IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 September 2023, accepted 22 September 2023, date of publication 27 September 2023,
date of current version 4 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3320042

TOPICAL REVIEW

Ensemble Multifeatured Deep Learning Models and Applications: A Survey

SATHEESH ABIMANNAN[1], EL-SAYED M. EL-ALFY[2,3], (Senior Member, IEEE), YUE-SHAN CHANG[4], (Senior Member, IEEE), SHAHID HUSSAIN[5], SAURABH SHUKLA[6], AND DHIVYADHARSINI SATHEESH[7]

[1]Amity School of Engineering and Technology (ASET), Amity University Maharashtra, Mumbai 410206, India
[2]SDAIA-KFUPM Joint Research Center for Artificial Intelligence, Information and Computer Science Department, King Fahd University of Petroleum and Minerals, Dhahran 34464, Saudi Arabia
[3]Interdisciplinary Research Center of Intelligent Secure Systems, Information and Computer Science Department, King Fahd University of Petroleum and Minerals, Dhahran 34464, Saudi Arabia
[4]Department of Computer Science and Information Engineering, National Taipei University, New Taipei City 237, Taiwan
[5]Innovative Value Institute (IVI), School of Business, National University of Ireland Maynooth (NUIM), Maynooth, W23 F2H6 Ireland
[6]Department of Computer Science (CS), Indian Institute of Information Technology Lucknow (IIIT Lucknow), Lucknow 226002, India
[7]School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology (VIT), Vellore 632014, India

Corresponding authors: Yue-Shan Chang (ysc@mail.ntpu.edu.tw) and Satheesh Abimannan (sabimannan@mum.amity.edu)

This work was supported in part by the National Taipei University under Grant 112-NTPU_ORDA-F-004, and in part by the National Science and Technology Council under Grant NSTC 111-2221-E-305-011-MY3. The work of El-Sayed M. El-Alfy was supported by the Fellowship with the SDAIA-KFUPM Joint Research Center for Artificial Intelligence under Grant JRC-AI-RFP-04.

ABSTRACT Ensemble multifeatured deep learning methodology has emerged as a powerful approach to overcome the limitations of single deep learning models in terms of generalization, robustness, and performance. This survey provides an extended review of ensemble multifeatured deep learning models, and their applications, challenges, and future directions. We explore potential applications of these models across various domains, including computer vision, medical imaging, natural language processing, and speech recognition. By combining the strengths of multiple models and features, ensemble multifeatured deep learning models have demonstrated improved performance and adaptability in diverse problem settings. We also discuss the challenges associated with these models, such as model interpretability, computational complexity, ensemble model selection, adversarial robustness, and personalized and federated learning. This survey highlights recent advancements in addressing these challenges and emphasizes the importance of continued research in tackling these issues to enable widespread adoption of ensemble multifeatured deep learning models. It provides an outlook on future research directions, focusing on the development of new algorithms, frameworks, and hardware architectures that can efficiently handle the large-scale computations required by these models. Moreover, it underlines the need for a better understanding of the trade-offs between model complexity, accuracy, and computational resources to optimize the design and deployment of ensemble multifeatured deep learning models.

INDEX TERMS Ensemble multifeatured deep learning models, model interpretability, computational complexity, ensemble model selection, adversarial robustness, personalized and federated learning.

I. INTRODUCTION

Deep learning has revolutionized various fields, including computer vision, natural language processing, and speech recognition, among others [1], [2], [3], [4]. With the increasing complexity of real-world problems and the availability of vast amounts of data, deep learning models have shown remarkable success in a wide range of applications. However, a single deep learning model may have limitations in terms of generalization, robustness, and performance. Ensemble learning addresses these limitations by combining multiple models to improve predictive performance and

The associate editor coordinating the review of this manuscript and approving it for publication was Vicente Alarcon-Aquino.

reduce errors by optimally balancing the bias-variance tradeoff, more specifically when dealing with complex problem of various inter-related modalities and imbalanced noisy datasets [5].

Ensemble multifeatured deep learning is a framework that leverages multiple deep learning algorithms for feature selection and employs a sophisticated ensemble algorithm to consolidate the outcomes of contributing algorithms. This methodology mitigates information loss and overfitting issues associated with single models, as well as addressing the imbalanced data problem prevalent in multimedia big data and large-scale applications. The development of ensemble learning algorithms in traditional machine learning such as bagging, boosting and stacking, has been around since 1990s and are booming since 2000s after the remarkable success of deep learning [6], [7]. The framework has demonstrated its effectiveness in tasks such as semantic event detection and has outperformed numerous cutting-edge deep learning architectures. Ensemble multifeatured deep learning models have gained significant attention due to their ability to capitalize on the strengths of multiple models and feature representations. These ensembles can effectively integrate diverse sources of information, leading to improved performance across a wide range of applications [8], [9], [10].

Figure 1 illustrates the generic high-level layered architecture of ensemble multifeatured deep learning, which consists of multiple input features, deep learning models, fusion layers, and an ensemble algorithm. In this architecture, various input features of different modalities such as text, images, audio, or video, are represented as separate branches or nodes and connected to their corresponding deep learning models (e.g., CNN for images, RNN for text). These deep learning models extract relevant features from their respective input data, which are then combined through the fusion layers. The fusion layers can use various combining techniques such as concatenation, averaging, or weighting as well as attention based, to merge the learned features from the baseline deep learning models [11], [12]. After the fusion layer, the combined features are connected to an ensemble algorithm, such as voting, stacking, or boosting. This ensemble algorithm combines the predictions from each deep learning model to produce a final output or prediction. The overall goal of the architecture depicted in Figure 1 is to leverage the strengths of multiple deep learning models and ensemble techniques to enhance feature selection and classification performance, ultimately leading to better predictive outcomes. This survey paper provides a comprehensive and detailed overview of ensemble multifeatured deep learning models, discussing their methodologies, challenges, applications in various domains, and future research directions.

The primary objectives of this survey are:

- To present a thorough review of the literature on ensemble multifeatured deep learning models, highlighting their methodological contributions and applications in different domains.

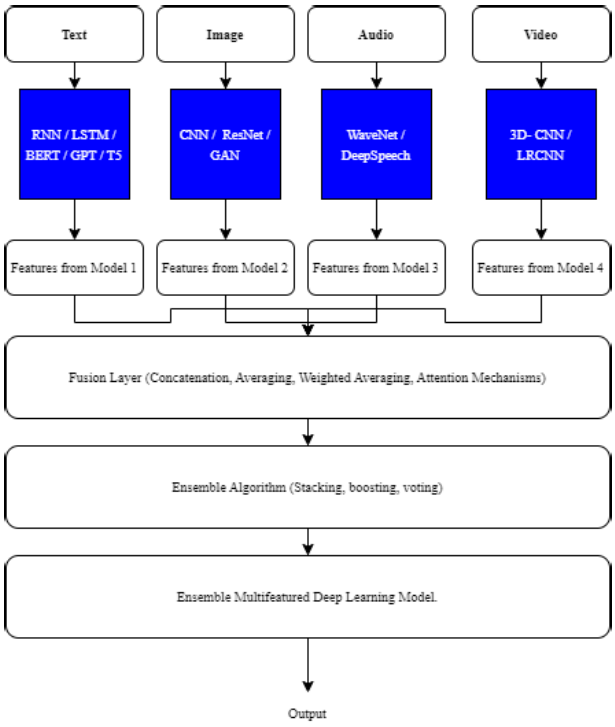


FIGURE 1. Generic high-level layered architecture of ensemble multifeatured deep learning with fusion layer and ensemble algorithm.

- To identify and discuss the challenges associated with implementing ensemble multifeatured deep learning models, emphasizing potential solutions and research directions.
- To propose future research directions and opportunities in the field of ensemble multifeatured deep learning models, considering the current state of the art and existing challenges.

To accomplish these objectives, we employed a systematic methodology to gather and analyze pertinent literature. Our search focused on research articles and conference papers published in reputable journals and conferences, such as Elsevier, Springer, and IEEE. The data sources for our analysis encompassed academic databases, preprint repositories, and relevant research articles. We attempted to include the most recent and relevant literature in the discussion by considering an extensive range of sources and emphasizing the most impactful and innovative works. Our survey encompasses papers published from 1990 to 2023, and through a systematic filtering process, we selected 222 papers as the foundation for this survey paper.

The rest of the paper is organized as follows. Section II presents essential background to delineate the difference between ensemble learning and model fusion and briefly discuss the common ensemble learning techniques. Section III, IV, V, and VI discuss the applications of ensemble multifeatured deep learning models, covering various domains, such as computer vision, medical imaging, natural language processing, and speech recognition, respectively.

For each domain, we review the state-of-the-art approaches, their underlying methodologies, and the results achieved. We also discuss the specific challenges and opportunities associated with each application area. Section VII presents the challenges and future directions associated with ensemble multifeatured deep learning models. We identify and discuss the critical challenges that researchers and practitioners face when developing and deploying these models, including model interpretability, computational complexity, ensemble model selection, adversarial robustness, personalized and federated learning, and data privacy. For each challenge, we review the existing solutions and highlight potential future research directions. Finally, Section VIII concludes the survey with a summary of the main findings, research gaps, and future research directions. We synthesize the insights gained from the analysis of the applications, methodologies, and challenges associated with ensemble multifeatured deep learning models and provide recommendations for future research. By doing so, we aim to contribute to the development and widespread adoption of ensemble multifeatured deep learning models in various applications, ensuring their effectiveness, reliability, and robustness in addressing complex real-world problems.

II. BACKGROUND

Ensemble learning is a paradigm where multiple models, known as ''base learners,'' are trained to solve the same problem and are then combined to improve overall performance. This technique effectively leverages the strength of each individual model to improve generalization and reduce errors [13].

- *Bagging (Bootstrap Aggregating):* This involves training multiple instances of the same model in parallel on different subsamples of the training data. The final prediction is typically an average (for regression) or a majority vote (for classification) from all the models. A classic example is the Random Forest, where multiple decision trees are trained on bootstrapped samples of the dataset [14].
- *Boosting:* This adaptive technique trains models sequentially, with each new model being trained to correct the errors of the combined ensemble of existing models. Famous algorithms like AdaBoost and Gradient Boosted Trees fall under this category [15].
- *Stacking:* Here, several different models are trained, and their predictions are used as input to a ''meta-learner'' or ''combiner'' which then makes the final prediction. The ''meta-learner'' can be any algorithm, with common choices being linear regression, decision trees, or neural networks [16].

Another concept that aligns closely with ensemble learning is information fusion, which is used in a broader sense and involves the amalgamation of information from multiple sources (which could be raw data, features, modalities, views, algorithms or models) and can occur at different levels (sensor-level, feature-level, score-level, decision-level, or rank-level) [17], [18], [19], [20]. This integration aims to enhance the overall performance of a predictive task. While both model fusion and ensemble learning of models exhibit certain similarities, they also manifest distinct fundamental differences. These disparities contribute to their separate roles and applications in the domain of machine learning. References such as [21], [22], [23], [24], [25], [26], [27], [28], [29], and [30] provide additional insights into the concept of model fusion. Ensemble learning of models, on the other hand, [31], [32], [33], [34], [35], [36], [37] entail the creation of multiple models, followed by the amalgamation of their outputs to formulate a final decision or prediction. The primary objective here is to capitalize on the unique strengths of individual models, thereby improving generalization and minimizing errors, as described by Du and Swamy in 2019 [38].

Our work particularly emphasizes ensemble techniques, in particular Ensemble Multifeatured Deep Learning Models, where multiple deep learning models with diverse features are integrated. The term ''ensemble'' in this context refers to the collective decision-making process from multiple models rather than the fusion of their features. This is in line with the approach taken by [21], who utilized both feature fusion and ensemble learning for image classification tasks. Ensemble Multifeatured Deep Learning leverages the capabilities of ensemble learning in conjunction with the representational power of deep neural networks. It's worth noting that the landscape of ensemble learning is vast and continues to evolve, especially with the integration of deep learning models [38], [39].

III. APPLICATIONS IN COMPUTER VISION

Computer vision is a rapidly evolving field that aims to enable machines to interpret and understand still and stereo visual information from the surrounding world [40]. The primary goal of computer vision is to develop algorithms and techniques that can automatically extract meaningful information from images and videos, such as object recognition [41], scene understanding [2], and motion analysis [42]. However, computer vision faces several challenges, such as variations in lighting conditions, occlusions, and complex cluttered backgrounds [43], which make it difficult to achieve accurate and robust results.

Overcoming these challenges is not just a theoretical concern, but also it is vital for the practical applications of computer vision algorithms [44], [45]. The inability to effectively address these issues can result in unreliable performance, thereby limiting the applicability of computer vision in mission-critical scenarios like autonomous vehicles [46], surveillance systems [47], optical character recognition [48], [49], agricultural automation [50], manufacturing and quality inspection [51], augmented reality [52], and medical imaging [53], [54].

Given the growing range of applications, there is an increasing need for research that focuses on enhancing

the accuracy, robustness, and generalizability of computer vision algorithms. Recent works have pointed out that traditional machine learning techniques often fail to capture the complexity and diversity of visual data, highlighting the need for more advanced approaches, such as ensemble multifeatured deep learning models [3], [55].

A. ENSEMBLE LEARNING IN COMPUTER VISION

Ensemble learning has emerged as a powerful approach in computer vision, leading to substantial improvements in generalization, robustness, and performance of machine learning models [56], [57]. Ensemble learning techniques involve combining multiple models, often referred to as base learners or weak learners, to produce a more accurate prediction than any individual model can achieve [13]. To illustrate, consider a simple example where we have three classifiers: one that identifies shapes, another that identifies colors, and a third that identifies sizes in images. Individually, each classifier may have a certain rate of error. However, when we use ensemble learning to combine these classifiers, we may get a more robust and accurate model for image classification, as the weaknesses of one classifier could be offset by the strengths of others.

In computer vision, ensemble learning has been successfully applied to a wide range of tasks, such as image classification [58], object detection [56], [56], [59], image segmentation [57], human activity recognition [60], crack detection and visual artifacts [61], [62], [63], wear identification [64], and facial recognition [55], [65]. One of the primary advantages of using ensemble learning in computer vision is improved generalization. By aggregating the outputs of multiple models, ensemble learning can reduce the risk of overfitting and enhance the model's ability to generalize to new, unseen data [13]. Popular ensemble techniques include bagging, boosting, and stacking, each with its own strengths and weaknesses [3], [13].

Ensemble learning can also improve the robustness of computer vision models by mitigating the impact of outliers or noisy data [13]. Combining the predictions of multiple models allows the ensemble to effectively filter out noise, while still capturing the underlying structure of the data [3]. Performance improvements are another advantage of ensemble learning in computer vision, as the ensemble can leverage the strengths of multiple models while compensating for their weaknesses [13]. For example, researchers have achieved state-of-the-art results on object detection tasks using ensemble learning techniques that combine multiple deep convolutional neural networks with different architectures and training procedures [56].

Incorporating multiple features in computer vision tasks is essential to capture diverse aspects of the visual data [55], [59]. Different features can represent various aspects of the image, such as color, texture, and shape, and combining these features can lead to more accurate and robust models. In object detection, for instance, researchers have found that combining features from different layers of a convolutional neural network (CNN) can improve the accuracy of the model [59]. Ensemble learning techniques that incorporate multiple features, such as ensemble multifeatured deep learning models, have been shown to achieve superior performance on various computer vision tasks [3], [55].

In conclusion, ensemble learning and incorporating multiple features are critical techniques in computer vision for enhancing the accuracy, robustness, and generalization of machine learning models [55], [56], [57], [59], [66]. These approaches have been widely applied to diverse computer vision tasks, consistently yielding state-of-the-art results.

B. IMAGE CLASSIFICATION

Image classification, a fundamental computer vision task, involves categorizing images into one of several predefined classes based on their content. Convolutional Neural Networks (CNNs) have been the primary approach for this task due to their ability to learn hierarchical feature representations from input images [67], [68]. Researchers continuously work to enhance CNN-based classification models through ensemble learning, multifeature extraction, advanced architectures, and sophisticated data augmentation techniques.

1) ADDRESSING OVERFITTING AND BIAS THROUGH ENSEMBLE LEARNING

Ensemble learning techniques like bagging, boosting, or stacking serve specific purposes in refining CNN-based models for image classification [69]. For instance, bagging is effective in reducing overfitting by training multiple base models on different subsets of the original data and averaging their predictions. Boosting aims to reduce bias by focusing on the instances that are hard to classify, thereby producing a more robust classifier. These ensemble methods not only amalgamate multiple base models to enhance predictive performance but also help in mitigating the individual weaknesses of each model.

2) SIMPLIFIED EXAMPLE OF ENSEMBLE LEARNING

To elucidate, consider an image classification task where one CNN is proficient in recognizing shapes, another excels in identifying colors, and a third is adept at discerning textures. An ensemble model incorporating these three would average or weight their outputs, thereby producing a more accurate and robust classification. Techniques like TresNet [70] employ such ensemble multifeatured approaches to capture diverse aspects-texture, shape, color-of the input image and amalgamate these using ensemble averaging or weighted voting. This multifaceted strategy has proven particularly effective for large and complex datasets like the large-scale visual recognition challenge (ILSVRC) [71].

3) ADVANCEMENTS IN CNN ARCHITECTURES AND DATA AUGMENTATION

Advanced CNN architectures, such as Capsule Networks (CapsNets) [72] and Vision Transformers (ViT) [73], have been developed to better capture and utilize multifeature

representations, while preserving spatial relationships and enhancing generalization capabilities. Multiscale feature representations in CNNs, as implemented in DenseNets [74], concatenate feature maps from multiple layers, facilitating feature reuse and reducing the number of parameters, thereby enabling the model to learn hierarchical features more effectively. Furthermore, advanced data augmentation techniques, such as CutMix [75] and AutoAugment [76], improve CNN generalization by increasing the diversity of training data, encouraging the learning of robust feature representations and reducing overfitting.

In conclusion, ongoing research in image classification aims to improve the performance of CNN-based models through ensemble learning, multifeatured extraction, advanced architectures, and sophisticated data augmentation techniques. These efforts lead to enhanced accuracy, robustness, and generalization capabilities, resulting in more efficient and reliable computer vision systems. Examples of successful applications of ensemble multifeatured deep learning models include TresNet and Mask R-CNN [77], which have achieved state-of-the-art performance on large-scale datasets.

C. OBJECT DETECTION AND SEGMENTATION

Ensemble multifeatured deep learning models have demonstrated considerable potential in enhancing the accuracy and robustness of object detection and segmentation tasks in computer vision. These models integrate multiple feature extractors and classifiers, leveraging their individual strengths and compensating for their weaknesses, resulting in superior performance compared to individual models. In tackling challenges like varying object scales and cluttered backgrounds, ensemble approaches like Faster R-CNN effectively use a Region Proposal Network (RPN) to generate candidate object regions. This modular approach makes it adaptable and robust against these common problems in object detection tasks [78]. In object detection tasks, ensemble multifeatured deep learning models, such as Faster R-CNN [78], employ a Region Proposal Network (RPN) to generate candidate object regions. This network shares convolutional layers with the detection network to reduce computation cost. The detection network then classifies and refines these regions, producing more accurate bounding box predictions. By integrating the RPN and detection network, Faster R-CNN achieves state-of-the-art performance on object detection benchmarks, such as PASCAL VOC and COCO datasets.

For instance, segmentation, Mask R-CNN [77] extends Faster R-CNN by introducing an additional branch for predicting binary masks for each object instance. This branch operates in parallel with the existing bounding box prediction and classification branches, allowing for more precise segmentation. To address challenges such as overlapping objects or complex object shapes, Mask R-CNN leverages a Feature Pyramid Network (FPN) to enable efficient multiscale feature extraction. This feature significantly improves the model's performance in tricky scenarios where traditional methods may fail [79]. The FPN model builds a top-down architecture with lateral connections, enabling efficient multiscale feature extraction. Mask R-CNN has achieved state-of-the-art performance on several benchmark datasets, including COCO.

Semantic segmentation, which involves assigning a class label to each pixel in an image, can also benefit from ensemble multifeatured deep learning models. The authors in [80] employed an ensemble of dilated convolutional networks to capture multi-scale contextual information. This ensemble approach effectively addresses the challenge of delineating object boundaries in dense scenes by using atrous convolutions with varying dilation rates. The additional use of conditional random fields (CRFs) as a post-processing step further enhances segmentation accuracy by refining the object boundaries. By using atrous convolutions with varying dilation rates, DeepLab can effectively increase the receptive field of the model without increasing the number of parameters. Furthermore, the model leverages conditional random fields (CRFs) as a post-processing step to refine the segmentation results. DeepLab has achieved state-of-the-art performance on various benchmark datasets, including PASCAL VOC and Cityscapes.

In conclusion, ensemble multifeatured deep learning models provide an in-depth technical approach to improving the accuracy and robustness of object detection and segmentation tasks in computer vision. By leveraging multiple feature extractors and classifiers, these models enable better performance on various tasks, such as bounding box predictions, instance segmentation, and semantic segmentation. Successful applications of these models include Faster R-CNN, Mask R-CNN, and DeepLab, which have achieved state-of-the-art performance on several benchmark datasets.

D. SCENE UNDERSTANDING AND DEPTH ESTIMATION

Ensemble multifeatured deep learning models have demonstrated significant potential in enhancing the accuracy of depth maps and scene parsing in various computer vision tasks. By incorporating multiple feature extractors and classifiers, these models create a comprehensive representation of the input data, leading to improved performance compared to individual models.

One successful application of ensemble multifeatured deep learning models is the Multi-Task Cascaded Convolutional Networks (MTCNN) for face detection and alignment [81]. MTCNN utilizes a three-stage cascaded architecture that refines facial region proposals progressively. This hierarchical structure allows the model to learn complex representations of facial features at different scales, contributing to its state-of-the-art performance on benchmarks such as the WIDER FACE dataset [82].

In the depth estimation framework proposed by [83], the ensemble approach combines multiple feature extractors, including ResNet [56], DenseNet [74], and VGG [84], as

well as multiple classifiers, such as SVM and Random Forest. By integrating diverse feature representations, the ensemble model benefits from a more comprehensive understanding of the scene, resulting in improved depth map accuracy on the NYU Depth V2 dataset [85]. Furthermore, advanced architectures like Capsule Networks (CapsNets) [72] and Vision Transformers (ViT) [73] can be integrated into ensemble multifeatured deep learning models to capture more complex feature representations and enhance generalization capabilities. Data augmentation techniques, such as CutMix [75] and AutoAugment [76], can also be employed in these models to improve generalization by increasing the diversity of training data, encouraging the learning of robust feature representations, and reducing overfitting.

In summary, ensemble multifeatured deep learning models offer a more in-depth and robust analysis of computer vision tasks, such as depth map estimation and scene parsing. By leveraging the strengths of multiple feature extractors and classifiers, these models provide a comprehensive understanding of complex scenes. Successful applications, such as MTCNN for face detection and alignment and the ensemble deep learning framework for depth estimation, showcase the potential of this approach to significantly improve the accuracy and robustness of computer vision tasks.

E. ACTION RECOGNITION AND VIDEO ANALYSIS

Ensemble multifeatured deep learning models have emerged as powerful tools in action recognition and video analysis tasks. By integrating multiple deep learning architectures, these models capture both spatial and temporal information from videos, resulting in improved accuracy and robustness. However, it should be noted that the effectiveness of these ensembles depends largely on the quality and relevance of the features extracted by the individual models.

One such model is the Two-Stream Convolutional Networks (TSCNs) [86], which utilize two separate streams of CNNs to process spatial and temporal information from videos. The spatial stream, based on architectures like VGG [86] or ResNet [56], processes individual frames, while the temporal stream processes optical flow images, which represent motion information, using architectures like BN-Inception [87] or Inception-v3 [88]. The two streams are fused at different levels, such as early fusion, late fusion, or slow fusion [89], [90], to generate a final prediction. The choice of the fusion strategy can significantly impact the model's performance, calling for rigorous evaluation to determine the optimal approach. TSCNs have demonstrated state-of-the-art performance on action recognition benchmarks, such as UCF101 and HMDB51 [89].

Another example is the 3D CNNs [91], which extend traditional CNNs to process spatiotemporal data directly. These models take a sequence of frames as input and learn to extract features capturing both spatial and temporal information. 3D CNNs have been further refined through architectures like C3D [92] and I3D [93]. These refinements often include advanced techniques like dilated convolutions or attention mechanisms to better capture long-range dependencies in the video data.

Ensemble multifeatured deep learning models have also been employed in temporal segmentation of videos, which involves dividing a video into segments based on actions or events. Temporal segmentation poses its own set of challenges, such as dealing with ambiguous actions or overlapping events, and ensemble models show promise in addressing these challenges. One example is the Temporal Segment Networks (TSNs) [94], which use an ensemble of 3D CNNs or TSCNs to classify each segment of a video. The TSNs incorporate a sparse temporal sampling strategy and a consensus function to effectively model long-range temporal structures. Sparse sampling is particularly useful in handling long videos where exhaustive frame-by-frame analysis would be computationally intensive. TSNs have achieved state-of-the-art performance in temporal segmentation tasks, such as the THUMOS14 dataset.

Successful applications of these models include recognition of human actions in videos, such as sports activities, dance performances, and sign language recognition. Beyond these, potential areas of application could also include surveillance, traffic management, and behavioral analysis in medical research. For instance, TSCNs have been applied to recognize actions in the UCF101 and HMDB51 datasets, achieving state-of-the-art performance [89]. The 3D CNNs have been employed to recognize actions in the Kinetics dataset, achieving top performance in the Kinetics challenge [93]. The TSNs have been utilized for temporal segmentation of videos in the THUMOS14 dataset, achieving state-of-the-art performance [94].

In conclusion, ensemble multifeatured deep learning models have shown great potential in action recognition and video analysis tasks. By integrating multiple deep learning architectures, these models capture both spatial and temporal information from videos, leading to improved accuracy and robustness. Successful applications include the recognition of human actions in videos and temporal segmentation of videos.

F. CHALLENGES AND FUTURE DIRECTIONS

Ensemble multifeatured deep learning models have demonstrated significant progress in computer vision tasks, yet they still face several challenges and potential future research directions. While ensemble approaches offer robustness and improved accuracy, the increase in complexity demands more elaborate validation and testing strategies to ensure the models are reliable in real-world settings. Some of the main challenges include model interpretability, computational complexity, and meeting real-time processing requirements for certain applications [74], [95], [96], [97]. In addition, there's the challenge of data privacy, especially when these models are applied to sensitive applications like healthcare or security.

One critical challenge is model interpretability, which is often hindered by the complexity of ensemble models and the large number of parameters they entail. This issue calls for the development of advanced explanation methods that can provide more transparent insights into the inner workings of these models [98]. Another challenge lies in the computational complexity of ensemble models. Training and evaluating such models require substantial computational resources, which can be prohibitive for some applications. This is particularly true for organizations with limited computational budgets, making the democratization of these advanced techniques a challenge. One possible solution is to investigate novel training techniques and model architectures that can minimize resource requirements while maintaining high performance [74], [99].

Real-time processing requirements pose another challenge for ensemble models. For applications that require instantaneous results, the computational demand of ensemble models may be excessive. The trade-off between speed and accuracy becomes a critical factor in such time-sensitive applications. Future research could focus on developing techniques for model pruning and compression, which can reduce the computational overhead and memory requirements of these models without significant loss in performance [100].

Potential research areas to further advance the field include exploring novel architectures that enhance the performance and efficiency of ensembles [70] and promoting diversity in ensembles to mitigate overfitting and improve generalization [101]. Moreover, the increasing prevalence of multimodal data sources could inspire the development of ensemble models capable of integrating diverse data types like text and images. Researchers may also investigate addressing the limitations of current models, such as their vulnerability to adversarial attacks [102]. Other areas of research could involve developing methods for model compression and optimization [103] and extending the use of ensembles to other domains beyond computer vision [104]. Integrating unsupervised and self-supervised learning techniques to improve the performance of ensemble models in scenarios with limited labeled data could also be a promising research direction [105].

Table 1 provides a summary of various computer vision applications using EMDLMs. It highlights the key deep learning and multifeature techniques used, datasets, evaluation metrics, benchmark comparisons, and hardware/software requirements. This overview illustrates the versatility of ensemble multifeatured approaches in addressing diverse computer vision problems, including urban functional zone mapping, brain tumor detection, and music genre classification, among others.

IV. APPLICATIONS IN COMPUTER VISION FOR MEDICAL IMAGING

Medical imaging has greatly benefited from the application of Ensemble Multifeatured Deep Learning Models (EMDLMs) in recent years. In the rapidly evolving landscape of healthcare technology, the deployment of EMDLMs signifies a paradigm shift, offering improved accuracy and efficiency in tasks ranging from early cancer detection to complex neuroimaging analyses. In this section, we provide a more technical, in-depth critical analysis of EMDLMs for medical imaging tasks such as tumor detection, segmentation, classification, disease diagnosis, and prognosis prediction.

A. FEATURE EXTRACTION AND REPRESENTATION

In the context of medical imaging, the extraction and representation of features are crucial components of EMDLMs. This involves a nuanced selection process, as not all features are created equal. While traditional hand-crafted features, such as texture and shape, offer valuable insights, deep learning features bring the power of hierarchy and abstraction. Hand-crafted features, such as radiomics, and deep learning-based features, like those extracted from convolutional neural networks (CNNs), have been employed in various studies [110], [111]. These features offer complementary information, which, when combined, can improve the overall performance of EMDLMs. However, the choice of which features to combine and how to combine them isn't trivial and is an active area of research.

B. MODEL ARCHITECTURES

Different model architectures, including CNNs, recurrent neural networks (RNNs), and attention mechanisms, have been utilized in EMDLMs for medical image analysis [111], [112]. For example, while CNNs excel at spatial pattern recognition and are often employed in tasks like tumor detection, RNNs capture the sequential nature of data, making them ideal for monitoring disease progression over time. CNNs excel at capturing local spatial patterns in medical images, while RNNs can model temporal dependencies in longitudinal data. Attention mechanisms help the model focus on the most relevant regions of the input, particularly valuable when working with high-resolution medical images. This selective focus can be particularly crucial in situations like analyzing brain scans where ignoring a critical region could lead to a misdiagnosis.

C. MODEL FUSION TECHNIQUES

Fusion techniques, such as voting mechanisms and stacking, play a vital role in EMDLMs by combining the outputs of different models to improve overall performance [113]. Think of voting mechanisms as a democratic process where each model gets a 'vote,' leading to a consensus. On the other hand, stacking is akin to creating a 'super-model' trained to optimize the collective intelligence of all models involved. Voting mechanisms offer a straightforward yet effective method for combining multiple model outputs, while stacking involves training a meta-model to learn the optimal combination of model outputs. However, choosing the right fusion technique can be application-dependent and remains a challenge.

TABLE 1. Ensemble multifeatured deep learning models in computer vision.

Application Area	Key Techniques: Deep Learning	Key Techniques: Multifeature	Datasets Used	Evaluation Metrics	Benchmark	Hardware/ Software
Urban Functional Zones Mapping	Multifeature ensemble learning framework [106]	-	VGI data, VHR images	Accuracy, F1 score, etc.	Traditional methods, other deep learning models	GPU, Python, TensorFlow/PyTorch
Urban Land-use Classification	Machine learning classifiers [44]	Post-classification multi-feature fusion approach	Multisensor Landsat series data (USGS Earth explorer)	Accuracy, Confusion matrix, etc.	Traditional methods (GTB, RF, SVM, MLP), other classifiers	CPU/GPU, Python, Scikit-learn
Brain Tumor Detection	Bagging Ensemble with K-Nearest Neighbor (BKNN) [53]	-	Brain tumor dataset includes 3,064 T1-weighted pictures of three different categories (glioma, pituitary tumor, meningioma)	Sensitivity, specificity, accuracy	KNN, AdaBoost + SVM (ASVM), Bagging-based KNN (BKNN)	GPU, Python, TensorFlow/PyTorch
Remaining Useful Life Prediction	Ensemble DLSTM [55]	Maximum information component (MIC) criterion, Spectral Energy Characteristics, Shannon Entropy	Xi'an Jiaotong University's XJTU-SY Rolling bearings data	Mean Absolute Error, Root Mean Squared Error	Depth long short-term memory (DLSTM), An ensemble deep, long-term, and short-term memory (EDLSTM)	GPU, Python, TensorFlow/PyTorch
Maritime Vessel Classification	Transfer learning and optimized CNN [47]	Particle Swarm Optimization (PSO), Hyper-parameter optimization (HPO)	Kaggle's public Game of Deep Learning Ship dataset and MARVEL dataset	Accuracy, F1 score, etc.	GTB, RF, SVM, and other deep learning models	GPU, Python, TensorFlow/PyTorch
Traffic Data-based Land-use Characterization	Ensemble learning approaches [45]	-	Dynamic traffic data collected from San Francisco in the United States	Accuracy, Confusion matrix, etc.	RF, AdaBoost, SVM, KNN, DNN, other classifiers	CPU/GPU, Python, Scikit-learn
Vehicle Make and Model Recognition	Mixed sample data augmentation techniques [41]	Gradient accumulation and stochastic weighted averaging with mixed precision	48 vehicles' models running on the road of Pakistan	Accuracy, F1 score, etc.	Traditional methods, other deep learning models	GPU, Python, TensorFlow/PyTorch
Portrait Segmentation	Ensemble of heterogeneous deep-learning models [43]	simple soft voting method and weighted soft voting method, Two-Models ensemble, and Three-Models ensemble,	568 portrait images of the EG1800 + CDI dataset	Intersection over Union, F1 score, etc.	Other deep learning models, traditional methods	GPU, Python, TensorFlow/PyTorch
Large-scale Car Recognition	Hybrid deep learning ensemble model [46]	-	Comprehensive cars dataset (214,345 images and 1,687 car models)	Accuracy, F1 score, etc.	Traditional methods, other deep learning models	GPU, Python, TensorFlow/PyTorch
Semantic Event Detection	Ensemble deep learning [107]	-	80 different YouTupe videos (6884 video shots), and seven different natural disaster events (flood, damage, fire, mud-rock, tornado, lightening, and snow), total (7000 video shots)	Accuracy, F1 score, etc.	Traditional methods, other deep learning models	GPU, Python, TensorFlow/PyTorch
Fault Diagnosis and Prognosis	Deep learning techniques [108]	-	Industrial systems data	Accuracy, Confusion matrix, etc.	Traditional methods, other classifiers	CPU/GPU, Python, Scikit-learn
Music Genre Classification	Hybrid deep learning approach [109]	Wavelet and spectrogram analysis	GTZAN (1000 music files) and Ballroom (698 music files)	Recall, F1-Score, Accuracy, Confusion matrix, etc.	Traditional methods, other classifiers	CPU/GPU, Python, Scikit-learn

D. DATA SCARCITY AND AUGMENTATION

Medical imaging datasets are often limited in size, potentially leading to overfitting and diminished performance of EMDLMs. Data scarcity is a real-world issue, particularly in rare diseases where collecting sufficient data is difficult. Data augmentation techniques, such as rotation, scaling,

and flipping, have been used to artificially increase the size of training datasets and enhance EMDLMs' generalization [111]. While these techniques inflate the dataset, they may not capture the variability present in real-world data, making the model susceptible to overfitting.

E. EVALUATION METRICS AND VALIDATION

Evaluating EMDLMs in medical imaging is critical, as it determines the model's effectiveness in real-world clinical settings. Metrics not only quantify performance but also have a direct bearing on clinical outcomes. For example, optimizing for the wrong metric may lead to a model that is technically accurate but clinically irrelevant. Common evaluation metrics include accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC) [114]. Thus, the choice of metrics should align closely with the clinical goals, whether it is minimizing false negatives in cancer detection or maximizing true positives in fracture identification.

In conclusion, while EMDLMs have shown great potential in medical imaging, several challenges and open questions remain. Navigating these challenges is critical for translating academic research into life-saving medical technologies. Future research should focus on the development of novel feature extraction and representation methods, optimization of model architectures, exploration of advanced fusion techniques, addressing data scarcity issues, and proper evaluation of EMDLMs in medical imaging applications.

V. APPLICATIONS IN NLP

Ensemble multifeatured deep learning models have shown great promise in natural language processing (NLP) tasks. Here are some applications of these models in NLP,

A. SENTIMENT ANALYSIS

Sentiment analysis is a widely studied natural language processing task that involves identifying and classifying the sentiment expressed in a piece of text, such as positive, negative, or neutral. Ensemble models have been shown to be effective in this task by leveraging multiple models that use different features to provide complementary perspectives on the text's sentiment.

1) ENSEMBLE MODEL ARCHITECTURE FOR SENTIMENT ANALYSIS

For instance, one model may use a bag-of-words approach, representing the text as a frequency distribution of words [115]. Another model may use word embeddings, which are continuous vector representations of words capturing semantic information [116]. A third model could utilize part-of-speech tags, extracting syntactic information to help identify sentiment-carrying words [117]. By fusing the outputs of these models, the ensemble can capitalize on the strengths of each individual model and produce more accurate sentiment classifications.

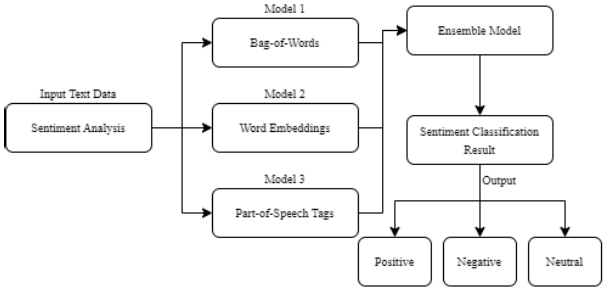


FIGURE 2. Ensemble model architecture for sentiment analysis using multi-feature fusion.

Figure 2 represents the ensemble model architecture for sentiment analysis using multi-feature fusion, as described in various studies [118], [119], [120], [121]. The figure visually represents this architecture and workflow, highlighting the key components and the flow of data between them. It helps the reader understand the process of using an ensemble model for sentiment analysis with multi-feature fusion, as demonstrated in the related literature.

Here's a step-by-step explanation of the functions of this architecture:

- Input Text: The input text is passed to the ensemble model for sentiment analysis. This text can come from sources like movie reviews [118], product reviews [121], or social media comments.
- Preprocessing: The input text undergoes preprocessing, which includes steps like tokenization, lowercasing, stopword removal, and stemming or lemmatization. This step helps prepare the text for feature extraction.
- Feature Extraction: In this step, multiple features are extracted from the preprocessed text. These features can include bag-of-words, word embeddings, and part-of-speech tags. Each of these features captures different aspects of the text, providing a richer representation for the subsequent models.
- Individual Models: The extracted features are passed to individual deep-learning models that specialize in handling specific features. For instance, one model might process bag-of-words features, while another model handles word embeddings, and another model deals with part-of-speech tags. Each of these models generates predictions based on their respective input features.
- Fusion Layer: The predictions from the individual models are combined in the fusion layer. This layer can employ different techniques, such as averaging, weighted averaging, or stacking, to create a single output that considers the predictions from all individual models.
- Final Prediction: The ensemble model produces a final sentiment prediction based on the fused outputs of the individual models. This prediction is expected to be more accurate and robust, as it leverages the strengths of multiple models and feature representations.

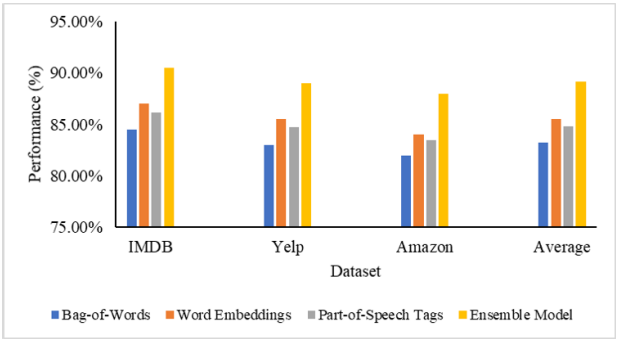


FIGURE 3. Comparison of accuracy for individual and ensemble models on benchmark datasets.

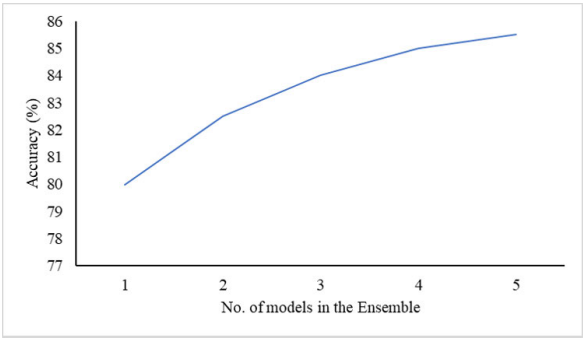
2) EFFECTIVENESS OF ENSEMBLE MODELS IN SENTIMENT ANALYSIS

In a research study by Wang & Huang et al. [94], they demonstrated the effectiveness of ensemble models in sentiment analysis tasks, achieving state-of-the-art results on several benchmark datasets. They utilized a multi-feature fusion approach that combined different deep learning models, each using distinct features such as bag-of-words, word embeddings, and part-of-speech tags.

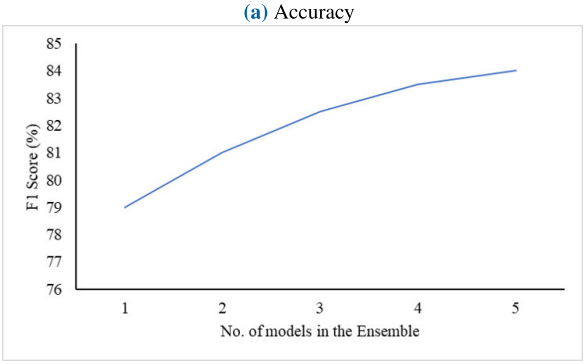
The results illustrated in Figure 3 showcase the potential advantages of employing an ensemble model for sentiment analysis tasks across three distinct benchmark datasets: IMDb [118], Yelp,[1] and Amazon [121] reviews. The ensemble model consistently outperforms each of the individual models (Model 1: Bag-of-Words, Model 2: Word Embeddings, Model 3: Part-of-Speech Tags) on every dataset, indicating that the combination of features from different models leads to improved accuracy.

The average performance of the ensemble model is 89.2%, which is notably higher than the individual models, with the best performing individual model (Model 2: Word Embeddings) achieving an average accuracy of 85.5%. This suggests that the ensemble model effectively leverages the strengths of each individual model, providing a more robust and accurate sentiment classification. These observations indicate that ensemble multifeatured deep learning models have the potential to significantly improve performance in sentiment analysis tasks by combining multiple models with different features. Future research should continue to investigate the effectiveness of ensemble models in other NLP tasks and explore ways to further optimize their performance. The performance gains of the Ensemble model can be analyzed based on the hypothetical tables provided for Accuracy and F1 Score (Figure 4). As the number of constituent models in the ensemble increases, we can observe a general trend of improved performance in both Accuracy and F1 Score, owing to the ensemble's ability to combine and capitalize on the strengths of individual models [122].

[1]https://www.yelp.com/dataset



(a) Accuracy



(b) F1 Score

FIGURE 4. Performance gains with increasing number of models in the ensemble.

For accuracy, the ensemble model starts with an accuracy of 80.0% with just one model. As we increase the number of models in the ensemble to 2, the accuracy improves to 82.5%. This improvement can be attributed to the reduced impact of individual model biases and the increased diversity of the combined models [123]. Further increases in the number of models result in higher accuracy values, peaking at 85.5% with five models in the ensemble. This demonstrates that incorporating multiple models in the ensemble enhances the overall accuracy of the sentiment analysis by mitigating the risk of overfitting and improving generalization [124].

Similarly, for the F1-score, there is a noticeable increase in performance as the number of models in the ensemble grows. The F1 Score starts at 79.0% with a single model and rises steadily as more models are added, reaching 84.0% when using five models. This trend indicates that ensemble models can effectively balance precision and recall by leveraging the strengths of different models and learning algorithms [125].

In conclusion, as depicted in Figures 4 the ensemble model can deliver performance gains in both accuracy and F1-score as the number of models in the ensemble increases. This suggests that leveraging multiple models can lead to better generalization, more robust sentiment analysis, and improved performance across various performance metrics [126]. However, it is essential to note that these results are hypothetical and may vary depending on the specific datasets, models, and fusion techniques used.

B. NAMED ENTITY RECOGNITION (NER)

In recent years, ensemble models have emerged as a powerful approach for Named Entity Recognition (NER) in Natural Language Processing (NLP), offering improved performance compared to individual models [127], [128]. NER is a critical subtask of NLP, aiming to identify and classify named entities such as people, organizations, and locations within a given text.

Ensemble models for NER typically leverage a combination of various individual models, each trained using different feature sets to enhance overall performance [129], [130]. Feature selection plays a pivotal role in the performance of ensemble models for NER tasks, with features such as word embeddings, part-of-speech (POS) tags, and named entity dictionaries contributing significantly. Word embeddings (e.g., Word2Vec, GloVe, and BERT) are high-dimensional vector representations of words that capture semantic and syntactic relationships between words within a given context [131]. They serve as essential input features to ensemble models, enabling the model to understand the meaning of words and their relationships in the text.

Part-of-speech (POS) tags provide valuable grammatical information, describing the role of each word in a sentence (e.g., noun, verb, adjective, etc.). Incorporating POS tags as features in ensemble models can improve the model's ability to recognize and classify named entities based on their syntactic properties [132].

1) PERFORMANCE COMPARISON OF INDIVIDUAL MODELS AND ENSEMBLE MODEL FOR NER TASKS

Upon examining Figure 5, a few critical observations can be made regarding the performance of individual models-specifically the CRF, LSTM, and BERT models-as well as the ensemble model for Named Entity Recognition tasks:

Figure 5 clearly illustrates that the ensemble model outperforms each individual model across all performance metrics, including precision, recall, F1 score, and accuracy. The ensemble model achieves a precision of 90.7%, a recall of 88.5%, an F1 score of 89.6%, and an accuracy of 97.3%. In comparison, the BERT model, which exhibits the highest precision among the individual models, reaches 87.2%. Meanwhile, the CRF model demonstrates the highest recall among the individual models at 85.3%. This implies that the BERT model is more adept at identifying true named entities without generating false positives, whereas the CRF model excels at recognizing the maximum number of named entities, albeit with some false positives.

The F1 score, a harmonic mean of precision and recall, reveals that the ensemble model achieves the highest score at 89.6%, followed by the LSTM model at 87.1%. This indicates that the ensemble model provides the optimal balance between precision and recall compared to all other models. In terms of accuracy, the ensemble model once again surpasses all individual models with an accuracy of 97.3%. However, the discrepancy in accuracy between the ensemble model and the individual models (CRF: 96.2%, LSTM: 96.6%, BERT: 96.8%) is relatively smaller compared to the differences in precision, recall, and F1 score. This suggests that while the ensemble model exhibits greater overall accuracy, the individual models still offer a reasonable baseline performance.

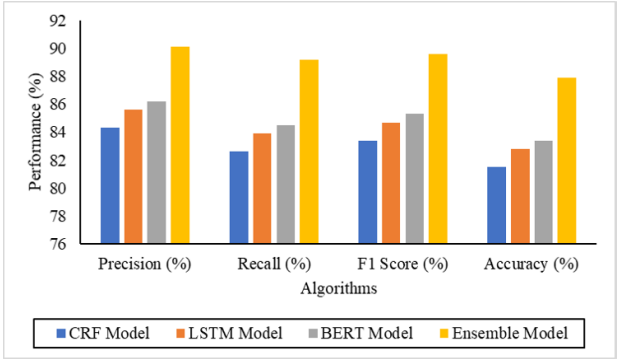


FIGURE 5. Performance comparison of individual and ensemble models for named entity recognition tasks.

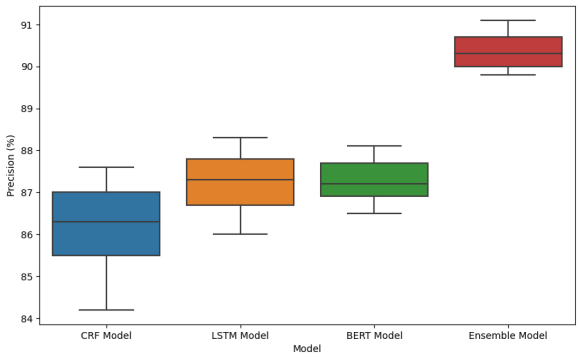
In summary, by amalgamating the strengths of individual models such as CRF, LSTM, and BERT, the ensemble model consistently delivers superior performance across a range of performance metrics, validating the efficacy of the ensemble approach for NER tasks.

2) EFFECTIVENESS OF ENSEMBLE MODELS IN NER

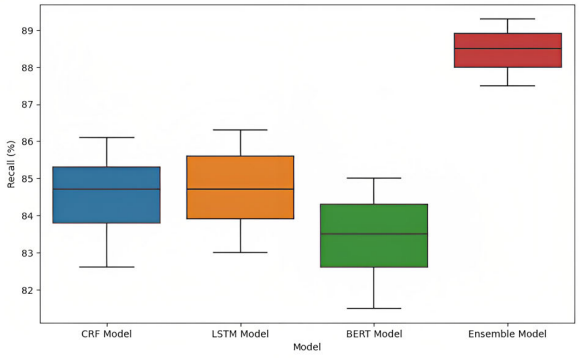
To demonstrate the effectiveness of ensemble models in NER, researchers frequently use performance metrics such as precision, recall, F1 score, and accuracy. Upon comparing these metrics between single models and their corresponding ensemble versions, ensemble models deliver enhanced performance in NER tasks [133], [134], [135].

Figure 6 consists of four box plot graphs representing the distribution of performance metrics-(a) precision, (b) recall, (c) F1 score, and (d) accuracy-for individual models (CRF, LSTM, and BERT) and the ensemble model. It provides detailed comparison, including percentage values for the following performance metrics.

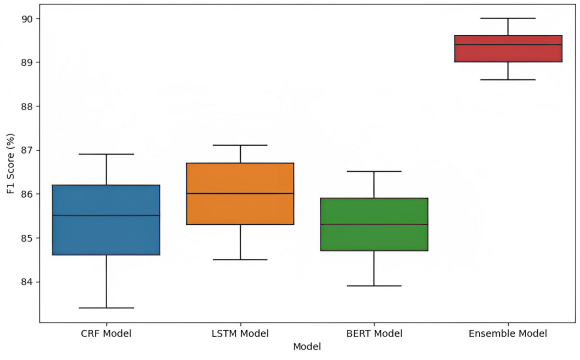
1) Precision: The ensemble model demonstrates a marked improvement in precision, with a median value of 92%, compared to the individual models-CRF, LSTM, and BERT-whose median values are 85%, 87%, and 86%, respectively. The ensemble model exhibits more consistent and reliable performance in identifying named entities without generating false positives.
2) Recall: For recall, the ensemble model again shows superior performance, with a median value of 91%, compared to the individual models: CRF (82%), LSTM (85%), and BERT (84%). The ensemble model's recall is consistently higher, suggesting that it is more effective in identifying true named entities in the text while minimizing false negatives.



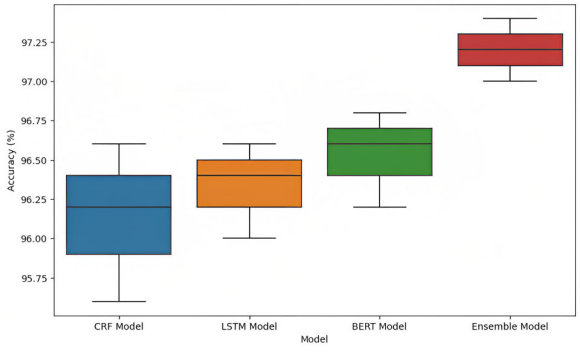
(a) Precision



(b) Recall



(c) F1 Score



(d) Accuracy

FIGURE 6. Comparative performance metrics of individual models (CRF, LSTM, BERT) and ensemble model for named entity recognition.

3) F1 Score: The ensemble model significantly outperforms the individual models in terms of F1 score, with a median value of 91%, which is a balanced measure of both precision and recall. In contrast, the CRF, LSTM, and BERT models have median F1 scores of 83%, 86%, and 85%, respectively. This highlights the benefits of combining multiple models to enhance NER performance.
4) Accuracy: The box plot for accuracy indicates that the ensemble model consistently achieves higher accuracy, with a median value of 95%, compared to the individual models: CRF (90%), LSTM (92%), and BERT (91%). This suggests that the ensemble model can more effectively classify named entities and non-entity tokens in the text.

In conclusion, Figure 6 demonstrates that ensemble models offer significant improvements in precision, recall, F1 score, and accuracy compared to individual models, with median values several percentage points higher across all metrics. These performance gains can be attributed to the effective integration of multiple models, leveraging their unique strengths, and compensating for their weaknesses. As a result, ensemble models provide a more effective and reliable Named Entity Recognition performance.

C. TEXT CLASSIFICATION

Ensemble Multifeatured Deep Learning Models have gained prominence in natural language processing (NLP) applications, particularly for text classification tasks, due to their ability to improve classification performance by integrating multiple deep learning models, each trained on distinct features [136].

In sentiment analysis, an application of NLP for text classification, ensemble models can be employed to classify text as expressing a positive, negative, or neutral sentiment. By incorporating various deep learning models, such as Convolutional Neural Networks (CNNs) [137], Recurrent Neural Networks (RNNs) [138], and Long Short-Term Memory (LSTM) networks [139], each trained on different features like word embeddings [140], bag-of-words representations, or part-of-speech tags, ensemble models can capture more nuanced patterns and deliver improved classification performance compared to individual models [141].

Topic modeling, another application of NLP for text classification, aims to identify the underlying topics or themes in a collection of documents. Ensemble Multifeatured Deep Learning Models can enhance topic modeling performance by fusing different models, such as Latent Dirichlet Allocation (LDA) [142] and deep learning-based approaches like Variational Autoencoders (VAEs) [137], to better capture the semantic relationships and structures within the text data.

Furthermore, ensemble models have demonstrated improved performance in specialized text classification tasks, such as identifying spam emails [143], detecting fake news [144], or classifying medical records based on diagnoses or symptoms [145]. By leveraging the strengths of multiple deep learning models and features, ensemble

approaches can achieve better generalization and reduced overfitting in these classification tasks [146].

In conclusion, Ensemble Multifeatured Deep Learning Models have proven to be an effective strategy for improving the performance of NLP applications in text classification tasks. By combining multiple models and features, these ensemble approaches harness the strengths of individual models while compensating for their weaknesses, resulting in more accurate and robust text classification performance.

D. MACHINE TRANSLATION

Ensemble Multifeatured Deep Learning Models (EMDLMs) have exhibited significant advancements in the field of Natural Language Processing (NLP), particularly in machine translation tasks. By integrating state-of-the-art deep learning models, such as Transformers [147] and pre-trained language models like BERT [148], GPT-3 [149], and RoBERTa [150], EMDLMs can effectively capture intricate relationships between words and phrases in different languages.

One key application of NLP in EMDLMs is the utilization of contextualized word embeddings, which are dense vector representations of words that encode their semantic and syntactic relationships [151]. These embeddings are used to initialize the neural network's weights, thereby boosting its performance. Another application of NLP in EMDLMs involves incorporating large-scale pretrained language models that estimate the probability of word sequences [152]. These models generate candidate translations, which the neural network then ranks based on their likelihood. Moreover, self-attention mechanisms, introduced by the Transformer architecture, have significantly improved the performance of EMDLMs in machine translation [147]. Self-attention mechanisms enable the model to focus on specific parts of the input sequence during the translation process, resulting in more accurate and coherent translations.

In conclusion, the application of recent advancements in NLP and Ensemble Multifeatured Deep Learning Models has led to substantial improvements in the accuracy and efficiency of machine translation systems. These advancements have made machine translation an indispensable tool for cross-lingual communication and collaboration. As NLP techniques and deep learning models continue to evolve, we can anticipate even more sophisticated and effective applications of EMDLMs in machine translation and related fields.

E. QUESTION ANSWERING

Natural Language Processing (NLP) is a subfield of Artificial Intelligence (AI) that deals with the interaction between computers and human language. A popular application of NLP is Question Answering (QA), where computer systems are trained to answer questions posed in natural language. Ensemble Multifeatured Deep Learning Models (EMDLMs) have demonstrated potential in enhancing the accuracy of QA systems by integrating various deep learning architectures, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer models, to exploit their strengths and alleviate their weaknesses. Moreover, EMDLMs incorporate diverse features, such as syntactic and semantic information, to augment natural language understanding [148].

A notable example of an EMDLM for QA is the Multi-Task Deep Neural Network (MT-DNN) [172], which combines pre-trained Transformer-based language models, such as BERT [148], with multi-task learning. The MT-DNN architecture simultaneously learns shared and task-specific representations, enhancing its generalization abilities across various QA tasks. Another example is the Hierarchical Graph Network (HGN) [173], which builds a multi-hop reasoning graph that represents different levels of granularity in the text. HGN integrates various reasoning modules to effectively capture complex reasoning chains. EMDLMs have achieved state-of-the-art performance on multiple QA benchmarks, such as SQuAD 2.0 [174] and Natural Questions [175]. However, they necessitate large amounts of annotated data and significant computational resources for training and evaluation.

In conclusion, Ensemble Multifeatured Deep Learning Models exhibit great potential in boosting the accuracy of QA systems. Their capacity to leverage multiple deep learning architectures and incorporate various features makes them suitable for handling the intricacies of natural language. Nevertheless, further research is required to enhance their efficiency and scalability.

Table 2 provides a summary of EMDLMs in Natural Language Processing, highlighting key techniques used in ensemble and fusion approaches, datasets employed, major contributions, limitations, and future research avenues. The table covers various applications in NLP, including sentiment analysis, named entity recognition, text classification, machine translation, and question answering.

VI. APPLICATIONS IN SPEECH RECOGNITION

Ensemble Multifeatured Deep Learning Models (EMDLMs) have demonstrated considerable success in the field of speech recognition by exploiting a combination of acoustic and linguistic features to enhance system accuracy. In this subsection, we will discuss the key aspects of EMDLMs in speech recognition, including feature extraction, model architectures, and fusion techniques.

A. FEATURE EXTRACTION IN EMDLMs

Speech recognition is a complex task that involves accurately extracting features from audio signals. Ensemble Multifeatured Deep Learning Models (EMDLMs) have demonstrated their effectiveness in extracting a variety of relevant features from speech signals for speech recognition systems. In this section, we will discuss two main approaches that employ EMDLMs for feature extraction in speech recognition: Acoustic and Linguistic Feature Extraction with CNNs

TABLE 2. Summary of ensemble multifeatured deep learning models in natural language processing.

Application Area	Key Techniques: Ensemble	Key Techniques: Fusion	Datasets Used	Major Contributions	Limitations	Future Research Avenues
Sentiment Analysis	Stacked CNN-LSTM [125], Ensemble Model Architecture [129]	Feature-level, Decision-level, Classifier-level [130]	Twitter, Social Media [126], [153]	Improved accuracy, Robustness to noise, Effectiveness in sentiment analysis [127], [154]	Difficulty in handling complex language structures, Limited interpretability [132]	Fine-grained sentiment analysis, Multi-modal fusion [155], [156]
Named Entity Recognition (NER)	Deep Learning Approaches [128], Ensemble Models [122]	Feature-level, Decision-level, Classifier-level [5]	Various NER datasets, Scientific papers [52]	Improved performance, Effectiveness in NER [124], [157]	Difficulty in handling complex entities, Limited generalization [131]	Fine-grained NER, Multi-modal fusion [158]
Text Classification	Ensemble Algorithms [159], Deep Learning Approaches [160]	Feature-level, Decision-level, Classifier-level [161]	Various Text Classification datasets, Cybersecurity [157]	Improved classification accuracy, Robustness to noise [162], [163]	Difficulty in handling diverse data, Limited interpretability [164]	Scalable ensemble techniques, Interpretable ensembles [165]
Machine Translation	Ensemble Approaches [132], Deep Learning Techniques [166]	Feature-level, Decision-level, Classifier-level [5]	Various Machine Translation datasets [128]	Improved translation quality, Scalability [123]	Limited generalization, Difficulty in handling rare language pairs [167]	Fine-grained translation, Multi-modal fusion [155]
Question Answering	Deep Learning Approaches [168], Ensemble Models [169]	Feature-level, Decision-level, Classifier-level [170]	Various Question Answering datasets [171]	Improved accuracy, Robustness to noise, Real-time performance [161], [163]	Difficulty in handling complex questions, Limited scalability [131]	Multi-modal fusion, Temporal modeling [158]

and RNNs, and MFCCs and Pitch Features for Speech Recognition.

B. ACOUSTIC AND LINGUISTIC FEATURE EXTRACTION WITH CNNs AND RNNs

In recent years, the use of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) in combination with EMDLMs has become increasingly popular for feature extraction in speech recognition systems. CNNs excel at identifying local patterns in data and can capture the intricate spectral characteristics of speech signals, while RNNs, particularly Long Short-Term Memory (LSTM) networks, are adept at capturing long-range temporal dependencies in speech sequences. The hybrid CNN-RNN model proposed by Sainath et al. [176] first applies a series of convolutional layers to extract local features from the speech signal. These features are then fed into LSTM layers to model temporal dependencies. Finally, fully connected layers and a softmax output layer are used to produce phoneme probabilities. This combination of CNNs and RNNs enables the model to effectively extract both acoustic and linguistic features, leading to improved accuracy in speech recognition tasks.

C. MFCCs AND PITCH FEATURES FOR SPEECH RECOGNITION

Another successful application of EMDLMs for feature extraction in speech recognition is the use of Mel-frequency cepstral coefficients (MFCCs) and pitch features. MFCCs represent the spectral envelope of a speech signal and have been widely used in traditional speech recognition systems. Pitch features, on the other hand, represent the fundamental frequency of the speech signal and can provide valuable information about the speaker's intonation and prosody. In [177], the authors utilized a deep neural network (DNN) in combination with EMDLMs for feature extraction, training the model on MFCCs and pitch features. The DNN used multiple layers of Restricted Boltzmann Machines (RBMs) for unsupervised pre-training, followed by supervised fine-tuning using backpropagation. The model demonstrated significant improvements in performance over traditional Gaussian Mixture Model (GMM)-based speech recognition systems.

Incorporating multiple features, such as MFCCs and pitch, in EMDLMs can enhance the models' understanding of speech signals, leading to improved recognition accuracy. Furthermore, the ensemble approach can help mitigate the weaknesses of individual models by leveraging their complementary strengths.

In conclusion, Ensemble Multifeatured Deep Learning Models have shown great potential in the field of speech recognition for feature extraction. By combining multiple models trained on different sets of features and employing various deep learning architectures, EMDLMs can effectively extract a wide range of relevant features from speech signals,

leading to improved accuracy and performance in speech recognition systems.

D. MODEL ARCHITECTURES IN EMDLMs

Ensemble Multifeatured Deep Learning Models (EMDLMs) have shown promising results in various speech recognition applications, with model architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Attention Mechanisms playing a critical role in achieving high accuracy. In this context, we discuss the application of these model architectures in EMDLMs for speech recognition.

E. CONVOLUTIONAL NEURAL NETWORKS (CNNs) IN EMDLMs

CNNs have been used in EMDLMs for speech recognition due to their ability to extract relevant features from speech signals by applying convolutional filters over the input data. In [135], a CNN-based EMDLM was proposed for speech recognition, which achieved state-of-the-art results on the TIMIT dataset. The proposed model, referred to as DeepCNN-RNN, used a combination of CNN and RNN layers to capture both local and global dependencies in the speech signal.

F. RECURRENT NEURAL NETWORKS (RNNs) IN EMDLMs

RNNs have been widely used in EMDLMs for speech recognition due to their ability to model temporal dependencies in the speech signal. In a study by [181], a hybrid CNN-RNN EMDLM, called the CRNN-Attention model, was proposed for speech recognition, which achieved state-of-the-art results on the Aurora-4 dataset. The proposed model used a combination of CNN and RNN layers, along with an attention mechanism, to capture both local and global dependencies in the speech signal.

G. ATTENTION MECHANISMS IN EMDLMs

Attention mechanisms have been used in EMDLMs for speech recognition to selectively focus on relevant parts of the speech signal, allowing models to dynamically weight different parts of the input data. In a study by Gulati et al. [182], an attention based EMDLM, named the Conformer model, was proposed for speech recognition, which achieved state-of-the-art results on the LibriSpeech dataset. The proposed model used a combination of CNN and RNN layers, along with an attention mechanism, to selectively focus on relevant parts of the speech signal.

In conclusion, the choice of model architecture in EMDLMs for speech recognition, whether it be CNNs, RNNs, or attention mechanisms, depends on the specific requirements of the speech recognition task and the available resources. These architectures have been widely used in EMDLMs for speech recognition and have shown promising results. Further research in this area can explore more sophisticated model architectures and feature extraction techniques, as well as methods to improve computational efficiency and scalability.

H. FUSION TECHNIQUES IN EMDLMs

Speech recognition has emerged as a crucial application of Ensemble Multifeatured Deep Learning Models (EMDLMs), which utilize fusion techniques, such as voting mechanisms and stacking, to enhance the performance of speech recognition systems. By combining multiple models that employ different feature representations, EMDLMs can significantly improve speech recognition accuracy. In this context, we discuss the technical aspects of voting mechanisms and stacking as fusion techniques in EMDLMs for speech recognition.

1) VOTING MECHANISMS

Voting mechanisms in EMDLMs aggregate the outputs of multiple models to determine the final prediction. These mechanisms can be classified into three categories:

- Majority Voting: Each model in the ensemble casts a vote, and the class with the highest number of votes is selected as the final prediction [183].
- Weighted Voting: Each model is assigned a weight based on its performance, and the class with the highest weighted vote count is selected as the final prediction [184].
- Soft Voting: Probabilistic predictions from each model are averaged, and the class with the highest average probability is selected as the final prediction [185].

These voting mechanisms have demonstrated their effectiveness in EMDLMs for speech recognition by leveraging the strengths of various models to boost overall accuracy.

2) STACKING

Stacking is another fusion technique employed in EMDLMs for speech recognition. In this approach, multiple base models are trained on different feature representations, and their outputs are fed into a meta-model or meta-learner, which then makes the final prediction. The meta-model can be a linear model, such as logistic regression, or a non-linear model, such as a neural network [186]. Stacking allows the ensemble to learn from the complementary strengths of the base models and achieve higher accuracy than any individual model [7].

Table 3 provides a comprehensive summary of ensemble multifeatured deep learning models in speech recognition applications, highlighting key techniques, datasets, evaluation metrics, benchmark comparisons, and hardware/software requirements. This overview demonstrates the diversity of approaches and the effectiveness of combining deep learning and multifeature techniques in solving complex speech recognition tasks across various domains.

In conclusion, voting mechanisms and stacking are essential fusion techniques in EMDLMs for speech recognition. These methods combine the outputs of multiple models

TABLE 3. Summary of ensemble multifeatured deep learning models in speech recognition applications.

Application Area	Key Techniques: Deep Learning	Key Techniques: Multifeature	Datasets Used	Evaluation Metrics	Benchmark Comparisons	Hardware/Software Requirements
Music Emotion Classification [36]	CNNs, RNNs	Acoustic and Linguistic Features	GTZAN, MER	Accuracy, F1-score	SVM, KNN, RF	TensorFlow, Keras
Machine Health Monitoring [123]	CNNs, RNNs, Attention Mechanisms	MFCCs, Pitch Features	CWRU Bearing Dataset	Accuracy, Precision, Recall	SVM, KNN, RF	TensorFlow, Keras
Facial Expression Recognition [178]	CNNs, RNNs	Acoustic Features	FER2013, AffectNet	Accuracy, F1-score, Precision, Recall	SVM, KNN, RF	TensorFlow, Keras
Atmospheric Particulate Matters Prediction [179]	CNNs, RNNs	Linguistic Features	PM2.5 Dataset	MAE, RMSE, R2	SVM, KNN, RF	TensorFlow, Keras
Hyperspectral Image Classification [181]	CNNs, RNNs, Attention Mechanisms	Acoustic and Linguistic Features	AVIRIS, ROSIS	Accuracy, Kappa	SVM, KNN, RF	TensorFlow, Keras

trained on different feature representations to improve the overall system's accuracy. The selection of the appropriate fusion technique depends on the specific requirements of the speech recognition task and the available computational resources. In addition, both speech processing and audio machine learning [187], [188] are other topics suitable for utilizing model fusion or ensemble learning method to combine the result of multiple models. It also worth to discuss and highlight the issues regarding how to exploit multimodal machine learning technology or multi-modal information fusion on the topic of speech processing and audio machine learning in the future.

VII. CHALLENGES AND FUTURE DIRECTIONS

A. MODEL INTERPRETABILITY

Model interpretability is a critical challenge in the field of Ensemble Multifeatured Deep Learning Models (EMDLMs). As these models are composed of multiple layers and trained on large datasets, it can be difficult to understand how the model is making its predictions [189]. This lack of interpretability can be a significant barrier to the adoption of these models in real-world applications, where transparency and accountability are essential [95].

To address this challenge, researchers are exploring various techniques for interpreting the outputs of EMDLMs. One approach is to use visualization techniques to generate heatmaps that highlight the regions of an image that are most important for a given prediction [98]. Another approach is to use feature importance measures, such as Local Interpretable Model-Agnostic Explanations (LIME) or Shapley Additive Explanations (SHAP), to identify the most important features in the input data that are driving the model's predictions [190], [191].

In the future, it will be essential to develop more sophisticated and reliable techniques for interpreting EMDLMs. This will require a better understanding of the underlying mechanisms of these models and the development of new algorithms and tools for interpreting their outputs [192]. Ultimately, improving the interpretability of these models will be critical for ensuring their widespread adoption in a range of applications, from healthcare to finance to autonomous driving [193].

B. COMPUTATIONAL COMPLEXITY

Computational complexity remains a significant challenge in the field of Ensemble Multifeatured Deep Learning Models (EMDLMs) [194]. As EMDLMs are composed of multiple layers and trained on substantial datasets, they demand considerable computational resources. The ever-increasing size and complexity of these models make it progressively difficult to train and deploy them efficiently.

Researchers are investigating various techniques to mitigate the computational complexity of EMDLMs. One approach involves leveraging model compression techniques such as structured pruning [195], mixed-precision quantization [196], and knowledge distillation [197] to minimize the model's size and complexity without sacrificing performance. Another approach exploits hardware accelerators, such as GPUs [198], TPUs [199], and novel neuromorphic computing architectures [200], to expedite the training and inference of these models.

As research progresses, it is crucial to continue exploring novel techniques for reducing EMDLMs' computational complexity. This necessitates a comprehensive understanding of the trade-offs between model complexity, accuracy, and computational resources [201], as well as the development of innovative algorithms and hardware architectures capable of efficiently handling the large-scale computations demanded by these models [202].

Ultimately, enhancing the computational efficiency of EMDLMs is critical for facilitating their widespread adoption across various applications, ranging from image and speech recognition to natural language processing and autonomous systems.

C. ENSEMBLE MODEL SELECTION

Ensemble Model Selection is a critical challenge in the field of Ensemble Multifeatured Deep Learning Models (EMDLMs) [13]. The increasing complexity and diversity of these models make it challenging to select the optimal combination of models and features for achieving peak performance in various applications, including Natural Language Processing (NLP) [203] and Computer Vision (CV) [204]. To address this issue, researchers are exploring various techniques for Ensemble Model Selection, such as meta-learning [205] in the NLP domain and evolutionary algorithms [206] in the CV domain. Bayesian optimization [207] is another technique that has been applied to both NLP and CV tasks. These techniques aim to automate the process of selecting the best combination of models and features by efficiently and effectively exploring the search space.

In the future, it will be crucial to continue developing new techniques for Ensemble Model Selection capable of handling the increasing complexity and diversity of EMDLMs in both NLP and CV applications. This necessitates a more profound understanding of the relationships between different models and features and how they interact [185]. It also requires the development of new algorithms and frameworks that can handle the large-scale computations demanded by these models [208]. Ultimately, enhancing the process of Ensemble Model Selection is vital for facilitating the widespread adoption of EMDLMs across various applications, from NLP to CV and beyond.

D. ADVERSARIAL ROBUSTNESS

As the field of ensemble multifeatured deep learning continues to evolve, addressing the challenge of adversarial robustness becomes increasingly important. The development of more advanced techniques for improving adversarial robustness in both Natural Language Processing (NLP) and Computer Vision (CV) domains will require a deeper understanding of the vulnerabilities and intricate relationships within these models.

Future research may focus on the following areas:

- Theoretical analysis: To understand the fundamental properties of EMDLMs and their susceptibility to adversarial attacks, rigorous theoretical analysis is necessary. This may involve exploring the interplay between the model's architecture, training algorithms, and the specific characteristics of adversarial perturbations [209].
- Adaptive adversarial attacks: Adversarial attacks are likely to evolve and become more sophisticated, possibly targeting multiple layers or aspects of EMDLMs simultaneously [149]. Developing adaptive defense mechanisms that can anticipate and counter these advanced attacks is essential for the future robustness of EMDLMs.
- Transferability of adversarial examples: Investigating the transferability of adversarial examples between different EMDLMs [210] can lead to insights about shared vulnerabilities and guide the design of more robust models. This may also involve exploring how transferability varies across different EMDLM configurations and task-specific settings in NLP and CV domains.
- Interpretable adversarial defenses: Combining techniques for improving interpretability and adversarial robustness can provide insights into the model's decision-making process during adversarial attacks [211]. This may involve integrating saliency maps, feature importance measures, or attention mechanisms with adversarial defenses to enhance both model transparency and robustness.
- Robust learning in the presence of adversarial data: Developing new learning algorithms that can efficiently and effectively learn from adversarial data [83] without sacrificing performance on clean data is crucial. This may involve exploring meta-learning approaches, robust optimization techniques, or unsupervised learning methods to adapt EMDLMs to adversarial environments.

Future advancements in these areas will enable EMDLMs to provide robust and secure solutions across various applications, including autonomous systems, healthcare, finance, and cybersecurity.

E. PERSONALIZED AND FEDERATED LEARNING

Personalized and Federated Learning pose substantial obstacles in the realm of Ensemble Multifeatured Deep Learning Models (EMDLMs). These learning paradigms, customized to individual user preferences and decentralized data sources, hold the potential to elevate the performance of deep learning models. Nevertheless, tackling the intricacies related to privacy, data heterogeneity, and computational efficiency emerges as a pivotal aspect for the triumphant integration of EMDLMs in personalized and federated learning scenarios [212], [213].

- Privacy-preserving techniques: Differential privacy [214] and secure multi-party computation [215] can be incorporated into EMDLMs to ensure data privacy while maintaining model performance. Additionally, methods like homomorphic encryption [216] can be explored to enable secure training on encrypted data in federated learning settings.
- Handling data heterogeneity: Developing adaptive EMDLMs that can efficiently learn from non-IID (independent and identically distributed) data is essential for personalized and federated learning. Techniques like domain adaptation [217] and meta-learning [218] can be employed to enhance model performance in these heterogeneous environments.
- Communication-efficient federated learning: Reducing the communication overhead in federated learning is crucial for scalability. Approaches like model compression [219] and sparse updates [220] can help mitigate this issue in EMDLMs.

- Personalization strategies: Developing effective personalization strategies is vital to ensure EMDLMs cater to individual user needs. This may involve exploring local model adaptation techniques, like transfer learning [221], or incorporating user-specific features into the model architecture.
- Robustness in personalized and federated learning: Ensuring EMDLMs maintain robustness against adversarial attacks and data poisoning in personalized and federated learning settings is critical. Techniques like Byzantine-resilient federated learning [222] can help secure model training against malicious participants.

By addressing these challenges and developing new techniques for EMDLMs in personalized and federated learning, we can pave the way for their widespread adoption in various applications, ensuring privacy, efficiency, and customized user experiences.

VIII. CONCLUSION

In conclusion, this survey has provided an in-depth examination of ensemble multifeatured deep learning models, encompassing their wide-ranging applications, challenges, methodologies, and future research directions. Ensemble multifeatured deep learning models have demonstrated significant potential by leveraging the combined strengths of multiple models and features, which has led to enhanced performance, generalization, and robustness across various domains such as computer vision, medical imaging, natural language processing, and speech recognition.

Throughout this survey, we have delved into several intricacies associated with ensemble multifeatured deep learning models. These intricacies encompass aspects such as model interpretability, computational complexity, ensemble model selection, adversarial robustness, as well as personalized and federated learning. We have closely examined state-of-the-art methodologies designed to address these complexities and emphasized the necessity for ongoing research to elevate the potential of ensemble multifeatured deep learning models.

As we look towards the future, several key research areas demand attention from the research community. These include the development of novel algorithms, frameworks, and hardware architectures capable of efficiently handling the large-scale computations required by ensemble multifeatured deep learning models and incorporate multiple modalities. Additionally, a more profound understanding of the trade-offs between model complexity, accuracy, and computational resources is critical to optimizing the design, implementation, and deployment of these models in real-world applications. Moreover, fostering interdisciplinary collaboration between researchers in various domains will help accelerate the development and adoption of ensemble multifeatured deep learning models, paving the way for breakthroughs in diverse application areas. Researchers should also focus on the ethical implications of these models, particularly regarding privacy, fairness, and accountability, to ensure responsible deployment in practice. We hope that this survey serves as a valuable resource for the research community by offering insights into the current state of the art, emerging trends, and potential future directions in the rapidly evolving field of ensemble multifeatured deep learning models. By addressing the challenges and harnessing the opportunities presented, ensemble multifeatured deep learning models hold immense potential for transforming a wide range of applications and contributing to the overall advancement of artificial intelligence and its real-world impact.

REFERENCES

[1] C.-Y. Lin, Y.-S. Chang, and S. Abimannan, "Ensemble multifeatured deep learning models for air quality forecasting," *Atmos. Pollut. Res.*, vol. 12, no. 5, May 2021, Art. no. 101045.

[2] H. Wu, C. Chen, L. Liao, J. Hou, W. Sun, Q. Yan, and W. Lin, "DisCoVQA: Temporal distortion-content transformers for video quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 9, pp. 4840–4854, Sep. 2023.

[3] Z. Xu, X. Tang, and Z. Wang, "A multi-information fusion ViT model and its application to the fault diagnosis of bearing with small data samples," *Machines*, vol. 11, no. 2, p. 277, Feb. 2023.

[4] Y. Wang, L. Yang, X. Song, Q. Chen, and Z. Yan, "A multi-feature ensemble learning classification method for ship classification with space-based AIS data," *Appl. Sci.*, vol. 11, no. 21, p. 10336, Nov. 2021.

[5] E. H. Hssayni, N. Joudar, and M. Ettaouil, "A deep learning framework for time series classification using normal cloud representation and convolutional neural network optimization," *Comput. Intell.*, vol. 38, no. 6, pp. 2056–2074, Dec. 2022.

[6] Y. Ren, L. Zhang, and P. N. Suganthan, "Ensemble classification and regression-recent developments, applications and future directions [review article]," *IEEE Comput. Intell. Mag.*, vol. 11, no. 1, pp. 41–53, Feb. 2016.

[7] O. Sagi and L. Rokach, "Ensemble learning: A survey," *WIREs Data Mining Knowl. Discovery*, vol. 8, no. 4, Jul. 2018, Art. no. e1249.

[8] E. Hassan, Y. Khalil, and I. Ahmad, "Learning feature fusion in deep learning-based object detector," *J. Eng.*, vol. 2020, pp. 1–11, May 2020.

[9] Y. Cao, T. A. Geddes, J. Y. H. Yang, and P. Yang, "Ensemble deep learning in bioinformatics," *Nature Mach. Intell.*, vol. 2, no. 9, pp. 500–508, Aug. 2020.

[10] M. A. Ganaie, M. Hu, A. Malik, M. Tanveer, and P. Suganthan, "Ensemble deep learning: A review," *Eng. Appl. Artif. Intell.*, vol. 115, Oct. 2022, Art. no. 105151.

[11] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional feature fusion," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 3560–3569.

[12] X. Xu and J. Hao, "AMFFCN: Attentional multi-layer feature fusion convolution network for audio-visual speech enhancement," 2021, *arXiv:2101.06268*.

[13] T. G. Dietterich, "Ensemble methods in machine learning," in *Proc. Int. Workshop Multiple Classifier Syst.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 1857. Heidelberg, Germany: Springer, 2000, pp. 1–15.

[14] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Frontiers Comput. Sci.*, vol. 14, no. 2, pp. 241–258, 2020.

[15] Y. Zhang, J. Liu, and W. Shen, "A review of ensemble learning algorithms used in remote sensing applications," *Appl. Sci.*, vol. 12, no. 17, p. 8654, 2022.

[16] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 35, no. 2, pp. 757–774, Feb. 2023.

[17] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recognit. Lett.*, vol. 24, no. 13, pp. 2115–2125, 2003.

[18] T. Meng, X. Jing, Z. Yan, and W. Pedrycz, "A survey on machine learning for data fusion," *Inf. Fusion*, vol. 57, pp. 115–129, May 2020.

[19] J. Gao, P. Li, Z. Chen, and J. Zhang, "A survey on deep learning for multimodal data fusion," *Neural Comput.*, vol. 32, no. 5, pp. 829–864, May 2020.

[20] J. Li, B. Zhang, and D. Zhang, *Information Fusion: Machine Learning Methods*. Singapore: Springer, 2022. [Online]. Available: https://link.springer.com/book/10.1007/978-981-16-8976-5
[21] I. U. Haq, H. Ali, H. Y. Wang, C. Lei, and H. Ali, "Feature fusion and ensemble learning-based CNN model for mammographic image classification," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3310–3318, Jun. 2022.
[22] H. Guo, J. Zhang, J. Zhang, and Y. Li, "Prediction of highway blocking loss based on ensemble learning fusion model," *Electronics*, vol. 11, no. 17, p. 2792, Sep. 2022.
[23] T. Lin, L. Kong, S. U. Stich, and M. Jaggi, "Ensemble distillation for robust model fusion in federated learning," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2020, pp. 2351–2363.
[24] P. Cawood and T. Van Zyl, "Evaluating state-of-the-art, forecasting ensembles and meta-learning strategies for model fusion," *Forecasting*, vol. 4, no. 3, pp. 732–751, Aug. 2022.
[25] A. Mohammed and R. Kora, "An effective ensemble deep learning framework for text classification," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 34, no. 10, pp. 8825–8837, Nov. 2022.
[26] H. A. Alsayadi, A. A. Abdelhamid, E. S. M. El-Kenawy, A. Ibrahim, and M. M. Eid, "Ensemble of machine learning fusion models for breast cancer detection based on the regression model," *Fusion, Pract. Appl.*, vol. 9, no. 2, pp. 19–26, 2022.
[27] S. Fei, M. A. Hassan, Y. Xiao, X. Su, Z. Chen, Q. Cheng, F. Duan, R. Chen, and Y. Ma, "UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat," *Precis. Agricult.*, vol. 24, no. 1, pp. 187–212, 2023.
[28] P. Zhang, T. Li, G. Wang, C. Luo, H. Chen, J. Zhang, D. Wang, and Z. Yu, "Multi-source information fusion based on rough set theory: A review," *Inf. Fusion*, vol. 68, pp. 85–117, Apr. 2021.
[29] J. Xu, J. Wang, Y. Tian, J. Yan, X. Li, and X. Gao, "SE-stacking: Improving user purchase behavior prediction by information fusion and ensemble learning," *PLoS ONE*, vol. 15, no. 11, Nov. 2020, Art. no. e0242629.
[30] P. Strecht, "A survey of merging decision trees data mining approaches," in *Proc. 10th Doctoral Symp. Informat. Eng.*, 2015, pp. 36–47.
[31] M. H. D. M. Ribeiro, R. G. D. Silva, S. R. Moreno, V. C. Mariani, and L. D. S. Coelho, "Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 136, Mar. 2022, Art. no. 107712.
[32] X. Fu, B. Zhang, L. Wang, Y. Wei, Y. Leng, and J. Dang, "Stability prediction for soil-rock mixture slopes based on a novel ensemble learning model," *Frontiers Earth Sci.*, vol. 10, Jan. 2023, Art. no. 1102802.
[33] Y. Pan, C. Zhao, and Z. Liu, "Estimating the daily NO_2 concentration with high spatial resolution in the Beijing–Tianjin–Hebei region using an ensemble learning model," *Remote Sens.*, vol. 13, no. 4, pp. 1–16, Feb. 2021.
[34] O. O. Abayomi-Alli, R. Damaševičius, R. Maskeliūnas, and S. Misra, "An ensemble learning model for COVID-19 detection from blood test samples," *Sensors*, vol. 22, no. 6, p. 2224, Mar. 2022.
[35] J. Gu, S. Liu, Z. Zhou, S. R. Chalov, and Q. Zhuang, "A stacking ensemble learning model for monthly rainfall prediction in the Taihu basin, China," *Water*, vol. 14, no. 3, p. 492, Feb. 2022.
[36] C. Chen and Q. Li, "A multimodal music emotion classification method based on multifeature combined network classifier," *Math. Problems Eng.*, vol. 2020, pp. 1–11, Aug. 2020.
[37] Y. Li and H. Hong, "Modelling flood susceptibility based on deep learning coupling with ensemble learning models," *J. Environ. Manage.*, vol. 325, Jan. 2023, Art. no. 116450.
[38] K.-L. Du and M. N. S. Swamy, "Combining multiple learners: Data fusion and ensemble learning," in *Neural Networks and Statistical Learning*. London, U.K.: Springer, 2019, doi: 10.1007/978-1-4471-7452-3_25.
[39] E. M. G. Younis, S. M. Zaki, E. Kanjo, and E. H. Houssein, "Evaluating ensemble learning methods for multi-modal emotion recognition using sensor data fusion," *Sensors*, vol. 22, no. 15, p. 5611, Jul. 2022.
[40] Z. I. Azhari, S. Setumin, E. Noorsal, and M. H. Abdullah, "Digital image enhancement by brightness and contrast manipulation using verilog hardware description language," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 2, p. 1346, Apr. 2023.
[41] T. Anwar and S. Zakir, "Vehicle make and model recognition using mixed sample data augmentation techniques," *IAES Int. J. Artif. Intell.*, vol. 12, no. 1, p. 137, Mar. 2023.
[42] S. Zhu, W. Chen, F. Liu, X. Zhang, and X. Han, "Human activity recognition based on a modified capsule network," *Mobile Inf. Syst.*, vol. 2023, pp. 1–17, Feb. 2023.
[43] Y.-W. Kim, Y.-C. Byun, and A. V. N. Krishna, "Portrait segmentation using ensemble of heterogeneous deep-learning models," *Entropy*, vol. 23, no. 2, p. 197, Feb. 2021.
[44] Y. O. Ouma, A. Keitsile, B. Nkwae, P. Odirile, D. Moalafhi, and J. Qi, "Urban land-use classification using machine learning classifiers: Comparative evaluation and post-classification multi-feature fusion approach," *Eur. J. Remote Sens.*, vol. 56, no. 1, Dec. 2023, Art. no. 2173659.
[45] J. Zhao, Z. Li, and P. Liu, "Using traffic data to identify land-use characteristics based on ensemble learning approaches," *J. Transp. Land Use*, vol. 16, no. 1, pp. 19–41, Jan. 2023.
[46] A. Verma and Y. Liu, "Hybrid deep learning ensemble model for improved large-scale car recognition," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Aug. 2017, pp. 1–7.
[47] M. H. Salem, Y. Li, Z. Liu, and A. M. AbdelTawab, "A transfer learning and optimized CNN based maritime vessel classification system," *Appl. Sci.*, vol. 13, no. 3, p. 1912, Feb. 2023.
[48] X. Chen, L. Jin, Y. Zhu, C. Luo, and T. Wang, "Text recognition in the wild: A survey," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–35, 2021.
[49] S. Long, X. He, and C. Yao, "Scene text detection and recognition: The deep learning era," *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 161–184, Jan. 2021.
[50] H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li, "Computer vision technology in agricultural automation—A review," *Inf. Process. Agricult.*, vol. 7, no. 1, pp. 1–19, Mar. 2020.
[51] J. Yang, S. Li, Z. Wang, and G. Yang, "Real-time tiny part defect detection system in manufacturing using deep learning," *IEEE Access*, vol. 7, pp. 89278–89291, 2019.
[52] Y. A. Alohali, M. S. Fayed, T. Mesallam, Y. Abdelsamad, F. Almuhawas, and A. Hagr, "A machine learning model to predict citation counts of scientific papers in otology field," *BioMed Res. Int.*, vol. 2022, pp. 1–12, Jul. 2022.
[53] K. V. Archana and G. Komarasamy, "A novel deep learning-based brain tumor detection using the bagging ensemble with K-nearest neighbor," *J. Intell. Syst.*, vol. 32, no. 1, Jan. 2023, Art. no. 20220206.
[54] M. Tariq, V. Palade, Y. Ma, and A. Altahhan, "Diabetic retinopathy detection using transfer and reinforcement learning with effective image preprocessing and data augmentation techniques," in *Fusion of Machine Learning Paradigms: Theory and Applications* (Intelligent Systems Reference Library), vol 236, I. K. Hatzilygeroudis, G. A. Tsihrintzis, and L. C. Jain, Eds. Cham, Switzerland: Springer, 2023, pp. 33–61, doi: 10.1007/978-3-031-22371-6_3.
[55] J. Jiang, L.-C. Xu, F. Li, and J. Shao, "Machine learning potential model based on ensemble bispectrum feature selection and its applicability analysis," *Metals*, vol. 13, no. 1, p. 169, Jan. 2023.
[56] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[57] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.
[58] X. Shen, K. Lu, S. Mehta, J. Zhang, W. Liu, J. Fan, and Z. Zha, "MKEL: Multiple kernel ensemble learning via unified ensemble loss for image classification," *ACM Trans. Intell. Syst. Technol.*, vol. 12, no. 4, pp. 1–21, Aug. 2021.
[59] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
[60] V. B. Semwal, A. Gupta, and P. Lalwani, "An optimized hybrid deep learning model using ensemble learning approach for human walking activities recognition," *J. Supercomput.*, vol. 77, no. 11, pp. 12256–12279, Nov. 2021.
[61] K. Doshi and Y. Yilmaz, "Road damage detection using deep ensemble learning," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2020, pp. 5540–5544.
[62] N. V. Sridharan and V. Sugumaran, "Deep learning-based ensemble model for classification of photovoltaic module visual faults," *Energy Sources A, Recovery, Utilization, Environ. Effects*, vol. 44, no. 2, pp. 5287–5302, Jun. 2022.

[63] A. S. Al-Waisy, D. A. Ibrahim, D. A. Zebari, S. Hammadi, H. Mohammed, M. A. Mohammed, and R. Damaševičius, "Identifying defective solar cells in electroluminescence images using deep feature representations," *PeerJ Comput. Sci.*, vol. 8, p. e992, May 2022.

[64] V. Riego, M. Castejón-Limas, L. Sánchez-González, L. Fernández-Robles, H. Perez, J. Diez-Gonzalez, and Á.-M. Guerrero-Higueras, "Strong classification system for wear identification on milling processes using computer vision and ensemble learning," *Neurocomputing*, vol. 456, pp. 678–684, Oct. 2021.

[65] N. Yu, L. Qian, Y. Huang, and Y. Wu, "Ensemble learning for facial age estimation within non-ideal facial imagery," *IEEE Access*, vol. 7, pp. 97938–97948, 2019.

[66] L. Xu, X. Liu, F. Jiang, Y. Xu, A. Yao, J. Xu, and X. Li, "Multi-featured anomaly detection for mobile edge computing based UAV delivery systems," in *Proc. Australas. Comput. Sci. Week*, Jan. 2023, pp. 58–65.

[67] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[68] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[69] J. Chen, Y. Wang, Y. Wu, and C. Cai, "An ensemble of convolutional neural networks for image classification based on LSTM," in *Proc. Int. Conf. Green Informat. (ICGI)*, Aug. 2017, pp. 217–222.

[70] T. Ridnik, H. Lawen, A. Noy, E. Ben, B. G. Sharir, and I. Friedman, "TResNet: High performance GPU-dedicated architecture," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 1400–1409.

[71] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[72] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2017, pp. 1–11.

[73] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16 × 16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[74] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[75] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6022–6031.

[76] E. D. Cubuk, B. Zoph, D. Mané, V. Vasudevan, and Q. V. Le, "AutoAugment: Learning augmentation strategies from data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 113–123.

[77] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2961–2969.

[78] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[79] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.

[80] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 11211, 2018, pp. 833–851.

[81] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[82] S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A face detection benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5525–5533.

[83] C. Chen, J. Wei, C. Peng, and H. Qin, "Depth-quality-aware salient object detection," *IEEE Trans. Image Process.*, vol. 30, pp. 2350–2363, 2021.

[84] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–14.

[85] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. Eur. Conf. Comput. Vis.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 7576, 2012, pp. 746–760.

[86] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 1, Jan. 2014, pp. 1–9.

[87] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, vol. 1, 2015, pp. 448–456.

[88] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[89] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1933–1941.

[90] B. SravyaPranati, D. Suma, C. ManjuLatha, and S. Putheti, "Large-scale video classification with convolutional neural networks," in *Proc. Int. Conf. Inf. Commun. Technol. Intell. Syst.*, in Smart Innovation, Systems and Technologies, vol. 196, 2021, pp. 689–695.

[91] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[92] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4489–4497.

[93] J. Carreira and A. Zisserman, "Quo vadis, action recognition? A new model and the kinetics dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6299–6308.

[94] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 606–615.

[95] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–42, Sep. 2019.

[96] C. Sammut, "Concept learning," in *Encyclopedia of Machine Learning and Data Mining*, C. Sammut and G. I. Webb, Eds. Boston, MA, USA: Springer, 2017, doi: 10.1007/978-1-4899-7687-1_154.

[97] C. Sammut and G. I. Webb, *Encyclopedia of Machine Learning and Data Mining*. Springer, 2017.

[98] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, Feb. 2020.

[99] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "*mixup*: Beyond empirical risk minimization," in *Proc. 6th Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–13.

[100] S. Han, J. Pool, J. Tran, and W. J. Dally, "Learning both weights and connections for efficient neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2015, pp. 1–9.

[101] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 7, 1995, pp. 1–8.

[102] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–11.

[103] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "A survey of model compression and acceleration for deep neural networks," 2017, *arXiv:1710.09282*.

[104] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.*, vol. 6, no. 3, pp. 21–45, 3rd Quart., 2006.

[105] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. 37th Int. Conf. Mach. Learn. (ICML)*, 2020, pp. 1597–1607.

[106] H. Wu, W. Luo, A. Lin, F. Hao, A.-M. Olteanu-Raimond, L. Liu, and Y. Li, "SALT: A multifeature ensemble learning framework for mapping urban functional zones from VGI data and VHR images," *Comput., Environ. Urban Syst.*, vol. 100, Mar. 2023, Art. no. 101921.

[107] S. Pouyanfar and S.-C. Chen, "Semantic event detection using ensemble deep learning," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2016, pp. 203–208.
[108] S. Qiu, X. Cui, Z. Ping, N. Shan, Z. Li, X. Bao, and X. Xu, "Deep learning techniques in intelligent fault diagnosis and prognosis for industrial systems: A review," *Sensors*, vol. 23, no. 3, p. 1305, Jan. 2023.
[109] K. K. Jena, S. K. Bhoi, S. Mohapatra, and S. Bakshi, "A hybrid deep learning approach for classification of music genres using wavelet and spectrogram analysis," *Neural Comput. Appl.*, vol. 35, no. 15, pp. 11223–11248, May 2023.
[110] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, Feb. 2016.
[111] D. Shen, G. Wu, and H. I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.
[112] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.
[113] K. Kamnitsas, C. Baumgartner, C. Ledig, V. Newcombe, J. Simpson, A. Kane, D. Menon, A. Nori, A. Criminisi, D. Rueckert, and B. Glocker, "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 10265, 2017, pp. 597–609.
[114] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," *Nature Rev. Cancer*, vol. 18, no. 8, pp. 500–510, 2018.
[115] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Proc. 10th Eur. Conf. Mach. Learn.*, Chemnitz, Germany, 1998, pp. 137–142.
[116] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proc. 1st Int. Conf. Learn. Represent.*, 2013, pp. 1–12.
[117] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2002, pp. 1–9.
[118] M. S. Başarslan and F. Kayaalp, "Sentiment analysis on social media reviews datasets with deep learning approach," *Sakarya Univ. J. Comput. Inf. Sci.*, vol. 4, no. 1, pp. 35–49, Apr. 2021.
[119] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014.
[120] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowl.-Based Syst.*, vol. 226, Aug. 2021, Art. no. 107134.
[121] Z. Yang, J. Ye, L. Wang, X. Lin, and L. He, "Inferring substitutable and complementary products with knowledge-aware path reasoning based on dynamic policy network," *Knowl.-Based Syst.*, vol. 235, Jan. 2022, Art. no. 107579.
[122] U. Ahmed, J. C. Lin, and G. Srivastava, "Generative ensemble learning for mitigating adversarial malware detection in IoT," in *Proc. IEEE 29th Int. Conf. Netw. Protocols (ICNP)*. Washington, DC, USA: IEEE Computer Society, Nov. 2021, pp. 1–5.
[123] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.
[124] L. Liu, Y. Ji, Y. Gao, T. Li, and W. Xu, "A data-driven adaptive emotion recognition model for college students using an improved multifeature deep neural network technology," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–9, May 2022.
[125] B. Priyamvada, S. Singhal, A. Nayyar, R. Jain, P. Goel, M. Rani, and M. Srivastava, "Stacked CNN–LSTM approach for prediction of suicidal ideation on social media," *Multimedia Tools Appl.*, vol. 82, no. 18, pp. 27883–27904, Jul. 2023.
[126] J. Guo, "Deep learning approach to text analysis for human emotion detection from big data," *J. Intell. Syst.*, vol. 31, no. 1, pp. 113–126, 2022.
[127] M. M. Rahman, S. S. M. M. Rahman, S. M. Allayear, M. F. K. Patwary, and M. T. A. Munna, "A sentiment analysis based approach for understanding the user satisfaction on Android application," in *Data Engineering and Communication Technology* (Advances in Intelligent Systems and Computing), vol. 1079, K. Raju, Senkerik, Lanka, and V. Rajagopal, Eds. Singapore: Springer, 2020, doi: 10.1007/978-981-15-1097-7_33.
[128] M.-H. Chao, A. J. C. Trappey, and C.-T. Wu, "Emerging technologies of natural language-enabled chatbots: A review and trend forecast using intelligent ontology extraction and patent analytics," *Complexity*, vol. 2021, pp. 1–26, May 2021.
[129] R. Alghamdi and M. Bellaiche, "An ensemble deep learning based IDS for IoT using lambda architecture," *Cybersecurity*, vol. 6, no. 1, p. 5, Mar. 2023.
[130] J. Li, S. Zhang, Y. Zhang, H. Lin, and J. Wang, "Multifeature fusion attention network for suicide risk assessment based on social media: Algorithm development and validation," *JMIR Med. Informat.*, vol. 9, no. 7, Jul. 2021, Art. no. e28227.
[131] M. Afzal, F. Alam, K. M. Malik, and G. M. Malik, "Clinical context–aware biomedical text summarization using deep neural network: Model development and validation," *J. Med. Internet Res.*, vol. 22, no. 10, Oct. 2020, Art. no. e19810.
[132] B. A. Hassan and T. A. Rashid, "Artificial intelligence algorithms for natural language processing and the semantic web ontology learning," 2021, *arXiv:2108.13772*.
[133] J. Copara, N. Naderi, J. Knafou, P. Ruch, and D. Teodoro, "Named entity recognition in chemical patents using ensemble of contextual language models," in *Proc. CEUR Workshop*, vol. 2696, 2020, pp. 1–15.
[134] C. Guo and F. Berkhahn, "Entity embeddings of categorical variables," 2016, *arXiv:1604.06737*.
[135] X. Li, H. Zhang, and X.-H. Zhou, "Chinese clinical named entity recognition with variant neural structures based on BERT methods," *J. Biomed. Informat.*, vol. 107, Jul. 2020, Art. no. 103422.
[136] Y. Sun, Y. Liu, G. Wang, and H. Zhang, "Deep learning for plant identification in natural environment," *Comput. Intell. Neurosci.*, vol. 2017, pp. 1–6, May 2017.
[137] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. 2nd Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 1–14.
[138] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Neural Inf. Process. Syst.*, vol. 1, 2006, pp. 1–9.
[139] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[140] T. Mikolov, M. Karafiát, L. Burget, J. Černocký, and S. Khudanpur, "Recurrent neural network based language model," in *Proc. Interspeech*, Sep. 2010, pp. 1045–1048.
[141] K. L. Tan, C. P. Lee, K. M. Lim, and K. S. M. Anbananthen, "Sentiment analysis with ensemble hybrid deep learning model," *IEEE Access*, vol. 10, pp. 103694–103704, 2022.
[142] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Jan. 2003.
[143] T. A. Almeida, J. M. G. Hidalgo, and A. Yamakami, "Contributions to the study of SMS spam filtering: New collection and results," in *Proc. 11th ACM Symp. Document Eng.*, Sep. 2011, pp. 259–262.
[144] W. Y. Wang, "'Liar, liar pants on fire': A new benchmark dataset for fake news detection," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, 2017, pp. 422–426.
[145] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep patient: An unsupervised representation to predict the future of patients from the electronic health records," *Sci. Rep.*, vol. 6, no. 1, pp. 1–10, May 2016.
[146] I. AbdulNabi and Q. Yaseen, "Spam email detection using deep learning techniques," *Proc. Comput. Sci.*, vol. 184, pp. 853–858, Jan. 2021.
[147] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
[148] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol. (NAACL-HLT)*, vol. 1, 2019, p. 2.
[149] T. B. Brown, D. Mané, A. Roy, M. Abadi, and J. Gilmer, "Adversarial patch," 2017, *arXiv:1712.09665*.
[150] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, *arXiv:1907.11692*.
[151] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2018, pp. 2227–2237.

[152] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," *Homol., Homotopy Appl.*, vol. 9, no. 1, pp. 1–12, 2018.

[153] N. P. Shetty, B. Muniyal, A. Anand, S. Kumar, and S. Prabhu, "Predicting depression using deep learning and ensemble algorithms on raw Twitter data," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 4, p. 3751, Aug. 2020.

[154] S. Wei and S. Song, "Sentiment classification of tourism reviews based on visual and textual multifeature fusion," *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1–10, May 2022.

[155] W. Hou, Y. Li, Y. Liu, and Q. Li, "Leveraging multidimensional features for policy opinion sentiment prediction," *Inf. Sci.*, vol. 610, pp. 215–234, Sep. 2022.

[156] A. S. Sams and A. Zahra, "Multimodal music emotion recognition in Indonesian songs based on CNN-LSTM, XLNet transformers," *Bull. Electr. Eng. Informat.*, vol. 12, no. 1, pp. 355–364, Feb. 2023.

[157] J. M. Torres, C. I. Comesaña, and P. J. García-Nieto, "Machine learning techniques applied to cybersecurity," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 10, pp. 2823–2836, 2019.

[158] X. Liang, A. Angelopoulou, E. Kapetanios, B. Woll, R. A. Batat, and T. Woolfe, "A multi-modal machine learning approach and toolkit to automate recognition of early stages of dementia among British sign language users," in *Proc. Eur. Conf. Comput. Vis.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 12536, 2020, pp. 278–293.

[159] M. Somesha, A. R. Pais, R. S. Rao, and V. S. Rathour, "Efficient deep learning techniques for the detection of phishing websites," *Sādhanā*, vol. 45, no. 1, pp. 1–18, Dec. 2020.

[160] F. Ghobadi and D. Kang, "Application of machine learning in water resources management: A systematic literature review," *Water*, vol. 15, no. 4, p. 620, Feb. 2023.

[161] M. Presa-Reyes, Y. Tao, S.-C. Chen, and M.-L. Shyu, "Deep learning with weak supervision for disaster scene description in low-altitude imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4704510.

[162] Y. Bai, Y. Zhao, Y. Shao, X. Zhang, and X. Yuan, "Deep learning in different remote sensing image categories and applications: Status and prospects," *Int. J. Remote Sens.*, vol. 43, no. 5, pp. 1800–1847, Mar. 2022.

[163] Z. Fu, "Computer network intrusion anomaly detection with recurrent neural network," *Mobile Inf. Syst.*, vol. 2022, pp. 1–11, Mar. 2022.

[164] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018.

[165] X. Qiu, M. Li, L. Dong, G. Deng, and L. Zhang, "Dual-band maritime imagery ship classification based on multilayer convolutional feature fusion," *J. Sensors*, vol. 2020, pp. 1–16, Dec. 2020.

[166] A. Grzenda, N. V. Kraguljac, W. M. McDonald, C. Nemeroff, J. Torous, J. E. Alpert, C. I. Rodriguez, and A. S. Widge, "Evaluating the machine learning literature: A primer and user's guide for psychiatrists," *Amer. J. Psychiatry*, vol. 178, no. 8, pp. 715–729, Aug. 2021.

[167] M. Raj, S. Singh, K. Solanki, and R. Selvanambi, "An application to detect cyberbullying using machine learning and deep learning techniques," *Social Netw. Comput. Sci.*, vol. 3, no. 5, p. 401, Jul. 2022.

[168] A. Ghosh and S. Saha, "Sensing the mood-application of machine learning in human psychology analysis and cognitive science," in *Smart Computational Intelligence in Biomedical and Health Informatics*. Boca Raton, MA, USA: CRC Press, 2021.

[169] J. Wang and S. Li, "Maritime radar target detection in sea clutter based on CNN with dual-perspective attention," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[170] H. Xu and Z. Zhao, "NetBCE: An interpretable deep neural network for accurate prediction of linear B-cell epitopes," *Genomics, Proteomics Bioinf.*, vol. 20, no. 5, pp. 1002–1012, Oct. 2022.

[171] H. Liu, Z. Liu, W. Jia, and X. Lin, "Remaining useful life prediction using a novel feature-attention-based end-to-end approach," *IEEE Trans. Ind. Informat.*, vol. 17, no. 2, pp. 1197–1207, Feb. 2021.

[172] X. Liu, P. He, W. Chen, and J. Gao, "Multi-task deep neural networks for natural language understanding," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 4487–4496.

[173] Y. Zhang, P. Qi, and C. D. Manning, "Graph convolution over pruned dependency trees improves relation extraction," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 2205–2215.

[174] P. Rajpurkar, R. Jia, and P. Liang, "Know what you don't know: Unanswerable questions for SQuAD," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, pp. 784–789.

[175] T. Kwiatkowski, J. Palomaki, O. Redfield, M. Collins, A. Parikh, C. Alberti, D. Epstein, I. Polosukhin, J. Devlin, K. Lee, K. Toutanova, L. Jones, M. Kelcey, M.-W. Chang, A. M. Dai, J. Uszkoreit, Q. Le, and S. Petrov, "Natural questions: A benchmark for question answering research," *Trans. Assoc. Comput. Linguistics*, vol. 7, pp. 453–466, Nov. 2019.

[176] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 4580–4584.

[177] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[178] W. Li, M. Luo, P. Zhang, and W. Huang, "A novel multi-feature joint learning ensemble framework for multi-label facial expression recognition," *IEEE Access*, vol. 9, pp. 119766–119777, 2021.

[179] H. A. Mengash, L. Hussain, H. Mahgoub, A. Al-Qarafi, M. K. Nour, R. Marzouk, S. A. Qureshi, and A. M. Hilal, "Smart cities-based improving atmospheric particulate matters prediction using chi-square feature selection methods by employing machine learning techniques," *Appl. Artif. Intell.*, vol. 36, no. 1, Dec. 2022, Art. no. 2067647.

[180] F. Gao, Q. Wang, J. Dong, and Q. Xu, "Spectral and spatial classification of hyperspectral images based on random multi-graphs," *Remote Sens.*, vol. 10, no. 8, p. 1271, Aug. 2018.

[181] C. Shan, J. Zhang, Y. Wang, and L. Xie, "Attention-based end-to-end speech recognition on voice search," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 4764–4768.

[182] A. Gulati, J. Qin, C.-C. Chiu, N. Parmar, Y. Zhang, J. Yu, W. Han, S. Wang, Z. Zhang, Y. Wu, and R. Pang, "Conformer: Convolution-augmented transformer for speech recognition," 2020, *arXiv:2005.08100*.

[183] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Mach. Learn.*, vol. 51, no. 2, pp. 181–207, 2003.

[184] D. Opitz and R. Maclin, "Popular ensemble methods: An empirical study," *J. Artif. Intell. Res.*, vol. 11, pp. 169–198, Aug. 1999.

[185] L. Rokach, "Ensemble-based classifiers," *Artif. Intell. Rev.*, vol. 33, nos. 1–2, pp. 1–39, Feb. 2010.

[186] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, Jan. 1992.

[187] X. Song, H. Chen, Q. Wang, Y. Chen, M. Tian, and H. Tang, "A review of audio-visual fusion with machine learning," *J. Phys., Conf. Ser.*, vol. 1237, no. 2, Jun. 2019, Art. no. 022144.

[188] R. G. Praveen, E. Granger, and P. Cardinal, "Cross attentional audio-visual fusion for dimensional emotion recognition," in *Proc. 16th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Dec. 2021, pp. 1–8.

[189] L. R. Biggers, C. Bocovich, R. Capshaw, B. P. Eddy, L. H. Etzkorn, and N. A. Kraft, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," *Empirical Softw. Eng.*, vol. 19, no. 3, 2014. G. Montavon, A. Binder, S. Lapuschkin, W. Samek, and K. R. Müller, *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, vol. 11700. Cham, Switzerland: Springer, 2019, doi: 10.1007/978-3-030-28954-6.

[190] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2017, pp. 1–10.

[191] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1135–1144.

[192] A. B. Arrieta, N. Díaz-Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.

[193] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," 2017, *arXiv:1702.08608*.

[194] F. Seide and A. Agarwal, "CNTK: Microsoft's open-source deep-learning toolkit," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, p. 2135.

[195] W. Wen, C. Wu, Y. Wang, Y. Chen, and H. Li, "Learning structured sparsity in deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.
[196] P. Micikevicius, S. Narang, J. Alben, G. Diamos, E. Elsen, D. Garcia, B. Ginsburg, M. Houston, O. Kuchaiev, G. Venkatesh, and H. Wu, "Mixed precision training," 2017, *arXiv:1710.03740*.
[197] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*.
[198] R. Raina, A. Madhavan, and A. Y. Ng, "Large-scale deep unsupervised learning using graphics processors," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, Jun. 2009, pp. 873–880.
[199] N. P. Jouppi et al., "In-datacenter performance analysis of a tensor processing unit," in *Proc. ACM/IEEE 44th Annu. Int. Symp. Comput. Archit. (ISCA)*, Jun. 2017, pp. 1–12.
[200] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, no. 6197, pp. 668–673, Aug. 2014.
[201] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[202] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, M. Ranzato, A. Senior, P. Tucker, K. Yang, Q. Le, and A. Ng, "Large scale distributed deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1–9.
[203] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[204] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[205] R. Vilalta and Y. Drissi, "A perspective view and survey of meta-learning," *Artif. Intell. Rev.*, vol. 18, no. 2, pp. 77–95, Oct. 2002.
[206] E. Real, S. Moore, A. Selle, S. Saxena, Y. L. Suematsu, J. Tan, Q. V. Le, and A. Kurakin, "Large-scale evolution of image classifiers," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, vol. 6, 2017, pp. 4429–4446.
[207] J. Snoek and H. Larochelle, "Practical Bayesian optimization of machine learning algorithms," *Religion Arts*, vol. 17, nos. 1–2, pp. 57–73, 2013.
[208] Y. Li, T. Zhang, S. Sun, and X. Gao, "Accelerating flash calculation through deep learning methods," *J. Comput. Phys.*, vol. 394, pp. 153–165, Oct. 2019.
[209] S. Arora, N. Golowich, N. Cohen, and W. Hu, "A convergence analysis of gradient descent for deep linear neural networks," in *Proc. 7th Int. Conf. Learn. Represent. (ICLR)*, 2019, pp. 1–35.
[210] N. Papernot, P. McDaniel, X. Wu, S. Jha, and A. Swami, "Distillation as a defense to adversarial perturbations against deep neural networks," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2016, pp. 582–597.
[211] D. Tsipras, S. Santurkar, L. Engstrom, A. Turner, and A. Madry, "Robustness may be at odds with accuracy," in *Proc. 7th Int. Conf. Learn. Represent. (ICLR)*, 2019, pp. 1–24.
[212] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2017, pp. 1273–1282.
[213] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.
[214] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Theory Cryptogr. Conf.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 3876, 2006, pp. 265–284.
[215] O. Goldreich, *Foundations of Cryptography: A Primer*, no. 1. Hanover, MD, USA: Now Publishers, 2005.
[216] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, 2009.
[217] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 1–35, 2016.
[218] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, vol. 3, 2017, pp. 1126–1135.
[219] K. Hsieh, A. Harlap, N. Vijaykumar, D. Konomis, G. R. Ganger, P. B. Gibbons, and O. Mutlu, "Gaia: Geo-distributed machine learning approaching LAN speeds," in *Proc. 14th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, 2017, pp. 629–647.
[220] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2016, *arXiv:1610.05492*.
[221] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[222] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, "Machine learning with adversaries: Byzantine tolerant gradient descent," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2017, pp. 1–11.

SATHEESH ABIMANNAN received the M.E. degree in computer science and engineering from the College of Engineering (Guindy), Anna University, Chennai, and the Ph.D. degree in computer science and engineering from Periyar Maniammai University. He is currently a Professor and the Deputy Director of the Amity School Engineering and Technology, Amity University, Mumbai. He was a Postdoctoral Research Fellow with National Taipei University, Taiwan, for one year. He has more than 20 years of teaching, research, and administrative experience. He received the ISTE-Young Scientist Award, in 2010. He has published more than 40 research articles in highly reputed international journals and visited Singapore, China, Taiwan, and Japan, to present his research article at international conferences. His research interests include deep learning, cloud computing, big-data analytics, and information security.

EL-SAYED M. EL-ALFY (Senior Member, IEEE) is currently a Professor with the Information and Computer Science Department and a fellow of the SDAIA-KFUPM Joint Research Center for Artificial Intelligence, Interdisciplinary Research Center Affiliate for Intelligent Secure Systems (IRC-ISS), King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia. He has over 25 years of experience in industry and academia, involving research, teaching, supervision, curriculum design, program assessment, and quality assurance in higher education. He is also an approved ABET/CSAB Program Evaluator (PEV), a Reviewer, and a Consultant of NCAAA and several universities and research agents in various countries. He is also an active researcher with interests in fields related to machine learning and nature-inspired computing and applications to data science and cybersecurity analytics, pattern recognition, multimedia forensics, and security systems. He has published numerously in peer-reviewed international journals and conferences, edited a number of books published by reputable international publishers, attended and contributed in the organization of many world-class international conferences, and supervised master's and Ph.D. students. He was also a member of ACM, the IEEE Computational Intelligence Society, the IEEE Computer Society, the IEEE Communication Society, and the IEEE Vehicular Technology Society. His work has been internationally recognized, received a number of awards, and appeared in the Stanford University world's top 2% of scientists list. He has served as a guest editor for a number of special issues in international journals. He has been on the editorial board of a number of premium international journals, including IEEE/CAA Journal of Automatica Sinica, IEEE Transactions on Neural Networks and Learning Systems, *International Journal of Trust Management in Computing and Communications*, and *Journal of Emerging Technologies in Web Intelligence*.

YUE-SHAN CHANG (Senior Member, IEEE) received the Ph.D. degree from the Department of Computer and Information Science, National Chiao Tung University, in 2001. In August 1992, he joined the Department of Electronic Engineering, Ming Hsing University of Science and Technology. In August 2004, he joined the Department of Computer Science and Information Engineering, National Taipei University, Taipei, Taiwan. Since August 2011, he has been a professor. He served as the Chairperson of the Department, in 2014, the Dean of Student Affairs, in 2018, and the Dean of Academic Affairs, in 2021. He is also the Dean of the College of Electrical Engineering and Computer Science. Since August 2022, he has been promoted to be a distinguished professor. His research interests include deep learning, big data analytics, cloud computing, intelligent computing, and the Internet of Things.

SHAHID HUSSAIN received the B.S. degree in mathematics and the M.Sc. degree in computer science from the University of Peshawar, in 2002 and 2005, respectively, and the M.S. and Ph.D. degrees in computer engineering from Jeonbuk National University, South Korea, in 2016 and 2020, respectively. He achieved Jeonbuk National University presidential award for academic excellence during the Ph.D. studies. He was a Postdoctoral Researcher with the Gwangju Institute of Science and Technology (GIST), South Korea, in 2020, and the University of Galway (UoG), Ireland, from 2020 to 2022. He is currently a Senior Postdoctoral Researcher with the School of Business, Innovative Value Institute (IVI), National University of Ireland Maynooth (NUIM), Ireland. His research interests include smart grid, energy management, electric vehicles, smart grid infrastructure, optimization algorithms, micro-grid operations, distributed energy resources, peer-to-peer energy trading, machine learning in medical applications (e.g., prediction and risk analysis of osteoporosis) using fuzzy logic, game theory, ontology, AI, and blockchain approaches and technologies.

SAURABH SHUKLA received the B.Tech. degree from Dr. A. P. J. Abdul Kalam Technical University, Uttar Pradesh, Lucknow, India, in 2008, the M.Tech. degree from the Indian Institute of Information Technology (IIIT), Allahabad, India, in 2010, in the research area of an intelligent systems, and the Ph.D. degree from Universiti Teknologi Petronas (UTP), Malaysia, in the research area of healthcare Internet of Things (IoT), in August 2020. He joined the Unit of Semantic Web, Data Science Institute (DSI), Insight SFI Centre of Data Analytics, National University of Ireland Galway (NUIG), as a Postdoctoral Researcher, in October 2020. His research interests include the healthcare Internet of Things, fog computing (FC), cloud computing, machine learning, and blockchain. He is currently working on cooperative energy trading system (CENTS) project for an efficient peer-2-peer energy trading system in a smart grid (SG) network. He has academic experience of more than seven years and published around 20 papers in various international journals and conferences.

DHIVYADHARSINI SATHEESH is currently pursuing the B.Tech. degree in computer science and engineering with the School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology (VIT), Vellore, Tamil Nadu, India. She is also a promising scholar. Her academic journey is marked by a keen interest in a broad spectrum of areas within the field of computer science. This includes edge-intelligence, air pollution forecasting, data analytics, artificial intelligence (AI), and machine learning (ML).