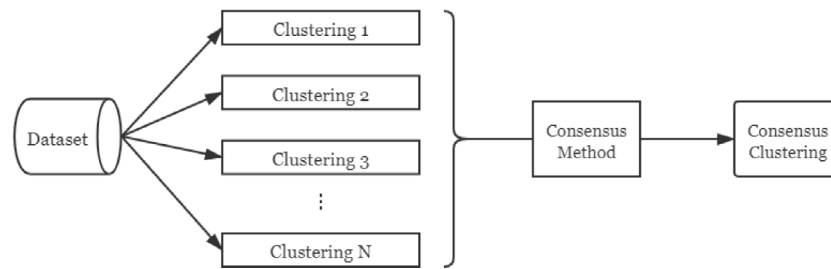
Fig. 7. The process of consensus clustering. An ensemble of different clustering results can be combined by a consensus approach.

a validation set is used to evaluate the performance instead of using the cross validation.

Another application field for super learner is in Reinforcement Learning With the development of Deep learning, some researchers have implemented deep reinforcement learning, which combines deep learning with a Q-learning algorithm (Mnih et al., 2013). Ensemble methods in deep Q learning have decent performance. Chen et al. (2018a) proposed an ensemble network architecture for deep reinforcement learning. The integrated network includes Temporal Ensemble and Target Values Ensemble. Develop a human-like chat robot is a challenging job, by incorporating deep reinforcement learning and ensemble method, Cuayáhuitl et al. (2019) integrated 100 deep reinforcement learning agents, the agents are trained based on clustered dialogues. They also demonstrate the ensemble of DRL agents has better performance than the single variant or Seq2Seq model. Stock trading is another topic where ensemble deep reinforcement learning has achieved a promising result. Carta et al. (2020) found the single supervised classifier is inadequate to deal with the complex and volatile stock market. They employed hundreds of neural networks to pre-process the data, then they combined several reward-based meta learners as a trading agency. Moreover, Yang et al. (2020) trained an ensemble trading agency based on three different metrics: Proximal Policy Optimization (PPO), Advantage Actor–Critic (A2C), and Deep Deterministic Policy Gradient (DDPG). The ensemble strategy combines the advantages of the three different algorithms. Besides, some researchers try to use ensemble strategy to solve the disease-prediction problem. The proposed model in Tang et al. (2016) consists of several sub-models which are in response to different anatomical parts.

4.7.6. Consensus

Unsupervised learning is another group of machine learning techniques. The fundamental difference between it and supervised learning is that unsupervised learning usually handles training samples without corresponding labels. Therefore, the primary usage of unsupervised learning is to do clustering. The reason why ensemble methods are employed is to combine some weak clusters into strong one. To create diverse clusters, several approaches can be applied: using different sampling data, using different subsets of the original features, and employing different clustering methods (Şenbabaoğlu et al., 2014). Sometimes, even some random noise can be added to these base models to increase randomness, which is good for ensemble methods according to Bian and Wang (2007). After receiving all the outputs from each cluster, various consensus functions can be chosen to obtain the final output based on the user's requirement (Vega-Pons and Ruiz-Shulcloper, 2011). The ensemble clustering is also known as consensus clustering Fig. 7.

Zhou and Tang (2006) explored ensemble methods for unsupervised learning and developed four different approaches to combine the outputs of these clusters. In recent years, some new ensemble clustering methods have been proposed that illustrated the priority of ensemble learning (Huang et al., 2017b; Zheng et al., 2010; Huang et al., 2016a). Most of the clustering ensemble method is based on the co-association matrix solution, which can be regarded as a graph partition problem. Besides, there is some research focus on integrating the deep structure

and ensemble clustering method. Liu et al. (2015, 2016) firstly showed that ensemble unsupervised representation learning with deep structure can be applied in large scale data. Then the author combined the method with auto-encoder and extends it to the vision field. Shaham et al. (2016) first demonstrated that some crowdsourcing algorithms can be replaced by a Restricted Boltzmann Machine with a single hidden neuron, then propose an RBM-based Deep Neural Net (DNN) used for unsupervised ensemble learning. The unsupervised ensemble method also makes some contribution to the field of Natural Language Processing. Alami et al. (2019) demonstrated that the ensemble of unsupervised deep neural network models that use Sentence2Vec representation as the input has the best performance according to the experiments. Hassan et al. (2019) proposed a module that includes four semantic similarity measures, which improves the performance on the semantic textual similarity (STS) task. The unsupervised ensemble method is also widely used for tasks that lack annotation, such as the medical image. Ahn et al. (2019) proposed an unsupervised feature learning method integrated ensemble approach with a traditional convolutional neural network. Lahiri et al. (2016) employed unsupervised hierarchical feature learning with ensemble sparsely autoencoder on retinal blood vessels segmentation task, meanwhile, Liu et al. (2019) also propose an unsupervised ensemble architecture to automatically segment retinal vessel. Besides, there are also some ensemble deep methods working on localization predicting for long non-coding RNAs Cao et al. (2018). Hu and Suganthan (2022) extended the ensemble random vector functional link to unsupervised tasks. The authors employ manifold regularization to re-represent the original features, and then use the Kuhn–Munkre algorithm with consensus clustering to ensemble the clustering results from multiple hidden layers.

4.7.7. Query-by-committee

Active Learning is another popular topic in the deep learning area, which is also often used in conjunction with semi-supervised learning and ensemble learning. The key sight of this is to make the algorithm learning from less annotated data. Some conventional active learning algorithms, such as Query-By-Committee (as shown in Fig. 8), have already adopted the idea of ensemble learning. Melville and Mooney (2003, 2004) explored an ensemble method that builds a diverse committee. Beluch et al. (2018) discussed the power of ensembles for active learning is significantly better than Monte-Carlo Dropout and geometric approaches. Sharma and Rani (2018) show some applications in drug-target interaction prediction. Ensemble active learning is also available to conquer the concept drift and class imbalance problem (Zhang et al., 2018).

5. Applications

In this section, we briefly present the applications of deep ensemble models across different domains in a tabular form. Ensemble deep models have been implemented in several domains and therefore, in broad sense, we have classified the application domains into five categories, i.e., health care, speech, image classification, forecasting

Table 4
Applications in health care.

| Year	Author	Title	Approach	Area
2014	Zheng et al. (2014)	HIBAG—HLA genotype imputation with attribute bagging	Bagging	Genotype Imputation
2014	Cortes et al. (2014)	Deep Boosting	Boosting	Classification
2015	Zhang et al. (2015)	Predicting drug side effects by multi-label learning and ensemble learning	Decision Fusion	Predict the drug side effects
2016	Guo et al. (2016)	Human protein subcellular localization with integrated source and multi-label ensemble classifier.	Decision Fusion	Protein subcellular localization prediction
2016	Lahiri et al. (2016)	Deep neural ensemble for retinal vessel segmentation in fundus images towards achieving label-free angiography	Decision Fusion	Medical image segmentation
2017	Cabria and Gondra (2017)	MRI segmentation fusion for brain tumor detection	Heterogeneous ensemble	MRI segmentation
2018	Grassmann et al. (2018)	A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography	Homogeneous ensemble	Disease prediction
2018	Cao et al. (2018)	The lnclocator: a subcellular localization predictor for long non-coding RNAs based on a stacked ensemble classifier	Decision Fusion	Subcellular localization predictor
2018	Sharma and Rani (2018)	Be-dti': Ensemble framework for drug target interaction prediction using dimensionality reduction and active learning	Active learning	Drug target interaction prediction
2019	Ahn et al. (2019)	Unsupervised feature learning with k-means and an ensemble of deep convolutional neural networks for medical image classification	Decision Fusion	Medical image classification
2019	Liu et al. (2019)	Unsupervised ensemble strategy for retinal vessel segmentation.	Unsupervised	Medical image classification
2020	Shalbaf and Vafaeezadeh (2021)	Automated detection of COVID-19 using ensemble of transfer learning with deep convolutional neural network based on CT scans	Heterogeneous Ensemble	Detection of COVID-19
2020	Ali et al. (2020)	A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion	Boosting	Heart disease prediction
2021	Zhou et al. (2021)	The ensemble deep learning model for novel COVID-19 on CT images	Heterogeneous Ensemble	Detection of COVID-19
2021	Li et al. (2020)	Intelligent Fault Diagnosis by Fusing Domain Adversarial Training and Maximum Mean Discrepancy via Ensemble Learning	Heterogeneous Ensemble	Fault diagnosis
2021	Das et al. (2021)	Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network	Heterogeneous Ensemble	Detection of COVID-19
2022	Sukegawa et al. (2022)	Identification of osteoporosis using ensemble deep learning model with panoramic radiographs and clinical covariates	Decision Fusion	Identification of osteoporosis
2022	Gao et al. (2022)	Vessel segmentation for X-ray coronary angiography using ensemble methods with deep learning and filter-based features	Boosting	Vessel segmentation
2022	Rath et al. (2022)	Improved heart disease detection from ECG signal using deep learning based ensemble model	Heterogeneous Ensemble	Heart disease detection
2022	Tanveer et al. (2022)	Classification of Alzheimer's Disease Using Ensemble of Deep Neural Networks Trained Through Transfer Learning	Heterogeneous Ensemble	Classification of Alzheimer's Disease
2022	Rai and Chatterjee (2022)	Hybrid CNN-LSTM deep learning model and ensemble technique for automatic detection of myocardial infarction using big ECG data	Heterogeneous Ensemble	Detection of myocardial infarction
2022	Ganaie and Tanveer (2022)	Ensemble deep random vector functional link network using privileged information for Alzheimer's disease diagnosis	Implicit ensemble	Diagnosis of Alzheimer's disease
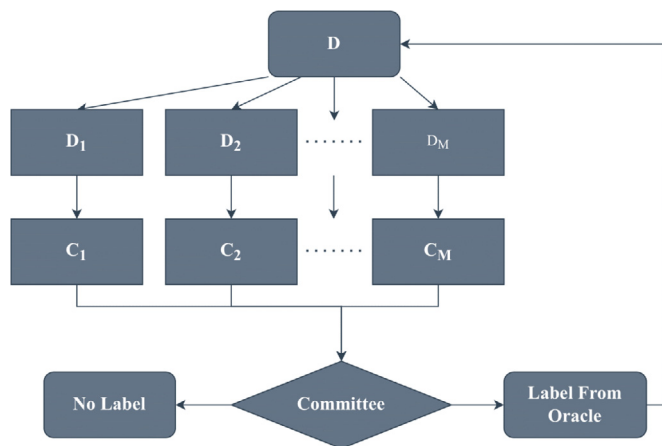
Fig. 8. Query-by-committee in Active Learning. Sampling with replacement is used to partition the labeled training data set into training splits. The committee determines whether to label the data based on the output of several algorithms.

and the rest models are listed in others category. Table 4 gives the information about the ensemble deep models that have been implemented in health care domain. Here, several papers are based on heterogeneous ensemble technique. It reveals that using different family's models into a single frame perform better in health care domain. Recently, ensemble deep techniques have been successful and have shown good

performance in health care domain. The models which have been implemented for speech task have been given in Table 5 and most of the ensemble approaches are based on stacking technique. Table 6 contains the ensemble deep models that have been implemented in speech areas. Models that have been implemented in forecasting and other domains have been given in Table 7 and Table 8, respectively. Fig. 9 shows the percentages of the application domains. The statistics reveals that different ensemble deep techniques have been used in different areas. A larger number of models, i.e. 27% of the ensemble deep models, have been implemented in health care domain and 5.6% percent of the models for speech application and 22.5% of the models for image classification task. Moreover, 9% models have been used in forecasting and 36% in other applications areas, i.e., information retrieval, emotion recognition, text categorization and so on. Fig. 10 shows the ensemble strategies in percentage. In ensemble learning, there are several ways to integrate the outcomes of the models in an ensemble. In the literature, researchers have proposed different techniques of decision fusion according to different areas of application. Bagging, boosting and stacking are the classical ensemble techniques. Based on these three techniques, researchers have developed several other techniques also. Boosting (18.2%), stacking (12.5%) and bagging (4.5%) techniques have been implemented in ensemble deep framework. Heterogeneous and implicit ensemble are also popular for making an efficient ensemble model and their contribution are as follows 11.4% and 10.2%, respectively. The rest ensemble techniques are, i.e. unsupervised (3.4%), NCL (3.4%), reinforcement (1.1%), active learning (1.1%), explicit ensemble (1.1%) and homogeneous ensemble (3.4%).

Table 5
Applications in speech.

Year	Author	Title	Approach	Area
2012	Tur et al. (2012)	Towards deeper understanding: Deep convex networks for semantic utterance classification	Stacking	Semantic Utterance Classification
2012	Deng et al. (2012b)	Use of kernel deep convex networks and end-to-end learning for spoken language understanding	Stacking	Spoken Language Understanding
2014	Deng and Platt (2014)	Ensemble deep learning for speech recognition	Stacking	Speech Recognition
2014	Palangi et al. (2014)	Recurrent Deep-Stacking Networks for sequence classification	Stacking	Sequence classification
2017	Li et al. (2017b)	Semi-supervised ensemble DNN acoustic model training	Decision Fusion	Speech Recognition

Table 6
Applications in image classification.

Year	Author	Title	Approach	Area
2012	Ciregan et al. (2012)	Multi-column deep neural networks for image classification	Homogeneous ensemble	Classification
2013	Wan et al. (2013)	Regularization of Neural Networks using DropConnect	Implicit ensemble	Image recognition
2014	Srivastava et al. (2014)	Dropout: a simple way to prevent neural networks from overfitting	Implicit Ensemble	Computer vision, speech recognition document classification and computational biology
2014	Liu et al. (2014)	Facial expression recognition via a boosted deep belief network	Boosting	Facial expression recognition
2015	Li et al. (2015)	Sparse deep stacking network for image classification	Stacking	Image Classification
2015	Yang et al. (2015)	Convolutional channel features	Boosting	Pedestrian detection, face detection, edge detection and object proposal generation
2016	Moghimi et al. (2016)	Boosted Convolutional Neural Networks	Boosting	Classification
2016	Huang et al. (2016b)	Deep networks with stochastic depth	Implicit ensemble	Classification
2016	He et al. (2016)	Deep residual learning for image recognition	Implicit ensemble	classification, and object detection
2016	Singh et al. (2016)	Swapout: Learning an ensemble of deep architectures	Implicit ensemble	classification
2016	Smith et al. (2016)	Gradual dropin of layers to train very deep neural networks	Implicit ensemble	Classification
2016	Laine and Aila (2016)	Temporal ensembling for semi-supervised learning	Homogeneous ensemble	Classification
2016	Tang et al. (2016)	Inquire and diagnose: Neural symptom checking ensemble using deep reinforcement learning	Decision Fusion	Inquire symptoms and diagnose diseases
2017	Huang et al. (2017a)	Snapshot ensembles: train 1, get M for free	Explicit ensemble	Classification
2017	Mosca and Magoulas (2017)	Deep incremental boosting	Boosting	Classification
2018	Beluch et al. (2018)	The power of ensembles for active learning in image classification	Decision Fusion	Image classification
2019	Amin-Naji et al. (2019)	Ensemble of CNN for multi-focus image fusion	Decision Fusion	Image Classification
2019	Li et al. (2019a)	Semi-supervised deep coupled ensemble learning with classification landmark exploration.	Decision Fusion	Image classification
2020	Wang et al. (2020)	Particle swarm optimization for evolving deep neural networks for image classification by evolving and stacking transferable blocks.	Stacking	Image Classification

Table 7
Applications in forecasting.

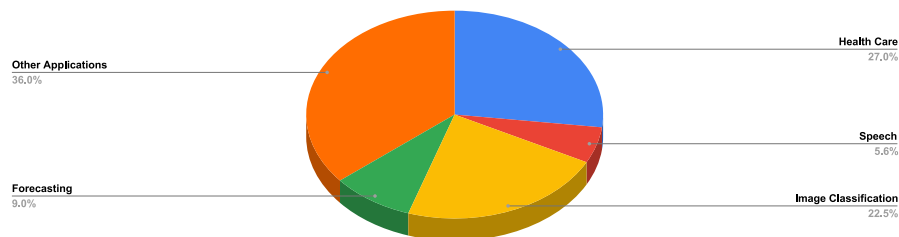| Year	Author	Title	Approach	Area
2014	Qiu et al. (2014)	Ensemble deep learning for regression and time series forecasting	Decision Fusion	Regression and Time Series Forecasting
2016	Grmanová et al. (2016)	Incremental ensemble learning for electricity load forecasting	Decision Fusion	Electricity load forecasting
2017	Qiu et al. (2017)	Empirical Mode Decomposition based ensemble deep learning for load demand time series forecasting	Decision Fusion	Load demand forecasting
2017	Liu et al. (2017)	A Flood Forecasting Model Based on Deep Learning Algorithm via Integrating Stacked Autoencoders with BP Neural Network	Stacking	Flood Forecasting
2018	Qiu et al. (2018)	Ensemble incremental learning random vector functional link network for short-term electric load forecasting.	Decision Fusion	Electric load forecasting
2020	Carta et al. (2020)	A multi-layer and multi-ensemble stock trader using deep learning and deep reinforcement learning	Implicit ensemble	Stock trader
2020	Yang et al. (2020)	Deep reinforcement learning for automated stock trading: An ensemble strategy	Decision Fusion	Stock trading agency
2021	Bhusal et al. (2021)	Deep ensemble learning-based approach to real-time power system state estimation	Stacking	Electric Power
2022	Singla et al. (2021)	An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network	Decision Fusion	Forecasting

Fig. 9. Ensemble-based approach in different areas. Data from Tables 4 to 8.

Table 8
Other applications.

Year	Author	Title	Approach	Area
2013	Deng et al. (2013)	Deep stacking networks for information retrieval	Stacking	Information Retrieval
2014	Kuznetsov et al. (2014)	Multi-class deep boosting	Boosting	Classification
2014	Wang et al. (2014)	Sentiment classification The contribution of ensemble learning	Bagging, Boosting	Sentiment classification
2015	Zareapoor and Shamsolmoali (2015)	Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier	Bagging	Credit Card Fraud Detection
2016	Yin et al. (2017)	Recognition of emotions using multimodal physiological signals and an ensemble deep learning model	Decision Fusion	Emotions Recognition
2016	Liu et al. (2016)	A deep learning approach to unsupervised ensemble learning	Decision Fusion	Clustering
2016	Walach and Wolf (2016)	Learning to count with CNN boosting	Boosting	Object counting in images
2017	Han et al. (2016)	Incremental boosting convolutional neural network for facial action unit recognition	Boosting	Facial action unit recognition
2017	Chen et al. (2017)	Ensemble application of convolutional and recurrent neural networks for multi-label text categorization	Decision Fusion	Text Categorization.
2017	Opitz et al. (2017)	Bier-boosting independent embeddings robustly	Boosting	Image retrieval
2018	Shi et al. (2018)	Crowd Counting with Deep Negative Correlation Learning	Negative correlation learning	Crowd Counting
2018	Kazemi et al. (2018)	Novel genetic-based negative correlation learning for estimating soil temperature	Negative correlation learning	Soil Temperature Estimation
2018	Randhawa et al. (2018)	Credit Card Fraud Detection Using AdaBoost and Majority Voting	Boosting	Credit Card Fraud Detection
2018	Sun et al. (2018)	Sparse Deep Stacking Network for Fault Diagnosis of Motor	Stacking	Fault Diagnosis
2018	Chen et al. (2018b)	Deep boosting for image denoising	Boosting	Image denoising
2018	Li et al. (2018)	Heterogeneous ensemble for default prediction of peer-to-peer lending in China	Heterogeneous ensemble	Default prediction
2018	Kilimci and Akyokus (2018)	Deep learning-and word embedding-based heterogeneous classifier ensembles for text classification	Heterogeneous ensemble	Classification
2018	Liu et al. (2018)	SSEL-ADE: a semi-supervised ensemble learning framework for extracting adverse drug events from social media	Decision Fusion	Extracting adverse drug events
2019	Martín et al. (2019)	Android malware detection through hybrid features fusion and ensemble classifiers: The AndroPyTool framework and the OmniDroid dataset	Decision Fusion	Android malware detection
2019	Cuayáhuitl et al. (2019)	Ensemble-based deep reinforcement learning for chatbots	Reinforcement	Chat robot
2019	Chen et al. (2019b)	Real-world image denoising with deep boosting	Boosting	Image denoising
2019	Waltner et al. (2019)	HiBsteR: Hierarchical Boosted Deep Metric Learning for Image Retrieval	Boosting	Image Retrieval
2019	Wang et al. (2019b)	AdaBoost-based security level classification of mobile intelligent terminals	AdaBoost	Security Level Classification
2019	Chen et al. (2019a)	Novel Hybrid Integration Approach of Bagging-Based Fisher's Linear Discriminant Function for Groundwater Potential Analysis	Bagging	Groundwater Potential Analysis
2019	Shi et al. (2021)	Random vector functional link neural network based ensemble deep learning	Implicit ensemble	Classification
2019	Zhang et al. (2019a)	Deep stacked hierarchical multi-patch network for image deblurring	Stacking	Deblurring Image
2019	Alami et al. (2019)	Enhancing unsupervised neural networks based text summarization with word embedding and ensemble learning	Decision Fusion	Text summarization
2019	Hassan et al. (2019)	Uests: An unsupervised ensemble semantic textual similarity method	Decision Fusion	Semantic textual similarity
2020	Zhang et al. (2020b)	Grasp for stacking via deep reinforcement learning	Stacking	Robotic arm control
2020	Zhang et al. (2020a)	Snapshot boosting: a fast ensemble framework for deep neural networks	Boosting	Computer vision (CV) and the natural language processing (NLP) tasks
2021	Tsogbaatar et al. (2021)	DeL-IoT: A deep ensemble learning approach to uncover anomalies in IoT	Decision Fusion	IoT
2022	Wen et al. (2020)	A new ensemble convolutional neural network with diversity regularization for fault diagnosis	Snapshot ensemble learning	Fault diagnosis
2022	Hu and Suganthan (2022)	Representation Learning Using Deep Random Vector Functional Link Networks for Clustering	Decision Fusion	Clustering
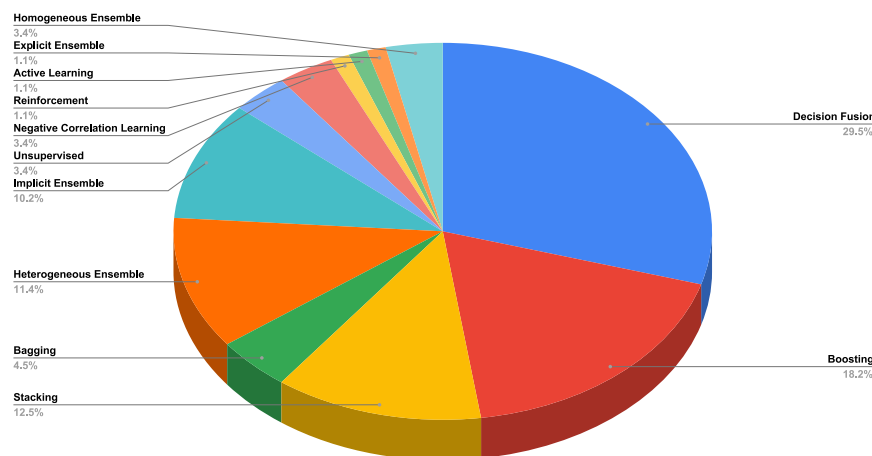
Fig. 10. Analysis of the applications of various ensemble methods. Data from Tables 4 to 8.

6. Conclusions and future works

In this paper, we reviewed the recent developments of ensemble deep learning models. The theoretical background of ensemble learning has been elaborated to understand the success of ensemble learning. The various approaches ranging from traditional ones like bagging, boosting to the recent novel approaches like implicit/explicit ensembles, heterogeneous ensembles, have led to better performance of deep ensemble models. We also reviewed the applications of the deep ensemble models in different domains.

Although deep ensemble models have been applied across different domains, there are several open problems which can be explored in the future to fill the gap. Big data (Zhou et al., 2014) is still a challenging problem, one can explore the benefits of deep ensemble models for learning the patterns using the techniques like implicit deep ensemble to maximize the performance in both time and generalization aspects.

Deep learning models are difficult to train than shallow models as large number of weights corresponding to different layers need to be tuned. Creating deep ensemble models may further complicate the problem. Hence, randomized models can be explored to overcome the training cost. Bagging based deep ensemble may incur heavy training time for optimizing the ensemble models. Hence, one can investigate the alternate ways of inducing diversity in the base models with lesser training cost. Randomized learning modules like random vector functional link network (Pao et al., 1994) are best suited for creating the ensemble models as randomized models lead to a significant variance reduction. Also, the hidden layers are randomly initialized, hence, can be used to create deep ensembles without incurring any additional cost of training (Shi et al., 2021). Randomized modules can be further explored using different techniques like implicit/explicit ensembles (Shi et al., 2021), stacking based ensembles (Katuwal and Suganthan, 2019). However, there are still open directions which can be worked upon like negative correlation learning, heterogeneous ensembles and so on.

Implicit/explicit ensembles are faster compared to training of multiple deep models. However, creating diversity within a single model is a big challenge. One can explore the methods to induce more diversity among the learners within these ensembles like branching based deep models (Han et al., 2017). Investigate the extension of explicit/implicit ensembles to traditional models.

Following the stacking based approach, Deep convex net (DCN) (Deng and Yu, 2011), traditional methods like random forest (Breiman, 2001; Zhou and Feng, 2017), support vector machines (Wang et al., 2019d,a; Li et al., 2019b) have been extended to deep learning architectures which resulted in improved performance. One can investigate these traditional models for creating the deep ensemble models.

Another big challenge of ensemble deep learning lies in model selection for building the ensemble architecture, homogeneous and heterogeneous ensembles represent two different ways for choosing the model. However, to answer how many different algorithms, and how many base learners in the ensemble architecture, are still problem-dependent. Finding a criterion for model selection in ensemble deep learning should be an important target for researchers in the next few years. Since most of the models focus on developing the architectures with little attention towards how to combine the base learners prediction is still unanswered. Hence, one can investigate the effect of different fusion strategies on the prediction of an ensemble output.

For unsupervised ensemble learning or consensus clustering, the ensemble approaches include but are not limited to: Hyper-graph partitioning, Voting approach, Mutual information, etc. Consensus clustering is a powerful tool and it can improve performance in most cases. However, there are many concerns remain to be tackled, it is exquisitely sensitive, which might assert as an apparent structure without obvious demarcation or declared cluster stable without cluster resistance. Besides, current method cannot handle some complex but possible scenarios, such as the boundary samples are assigned to the single cluster, clusters do not intersect and the methods are not able to represent outliers. These are the possible research directions for future work.

The problem of semi-supervised ensemble domains has not been extensively studied yet, and most of the literature shows that semi-supervised ensemble methods are mainly used in cases where there is insufficient labeling data. Also, combining the semi-supervision with some other machine learning methods, such as active learning, is a direction for future research.

Reinforcement learning is another popular topic recently. The idea of integrating model-based reinforcement learning with ensemble learning has been used with promising results in many applications, but there is little integration of planning & learning-based reinforcement learning with ensemble learning methods.

CRediT authorship contribution statement

M.A. Ganaie: Idea and Conceptualization, Methodology, Writing – original draft. **Minghui Hu:** Idea and Conceptualization, Methodology, Writing – original draft . **A.K. Malik:** Idea and Conceptualization, Methodology, Writing – original draft. **M. Tanveer:** Idea and Conceptualization, Validation, Reviewing, Supervision. **P.N. Suganthan:** Idea and Conceptualization, Writing – Reviewing, Supervision.

Declaration of competing interest

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

Acknowledgments

References

Ahn, E., Kumar, A., Feng, D., Fulham, M., Kim, J., 2019. Unsupervised feature learning with K-means and an ensemble of deep convolutional neural networks for medical image classification. arXiv preprint arXiv:1906.03359.

Alami, N., Meknassi, M., En-nahnahi, N., 2019. Enhancing unsupervised neural networks based text summarization with word embedding and ensemble learning. Expert Syst. Appl. 123, 195–211.

Ali, F., El-Sappagh, S., Islam, S.R., Kwak, D., Ali, A., Imran, M., Kwak, K.-S., 2020. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. Inf. Fusion 63, 208–222.

Alvear-Sandoval, R.F., Figueiras-Vidal, A.R., 2018. On building ensembles of stacked denoising auto-encoding classifiers and their further improvement. Inf. Fusion 39, 41–52.

Amin-Naji, M., Aghagolzadeh, A., Ezoji, M., 2019. Ensemble of CNN for multi-focus image fusion. Inf. Fusion 51, 201–214.

Barandiaran, I.n., 1998. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20.

Beluch, W.H., Genewein, T., Nürnberger, A., Köhler, J.M., 2018. The power of ensembles for active learning in image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9368–9377.

Beygelzimer, A., Hazan, E., Kale, S., Luo, H., 2015. Online gradient boosting. Adv. Neural Inf. Process. Syst. 28.

Bhusal, N., Shukla, R.M., Gautam, M., Benidris, M., Sengupta, S., 2021. Deep ensemble learning-based approach to real-time power system state estimation. Int. J. Electr. Power Energy Syst. 129, 106806.

Bian, S., Wang, W., 2007. On diversity and accuracy of homogeneous and heterogeneous ensembles. Int. J. Hybrid Intell. Syst. 4, 103–128.

Błaszczyński, J., Stefanowski, J., 2015. Neighbourhood sampling in bagging for imbalanced data. Neurocomputing 150, 529–542.

Breiman, L., 1996a. Bias, Variance, and Arcing Classifiers. Technical Report 460, Statistics Department, University of California, Berkeley.

Breiman, L., 1996b. Bagging predictors. Mach. Learn. 24, 123–140.

Breiman, L., 1996c. Stacked regressions. Mach. Learn. 24, 49–64.

Breiman, L., 1998. Arcing classifier (with discussion and a rejoinder by the author). Ann. Statist. 26, 801–849.

Breiman, L., 2000. Randomizing outputs to increase prediction accuracy. Mach. Learn. 40, 229–242.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brown, G., Wyatt, J., Harris, R., Yao, X., 2005a. Diversity creation methods: a survey and categorisation. Inf. Fusion 6, 5–20.

Brown, G., Wyatt, J.L., Tiňo, P., 2005b. Managing diversity in regression ensembles. J. Mach. Learn. Res. 6, 1621–1650.

Bühlmann, P., Yu, B., 2002. Analyzing bagging. Ann. Statist. 30, 927–961.

Buja, A., Stuetzle, W., 2000. Smoothing effects of bagging. In: Preprint. AT&T Labs-Research, Citeseer.

Buschjäger, S., Pfahler, L., Morik, K., 2020. Generalized negative correlation learning for deep ensembling. arXiv preprint arXiv:2011.02952.

Cabria, I., Gondra, I., 2017. MRI segmentation fusion for brain tumor detection. Inf. Fusion 36, 1–9.

Cao, Y., Geddes, T.A., Yang, J.Y.H., Yang, P., 2020. Ensemble deep learning in bioinformatics. Nat. Mach. Intell. 2, 500–508.

Cao, Z., Pan, X., Yang, Y., Huang, Y., Shen, H.-B., 2018. The lnclocator: a subcellular localization predictor for long non-coding RNAs based on a stacked ensemble classifier. Bioinformatics 34, 2185–2194.

Carta, S., Corriga, A., Ferreira, A., Podda, A.S., Recupero, D.R., 2020. A multi-layer and multi-ensemble stock trader using deep learning and deep reinforcement learning. Appl. Intell. 1–17.

Chen, X.-l., Cao, L., Li, C.-x., Xu, Z.-x., Lai, J., 2018a. Ensemble network architecture for deep reinforcement learning. Math. Probl. Eng. 2018.

Chen, W., Pradhan, B., Li, S., Shahabi, H., Rizeei, H.M., Hou, E., Wang, S., 2019a. Novel hybrid integration approach of bagging-based fisher's linear discriminant function for groundwater potential analysis. Nat. Resources Res. 28, 1239–1258.

Chen, C., Xiong, Z., Tian, X., Wu, F., 2018b. Deep boosting for image denoising. In: Proceedings of the European Conference on Computer Vision, ECCV. pp. 3–18.

Chen, C., Xiong, Z., Tian, X., Zha, Z.-J., Wu, F., 2019b. Real-world image denoising with deep boosting. IEEE Trans. Pattern Anal. Mach. Intell..

Chen, G., Ye, D., Xing, Z., Chen, J., Cambria, E., 2017. Ensemble application of convolutional and recurrent neural networks for multi-label text categorization. In: 2017 International Joint Conference on Neural Networks, IJCNN. IEEE pp. 2377–2383.

Choromanska, A., Henaff, M., Mathieu, M., Arous, G.B., LeCun, Y., 2015. The loss surfaces of multilayer networks. In: Artificial Intelligence and Statistics pp. 192–204.

Ciregan, D., Meier, U., Schmidhuber, J., 2012. Multi-column deep neural networks for image classification. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 3642–3649.

Condorcet, M.d., 1785. Essay on the Application of Analysis to the Probability of Majority Decisions. Imprimerie Royale, Paris.

Cortes, C., Gonzalvo, X., Kuznetsov, V., Mohri, M., Yang, S., 2017. Adanet: Adaptive structural learning of artificial neural networks. In: International Conference on Machine Learning. PMLR, pp. 874–883.

Cortes, C., Mohri, M., Syed, U., 2014. Deep boosting. In: 31st International Conference on Machine Learning, ICML 2014.

Şenbabaoğlu, Y., Michailidis, G., Li, J.Z., 2014. Critical limitations of consensus clustering in class discovery. Sci. Rep. 4, 1–13.

Cuayáhuitl, H., Lee, D., Ryu, S., Cho, Y., Choi, S., Indurthi, S., Yu, S., Choi, H., Hwang, I., Kim, J., 2019. Ensemble-based deep reinforcement learning for chatbots. Neurocomputing 366, 118–130.

Das, R.T., Ang, K.K., Quek, C., 2016. ieRSPOP: A novel incremental rough set-based pseudo outer-product with ensemble learning. Appl. Soft Comput. 46, 170–186. http://dx.doi.org/10.1016/j.asoc.2016.04.015.

Das, A.K., Ghosh, S., Thunder, S., Dutta, R., Agarwal, S., Chakrabarti, A., 2021. Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network. Pattern Anal. Appl. 24, 1111–1124.

Deng, L., He, X., Gao, J., 2013. Deep stacking networks for information retrieval. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, pp. 3153–3157.

Deng, L., Platt, J.C., 2014. Ensemble deep learning for speech recognition. In: Fifteenth Annual Conference of the International Speech Communication Association.

Deng, L., Tur, G., He, X., Hakkani-Tur, D., 2012a. Use of kernel deep convex networks and end-to-end learning for spoken language understanding. In: 2012 IEEE Workshop on Spoken Language Technology, SLT 2012 - Proceedings, IEEE, pp. 210–215, http://dx.doi.org/10.1109/SLT.2012.6424224.

Deng, L., Tur, G., He, X., Hakkani-Tur, D., 2012b. Use of kernel deep convex networks and end-to-end learning for spoken language understanding. In: 2012 IEEE Spoken Language Technology Workshop, SLT, IEEE, pp. 210–215.

Deng, L., Yu, D., 2011. Deep convex net: A scalable architecture for speech pattern classification. Proc. Ann. Conf. Int. Speech Commun. Assoc. 2285–2288.

Deng, L., Yu, D., 2014. Deep learning: methods and applications. Foundations and Trends® in Signal Processing 7, 197–387.

Deng, L.L., Yu, D., Platt, J., 2012c. Scalable stacking and learning for building deep architectures. IEEE Int. Conf. Acoustics Speech Signal Process. Proc. 2133–2136. http://dx.doi.org/10.1109/ICASSP.2012.6288333.

Dietterich, T.G., 2000. Ensemble methods in machine learning. In: International Workshop on Multiple Classifier Systems, Springer, pp. 1–15.

Duan, F., Dai, L., 2017. Recognizing the gradual changes in sEMG characteristics based on incremental learning of wavelet neural network ensemble. IEEE Trans. Ind. Electron. 64, 4276–4286. http://dx.doi.org/10.1109/TIE.2016.2593693.

Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm. In: Icml, vol. 96. Citeseer, pp. 148–156.

Friedman, J.H., 1997. On bias, variance, 0/1—loss, and the curse-of-dimensionality. Data Min. Knowl. Discov. 1, 55–77.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Statist. 1189–1232.

Ganaie, M., Tanveer, M., 2022. Ensemble deep random vector functional link network using privileged information for Alzheimer's disease diagnosis. IEEE/ACM Trans. Comput. Biol. Bioinform. http://dx.doi.org/10.1109/TCBB.2022.3170351.

Gao, Z., Wang, L., Soroushmehr, R., Wood, A., Gryak, J., Nallamothu, B., Najarian, K., 2022. Vessel segmentation for X-ray coronary angiography using ensemble methods with deep learning and filter-based features. BMC Medical Imaging 22, 1–17.

Geman, S., Bienenstock, E., Doursat, R., 1992. Neural networks and the bias/variance dilemma. Neural Comput. 4, 1–58.

Gençay, R., Qi, M., 2001. Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging. IEEE Trans. Neural Netw. 12, 726–734.

Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. pp. 249–256.

González, S., García, S., Del Ser, J., Rokach, L., Herrera, F., 2020. A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. Inf. Fusion 64, 205–237.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, (in press) URL: http://www.deeplearningbook.org.

Gopika, D., Azhagusundari, B., 2014. An analysis on ensemble methods in classification tasks. Int. J. Adv. Res. Comput. Commun. Eng. 3, 7423–7427.

Grassmann, F., Mengelkamp, J., Brandl, C., Harsch, S., Zimmermann, M.E., Linkohr, B., Peters, A., Heid, I.M., Palm, C., Weber, B.H., 2018. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. Ophthalmology 125, 1410–1420.

Grmanová, G., Laurinec, P., Rozinajová, V., Ezzeddine, A.B., Lucká, M., Lacko, P., Vrablecová, P., Návrat, P., 2016. Incremental ensemble learning for electricity load forecasting. 13, pp. 97–117,

Guo, X., Liu, F., Ju, Y., Wang, Z., Wang, C., 2016. Human protein subcellular localization with integrated source and multi-label ensemble classifier. Sci. Rep. 6, 28087.

Ha, K., Cho, S., MacLachlan, D., 2005. Response models based on bagging neural networks. J. Interactive Marketing 19, 17–30.

Han, S., Meng, Z., Khan, A.S., Tong, Y., 2016. Incremental boosting convolutional neural network for facial action unit recognition. Adv. Neural Inf. Process. Syst. 29.

Han, B., Sim, J., Adam, H., 2017. Branchout: Regularization for online ensemble tracking with convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3356–3365.

Hansen, L.K., Salamon, P., 1990. Neural network ensembles. IEEE Trans. Pattern Anal. Mach. Intell. 12, 993–1001.

Hassan, B., Abdelrahman, S.E., Bahgat, R., Farag, I., 2019. UESTS: An unsupervised ensemble semantic textual similarity method. IEEE Access 7, 85462–85482.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hido, S., Kashima, H., Takahashi, Y., 2009. Roughly balanced bagging for imbalanced data. Statist. Anal. Data Mining: ASA Data Sci. J. 2, 412–426.

Hinton, G., Vinyals, O., Dean, J., 2015. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

Hochreiter, S., 1991. Untersuchungen zu dynamischen neuronalen netzen. Diploma, Technische UniversitÄT MÜNchen 91.

Hothorn, T., Lausen, B., Benner, A., Radespiel-Tröger, M., 2004. Bagging survival trees. Stat. Med. 23, 77–91.

Hu, M., Suganthan, P., 2022. Representation learning using deep random vector functional link networks for clustering. Pattern Recognit. 108744.

Huang, F., Ash, J., Langford, J., Schapire, R., 2018. Learning deep resnet blocks sequentially using boosting theory. In: International Conference on Machine Learning. PMLR, pp. 2058–2067.

Huang, P.-S., Deng, L., Hasegawa-Johnson, M., He, X., 2013. Random features for kernel deep convex network. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, pp. 3143–3147. http://dx.doi.org/10.1109/ICASSP.2013.6638237.

Huang, D., Lai, J., Wang, C.-D., 2016a. Ensemble clustering using factor graph. Pattern Recognit. 50, 131–142.

Huang, G., Li, Y., Pleiss, G., Liu, Z., Hopcroft, J.E., Weinberger, K.Q., 2017a. Snapshot ensembles: Train 1, get M for free. arXiv preprint arXiv:1704.00109.

Huang, G., Sun, Y., Liu, Z., Sedra, D., Weinberger, K.Q., 2016b. Deep networks with stochastic depth. In: European Conference on Computer Vision. Springer pp. 646–661. http://dx.doi.org/10.1007/978-3-319-46493-0_39.

Huang, D., Wang, C.-D., Lai, J.-H., 2017b. Locally weighted ensemble clustering. IEEE Trans. Cybern. 48, 1460–1473.

Hutchinson, B., Deng, L., Yu, D., 2012. A deep architecture with bilinear modeling of hidden representations: Applications to phonetic recognition. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP. IEEE, pp. 4805–4808. http://dx.doi.org/10.1109/ICASSP.2012.6288994.

Hutchinson, B., Deng, L.L., Yu, D., 2013. Tensor deep stacking networks. IEEE Trans. Pattern Anal. Mach. Intell. 35, 1944–1957. http://dx.doi.org/10.1109/TPAMI.2012.268.

James, G.M., 2003. Variance and bias for general loss functions. Mach. Learn. 51, 115–135.

Ju, C., Bibaut, A., van der Laan, M., 2018. The relative performance of ensemble methods with deep convolutional neural networks for image classification. J. Appl. Stat. 45, 2800–2818.

Ju, C., Combs, M., Lendle, S.D., Franklin, J.M., Wyss, R., Schneeweiss, S., van der Laan, M.J., 2019. Propensity score prediction for electronic healthcare databases using super learner and high-dimensional propensity score methods. J. Appl. Stat. 46, 2216–2236.

Kang, T., Chen, P., Quackenbush, J., Ding, W., 2020. A novel deep learning model by stacking conditional restricted Boltzmann machine and deep neural network. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1316–1324.

Katuwal, R., Suganthan, P.N., 2019. Stacked autoencoder based deep random vector functional link neural network for classification. Appl. Soft Comput. 85, 105854.

Kazemi, S., Minaei Bidgoli, B., Shamshirband, S., Karimi, S.M., Ghorbani, M.A., Chau, K.-w., Kazem Pour, R., 2018. Novel genetic-based negative correlation learning for estimating soil temperature. Eng. Appl. Comput. Fluid Mech. 12, 506–516.

Khwaja, A., Naeem, M., Anpalagan, A., Venetsanopoulos, A., Venkatesh, B., 2015. Improved short-term load forecasting using bagged neural networks. Electr. Power Syst. Res. 125, 109–115.

Kilimci, Z.H., Akyokus, S., 2018. Deep learning-and word embedding-based heterogeneous classifier ensembles for text classification. Complexity 2018.

Kim, H.-C., Pang, S., Je, H.-M., Kim, D., Bang, S.-Y., 2002. Support vector machine ensemble with bagging. In: International Workshop on Support Vector Machines, Springer, pp. 397–408.

Kimura, K., Kudo, M., Sun, L., Koujaku, S., 2016. Fast random k-labelsets for large-scale multi-label classification. In: 2016 23rd International Conference on Pattern Recognition, ICPR. IEEE, pp. 438–443.

Kleinberg, E., 1990. Stochastic discrimination. Ann. Math. Artif. Intell. 1, 207–239.

Kohavi, R., Wolpert, D.H., 1996. Bias plus variance decomposition for zero-one loss functions. In: ICML, vol. 96. pp. 275–283.

Kong, E.B., Dietterich, T.G., 1995. Error-correcting output coding corrects bias and variance. In: Machine Learning Proceedings 1995. Elsevier, pp. 313–321.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 1097–1105.

Krogh, A., Vedelsby, J., 1995. Neural network ensembles, cross validation, and active learning. In: Advances in Neural Information Processing Systems. pp. 231–238.

Kuncheva, L.I., Whitaker, C.J., Shipp, C.A., Duin, R.P., 2003. Limits on the majority vote accuracy in classifier fusion. Pattern Anal. Appl. 6, 22–31.

Kuznetsov, V., Mohri, M., Syed, U., 2014. Multi-class deep boosting. Adv. Neural Inf. Process. Syst. 3, 2501–2509.

Van der Laan, M.J., Polley, E.C., Hubbard, A.E., 2007. Super learner. Statist. Appl. Genetics Molecul. Biol. 6.

Lahiri, A., Roy, A.G., Sheet, D., Biswas, P.K., 2016. Deep neural ensemble for retinal vessel segmentation in fundus images towards achieving label-free angiography. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC. IEEE, pp. 1340–1343.

Laine, S., Aila, T., 2016. Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242.

Leblanc, M., Tibshirani, R., 1996. Combining estimates in regression and classification. J. Amer. Statist. Assoc. 91, 1641–1650. http://dx.doi.org/10.1080/01621459.1996.10476733.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Li, J., Chang, H., Yang, J., 2015. Sparse deep stacking network for image classification. In: Twenty-Ninth AAAI Conference on Artificial Intelligence.

Li, J., Chang, H., Yang, J., Luo, W., Fu, Y., 2017a. Visual representation and classification by learning group sparse deep stacking network. IEEE Trans. Image Process. 27, 464–476.

Li, W., Ding, S., Chen, Y., Yang, S., 2018. Heterogeneous ensemble for default prediction of peer-to-peer lending in China. IEEE Access 6, 54396–54406.

Li, P., Li, H., Wu, M., 2013. Multi-label ensemble based on variable pairwise constraint projection. Inform. Sci. 222, 269–281. http://dx.doi.org/10.1016/j.ins.2012.07.066, Including Special Section on New Trends in Ambient Intelligence and Bio-inspired Systems.

Li, S., Lu, X., Sakai, S., Mimura, M., Kawahara, T., 2017b. Semi-supervised ensemble DNN acoustic model training. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP. IEEE, pp. 5270–5274.

Li, Y., Song, Y., Jia, L., Gao, S., Li, Q., Qiu, M., 2020. Intelligent fault diagnosis by fusing domain adversarial training and maximum mean discrepancy via ensemble learning. IEEE Trans. Ind. Inf. 17, 2833–2841.

Li, J., Wu, S., Liu, C., Yu, Z., Wong, H.-S., 2019a. Semi-supervised deep coupled ensemble learning with classification landmark exploration. IEEE Trans. Image Process. 29, 538–550.

Li, X., Yang, Y., Pan, H., Cheng, J., Cheng, J., 2019b. A novel deep stacking least squares support vector machine for rolling bearing fault diagnosis. Comput. Ind. 110, 36–47. http://dx.doi.org/10.1016/j.compind.2019.05.005.

Liu, B., Gu, L., Lu, F., 2019. Unsupervised ensemble strategy for retinal vessel segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 111–119.

Liu, P., Han, S., Meng, Z., Tong, Y., 2014. Facial expression recognition via a boosted deep belief network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1805–1812.

Liu, H., Liu, T., Wu, J., Tao, D., Fu, Y., 2015. Spectral ensemble clustering. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 715–724.

Liu, H., Shao, M., Li, S., Fu, Y., 2016. Infinite ensemble for image clustering. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1745–1754.

Liu, F., Xu, F., Yang, S., 2017. A flood forecasting model based on deep learning algorithm via integrating stacked autoencoders with BP neural network. In: 2017 IEEE Third International Conference on Multimedia Big Data, BigMM. IEEE pp. 58–61.

Liu, Y., Yao, X., 1999. Ensemble learning via negative correlation. Neural Netw. 12, 1399–1404. http://dx.doi.org/10.1016/S0893-6080(99)00073-8.

Liu, J., Zhao, S., Wang, G., 2018. SSEL-ADE: a semi-supervised ensemble learning framework for extracting adverse drug events from social media. Artif. Intell. Med. 84, 34–49.

Low, C.-Y., Park, J., Teoh, A.B.-J., 2019. Stacking-based deep neural network: Deep analytic network for pattern classification. IEEE Trans. Cybern. 50, 5021–5034.

Malik, A.K., Gao, R., Ganaie, M.A., Tanveer, M., Suganthan, P.N., 2022. Random vector functional link network: recent developments, applications, and future directions. arXiv preprint arXiv:2203.11316.

Mao, J., 1998. A case study on bagging, boosting and basic ensembles of neural networks for OCR. In: 1998 IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence, Cat. No. 98CH36227, vol. 3. IEEE, pp. 1828–1833.

Martín, A., Lara-Cabrera, R., Camacho, D., 2019. Android malware detection through hybrid features fusion and ensemble classifiers: The AndroPyTool framework and the OmniDroid dataset. Inf. Fusion 52, 128–142.

Melville, P., Mooney, R.J., 2003. Constructing diverse classifier ensembles using artificial training examples. In: IJCAI, vol. 3. pp. 505–510.

Melville, P., Mooney, R.J., 2004. Diverse ensembles for active learning. In: Proceedings of the Twenty-First International Conference on Machine Learning. p. 74.

Mendes-Moreira, J., Soares, C., Jorge, A.M., Sousa, J.F.D., 2012. Ensemble approaches for regression: A survey. ACM Comput. Surv. 45, 10.

Minku, L.L., White, A.P., Yao, X., 2009. The impact of diversity on online ensemble learning in the presence of concept drift. IEEE Trans. Knowl. Data Eng. 22, 730–742.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., Riedmiller, M., 2013. Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602.

Moghimi, M., Belongie, S.J., Saberian, M.J., Yang, J., Vasconcelos, N., Li, L.-J., 2016. Boosted convolutional neural networks. In: BMVC. pp. 1–24.

Mosca, A., Magoulas, G.D., 2017. Deep incremental boosting. arXiv preprint arXiv:1708.03704.

Moyano, J.M., Gibaja, E.L., Cios, K.J., Ventura, S., 2019. An evolutionary approach to build ensembles of multi-label classifiers. Inf. Fusion 50, 168–180.

Muhlbaier, M.D., Polikar, R., 2007. An ensemble approach for incremental learning in nonstationary environments. In: Haindl, M., Kittler, J., Roli, F. (Eds.), Multiple Classifier Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 490–500.

Opitz, M., Waltner, G., Possegger, H., Bischof, H., 2017. Bier-boosting independent embeddings robustly. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 5189–5198.

Oza, N.C., 2005. Online bagging and boosting. In: 2005 IEEE International Conference on Systems, Man and Cybernetics, vol. 3. Ieee, pp. 2340–2345.

Palangi, H., Deng, L., Ward, R.K., 2014. Recurrent deep-stacking networks for sequence classification. In: 2014 IEEE China Summit and International Conference on Signal and Information Processing, IEEE ChinaSIP 2014 - Proceedings. pp. 510–514. http://dx.doi.org/10.1109/ChinaSIP.2014.6889295.

Pao, Y.-H., Park, G.-H., Sobajic, D.J., 1994. Learning and generalization characteristics of the random vector functional-link net. Neurocomputing 6, 163–180.

Pedro, D., 2000. A unified bias-variance decomposition and its applications. In: 17th International Conference on Machine Learning. pp. 231–238.

Pio, G., Malerba, D., D'Elia, D., Ceci, M., 2014. Integrating microRNA target predictions for the discovery of gene regulatory networks: a semi-supervised ensemble learning approach. BMC Bioinformatics 15, S4.

Pisetta, V., 2012. New Insights into Decision Trees Ensembles. (Ph.D. thesis). Lyon 2.

Qiu, X., Ren, Y., Suganthan, P.N., Amaratunga, G.A., 2017. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. Appl. Soft Comput. 54, 246–255.

Qiu, X., Suganthan, P.N., Amaratunga, G.A., 2018. Ensemble incremental learning random vector functional link network for short-term electric load forecasting. Knowl.-Based Syst. 145, 182–196. http://dx.doi.org/10.1016/j.knosys.2018.01.015.

Qiu, X., Zhang, L., Ren, Y., Suganthan, P.N., Amaratunga, G., 2014. Ensemble deep learning for regression and time series forecasting. In: 2014 IEEE Symposium on Computational Intelligence in Ensemble Learning, CIEL. IEEE, pp. 1–6.

Rai, H.M., Chatterjee, K., 2022. Hybrid CNN-LSTM deep learning model and ensemble technique for automatic detection of myocardial infarction using big ECG data. Appl. Intell. 52, 5366–5384.

Randhawa, K., Loo, C.K., Seera, M., Lim, C.P., Nandi, A.K., 2018. Credit card fraud detection using AdaBoost and majority voting. IEEE Access 6, 14277–14284.

Rath, A., Mishra, D., Panda, G., Satapathy, S.C., Xia, K., 2022. Improved heart disease detection from ECG signal using deep learning based ensemble model. Sustain. Comput.: Inform. Syst. 35, 100732.

Read, J., Pfahringer, B., Holmes, G., Frank, E., 2011. Classifier chains for multi-label classification. Mach. Learn. 85, 333.

Ren, Y., Suganthan, P., Srikanth, N., 2015. Ensemble methods for wind and solar power forecasting—A state-of-the-art review. Renew. Sustain. Energy Rev. 50, 82–91.

Ren, Y., Zhang, L., Suganthan, P.N., 2016. Ensemble classification and regression-recent developments, applications and future directions. IEEE Comput. Intell. Mag. 11, 41–53.

Rokach, L., 2010. Ensemble-based classifiers. Artif. Intell. Rev. 33, 1–39.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., 2015. Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. 115, 211–252.

Sagi, O., Rokach, L., 2018. Ensemble learning: A survey. Wiley Interdiscip. Rev.: Data Mining Knowl. Discov. 8, e1249.

Schapire, R.E., Freund, Y., Bartlett, P., Lee, W.S., 1998. Boosting the margin: A new explanation for the effectiveness of voting methods. Ann. Statist. 26, 1651–1686.

Schietgat, L., Vens, C., Struyf, J., Blockeel, H., Kocev, D., Džeroski, S., 2010. Predicting gene function using hierarchical multi-label decision tree ensembles. BMC Bioinformatics 11, 2.

Seijo-Pardo, B., Porto-Díaz, I., Bolón-Canedo, V., Alonso-Betanzos, A., 2017. Ensemble feature selection: homogeneous and heterogeneous approaches. Knowl.-Based Syst. 118, 124–139.

Shaham, U., Cheng, X., Dror, O., Jaffe, A., Nadler, B., Chang, J., Kluger, Y., 2016. A deep learning approach to unsupervised ensemble learning. In: International Conference on Machine Learning. pp. 30–39.

Shalbaf, A., Vafaeezadeh, M., 2021. Automated detection of COVID-19 using ensemble of transfer learning with deep convolutional neural network based on CT scans. Int. J. Comput. Assist. Radiol. Surg. 16, 115–123.

Sharma, A., Rani, R., 2018. BE-DTI': Ensemble framework for drug target interaction prediction using dimensionality reduction and active learning. Comput. Methods Programs Biomed. 165, 151–162.

Shi, Q., Katuwal, R., Suganthan, P.N., Tanveer, M., 2021. Random vector functional link neural network based ensemble deep learning. Pattern Recognit. 117, 107978.

Shi, C., Kong, X., Yu, P.S., Wang, B., 2011. Multi-label ensemble learning. In: Gunopulos, D., Hofmann, T., Malerba, D., Vazirgiannis, M. (Eds.), Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 223–239.

Shi, Z., Zhang, L., Liu, Y., Cao, X., Ye, Y., Cheng, M.-M., Zheng, G., 2018. Crowd counting with deep negative correlation learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5382–5390.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Singh, S., Hoiem, D., Forsyth, D., 2016. Swapout: Learning an ensemble of deep architectures. In: Lee, D.D., Sugiyama, M., Luxburg, U.V., Guyon, I., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 29. Curran Associates, Inc., pp. 28–36.

Singla, P., Duhan, M., Saroha, S., 2021. An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network. Earth Sci. Inform. 1–16.

Siu, C., 2019. Residual networks behave like boosting algorithms. In: 2019 IEEE International Conference on Data Science and Advanced Analytics, DSAA. IEEE, pp. 31–40.

Smith, L.N., Hand, E.M., Doster, T., 2016. Gradual dropin of layers to train very deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4763–4771.

Srivastava, R.K., Greff, K., Schmidhuber, J., 2015. Training very deep networks. In: Advances in Neural Information Processing Systems. pp. 2377–2385.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., Mele, B., Altarelli, G., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958. http://dx.doi.org/10.1016/0370-2693(93)90272-J.

Sukegawa, S., Fujimura, A., Taguchi, A., Yamamoto, N., Kitamura, A., Goto, R., Nakano, K., Takabatake, K., Kawai, H., Nagatsuka, H., Furuki, Y., 2022. Identification of osteoporosis using ensemble deep learning model with panoramic radiographs and clinical covariates. Sci. Rep. 12, 1–10.

Sun, C., Ma, M., Zhao, Z., Chen, X., 2018. Sparse deep stacking network for fault diagnosis of motor. IEEE Trans. Ind. Inf. 14, 3261–3270.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–9.

Tabik, S., Alvear-Sandoval, R.F., Ruiz, M.M., Sancho-Gómez, J.-L., Figueiras-Vidal, A.R., Herrera, F., 2020. MNIST-NET10: A heterogeneous deep networks fusion based on the degree of certainty to reach 0.1% error rate. ensembles overview and proposal. Inf. Fusion 62, 73–80. http://dx.doi.org/10.1016/j.inffus.2020.04.002.

Tang, K.-F., Kao, H.-C., Chou, C.-N., Chang, E.Y., 2016. Inquire and diagnose: Neural symptom checking ensemble using deep reinforcement learning. In: Proceedings of NIPS Workshop on Deep Reinforcement Learning.

Tang, K., Lin, M., Minku, F.L., Yao, X., 2009. Selective negative correlation learning approach to incremental learning. Neurocomputing 72, 2796–2805. http://dx.doi.org/10.1016/j.neucom.2008.09.022.

Tanveer, M., Rashid, A., Ganaie, M., Razzak, I., Hua, K.-L., 2022. Classification of Alzheimer's disease using ensemble of deep neural networks trained through transfer learning. IEEE J. Biomed. Health Inf. 26, 1453–1463. http://dx.doi.org/10.1109/JBHI.2021.3083274.

Tao, D., Tang, X., Li, X., Wu, X., 2006. Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. IEEE Trans. Pattern Anal. Mach. Intell. 28, 1088–1099.

Tsogbaatar, E., Bhuyan, M.H., Taenaka, Y., Fall, D., Gonchigsumlaa, K., Elmroth, E., Kadobayashi, Y., 2021. Del-IoT: A deep ensemble learning approach to uncover anomalies in IoT. Internet of Things 14, 100391.

Tsoumakas, G., Katakis, I., 2007. Multi-label classification: An overview. Int. J. Data Warehousing Mining 3, 1–13.

Tsoumakas, G., Vlahavas, I., 2007. Random k-labelsets: An ensemble method for multilabel classification. In: European Conference on Machine Learning. Springer, pp. 406–417.

Tur, G., Deng, L., Hakkani-Tür, D., He, X., 2012. Towards deeper understanding: Deep convex networks for semantic utterance classification. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP. IEEE, pp. 5045–5048.

Vega-Pons, S., Ruiz-Shulcloper, J., 2011. A survey of clustering ensemble algorithms. Int. J. Pattern Recognit. Artif. Intell. 25, 337–372.

Veit, A., Wilber, M.J., Belongie, S., 2016. Residual networks behave like ensembles of relatively shallow networks. In: Advances in Neural Information Processing Systems. pp. 550–558.

Walach, E., Wolf, L., 2016. Learning to count with cnn boosting. In: European Conference on Computer Vision. Springer, pp. 660–676. http://dx.doi.org/10.1007/978-3-319-46475-6_41.

Waltner, G., Opitz, M., Possegger, H., Bischof, H., 2019. HiBsteR: Hierarchical boosted deep metric learning for image retrieval. In: 2019 IEEE Winter Conference on Applications of Computer Vision, WACV. IEEE, pp. 599–608.

Wan, L., Zeiler, M., Zhang, S., Cun, Y.L., Fergus, R., 2013. Regularization of neural networks using DropConnect. In: Dasgupta, S., McAllester, D. (Eds.), Proceedings of the 30th International Conference on Machine Learning. In: Proceedings of Machine Learning Research, vol. 28, PMLR, Atlanta, Georgia, USA, pp. 1058–1066. http://dx.doi.org/10.1109/TPAMI.2017.2703082.

Wang, J., Feng, K., Wu, J., 2019a. SVM-based deep stacking networks. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33. pp. 5273–5280. http://dx.doi.org/10.1609/aaai.v33i01.33015273.

Wang, F., Jiang, D., Wen, H., Song, H., 2019b. Adaboost-based security level classification of mobile intelligent terminals. J. Supercomput. 75, 7460–7478.

Wang, X., Kihara, D., Luo, J., Qi, G.-J., 2019c. ENAET: self-trained ensemble autoencoding transformations for semi-supervised learning. arXiv preprint arXiv:1911.09265.

Wang, R., Kwong, S., Wang, X., Jia, Y., 2021. Active k-labelsets ensemble for multi-label classification. Pattern Recognit. 109, 107583.

Wang, G., Sun, J., Ma, J., Xu, K., Gu, J., 2014. Sentiment classification: The contribution of ensemble learning. Decis. Support Syst. 57, 77–93.

Wang, B., Xue, B., Zhang, M., 2020. Particle swarm optimisation for evolving deep neural networks for image classification by evolving and stacking transferable blocks. In: 2020 IEEE Congress on Evolutionary Computation, CEC. IEEE, pp. 1–8.

Wang, G., Zhang, G., Choi, K.S., Lu, J., 2019d. Deep additive least squares support vector machines for classification with model transfer. IEEE Trans. Syst. Man Cybernet.: Syst. 49, 1527–1540. http://dx.doi.org/10.1109/TSMC.2017.2759090.

Welchowski, T., Schmid, M., 2016. A framework for parameter estimation and model selection in kernel deep stacking networks. Artif. Intell. Med. 70, 31–40. http://dx.doi.org/10.1016/j.artmed.2016.04.002.

Wen, L., Xie, X., Li, X., Gao, L., 2020. A new ensemble convolutional neural network with diversity regularization for fault diagnosis. J. Manuf. Syst. 62, 964–971. http://dx.doi.org/10.1016/j.jmsy.2020.12.002.

Wolpert, D.H., 1992. Stacked generalization. Neural Netw. 5, 241–259.

Wolpert, D.H., 1997. On bias plus variance. Neural Comput. 9, 1211–1243.

Xia, Y., Chen, K., Yang, Y., 2021. Multi-label classification with weighted classifier selection and stacked ensemble. Inform. Sci. 557, 421–442.

Xie, J., Xu, B., Chuang, Z., 2013. Horizontal and vertical ensemble with deep representation for classification. arXiv preprint arXiv:1306.2759.

Xue, J., Wang, Z., Kong, D., Wang, Y., Liu, X., Fan, W., Yuan, S., Niu, S., Li, D., 2021. Deep ensemble neural-like P systems for segmentation of central serous chorioretinopathy lesion. Inf. Fusion 65, 84–94.

Yang, P., Hwa Yang, Y., B. Zhou, B., Y. Zomaya, A., 2010. A review of ensemble methods in bioinformatics. Curr. Bioinform. 5, 296–308.

Yang, H., Liu, X.-Y., Zhong, S., Walid, A., 2020. Deep reinforcement learning for automated stock trading: An ensemble strategy. In: Proceedings of the First ACM International Conference on AI in Finance. pp. 1–8.

Yang, B., Yan, J., Lei, Z., Li, S.Z., 2015. Convolutional channel features. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 82–90.

Yin, Z., Zhao, M., Wang, Y., Yang, J., Zhang, J., 2017. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. Comput. Methods Programs Biomed. 140, 93–110.

Yu, G., Domeniconi, C., Rangwala, H., Zhang, G., Yu, Z., 2012. Transductive multi-label ensemble classification for protein function prediction. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1077–1085.

Zareapoor, M., Shamsolmoali, P., 2015. Application of credit card fraud detection: Based on bagging ensemble classifier. Procedia Comput. Sci. 48, 679–685.

Zhang, H., Dai, Y., Li, H., Koniusz, P., 2019a. Deep stacked hierarchical multi-patch network for image deblurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5978–5986.

Zhang, W., Jiang, J., Shao, Y., Cui, B., 2020a. Snapshot boosting: a fast ensemble framework for deep neural networks. Sci. China Inf. Sci. 63, 112102.

Zhang, W., Liu, F., Luo, L., Zhang, J., 2015. Predicting drug side effects by multi-label learning and ensemble learning. BMC Bioinformatics 16, 365.

Zhang, H., Liu, W., Shan, J., Liu, Q., 2018. Online active learning paired ensemble for concept drift and class imbalance. IEEE Access 6, 73815–73828. http://dx.doi.org/10.1109/ACCESS.2018.2882872.

Zhang, L., Shi, Z., Cheng, M.-M., Liu, Y., Bian, J.-W., Zhou, J.T., Zheng, G., Zeng, Z., 2019b. Nonlinear regression via deep negative correlation learning. IEEE Trans. Pattern Anal. Mach. Intell. PP, 1. http://dx.doi.org/10.1109/tpami.2019.2943860.

Zhang, C.-X., Zhang, J.-S., 2008. RotBoost: a technique for combining rotation forest and AdaBoost. Pattern Recognit. Lett. 29, 1524–1536.

Zhang, J., Zhang, W., Song, R., Ma, L., Li, Y., 2020b. Grasp for stacking via deep reinforcement learning. In: 2020 IEEE International Conference on Robotics and Automation, ICRA. IEEE, pp. 2543–2549.

Zhao, Y., Gao, J., Yang, X., 2005. A survey of neural network ensembles. In: 2005 International Conference on Neural Networks and Brain, vol. 1. IEEE, pp. 438–442.

Zhao, Q.L., Jiang, Y.H., Xu, M., 2010. Incremental learning by heterogeneous bagging ensemble. In: Cao, L., Zhong, J., Feng, Y. (Eds.), Advanced Data Mining and Applications. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–12.

Zheng, L., Li, T., Ding, C., 2010. Hierarchical ensemble clustering. In: 2010 IEEE International Conference on Data Mining. IEEE, pp. 1199–1204.

Zheng, X., Shen, J., Cox, C., Wakefield, J.C., Ehm, M.G., Nelson, M.R., Weir, B.S., 2014. HIBAG—HLA genotype imputation with attribute bagging. The Pharmacogenomics J. 14, 192–200.

Zhou, Z.-H., Chawla, N.V., Jin, Y., Williams, G.J., 2014. Big data opportunities and challenges: Discussions from data analytics perspectives [discussion forum]. IEEE Comput. Intell. Mag. 9, 62–74.

Zhou, Z.-H., Feng, J., 2017. Deep forest. arXiv preprint arXiv:1702.08835.

Zhou, H., Huang, G.B., Lin, Z., Wang, H., Soh, Y.C., 2015. Stacked extreme learning machines. IEEE Trans. Cybern. 45, 2013–2025. http://dx.doi.org/10.1109/TCYB.2014.2363492.

Zhou, T., Lu, H., Yang, Z., Qiu, S., Huo, B., Dong, Y., 2021. The ensemble deep learning model for novel COVID-19 on CT images. Appl. Soft Comput. 98, 106885.

Zhou, Z.-H., Tang, W., 2006. Clusterer ensemble. Knowl.-Based Syst. 19, 77–83.