Comparative Analysis of Performance of Deep Learning Classification Approach based on LSTM-RNN for Textual and Image Datasets

Alaa Sahl Gaafar and Jasim Mohammed Dahr
Directorate of Education in Basrah, Basrah, Iraq.
E-mail: alaasy.2040@gmail.com and Jmd20586@gmail.com

Alaa Khalaf Hamoud
Department of Computer Information Systems, University of Basrah, Basrah, Iraq
E-mail: alaa.hamoud@uobasrah.edu.iq

Deep learning approaches can be applied to a large amount of data for the purpose of simplifying and improving the engineering practice of automated decision-making activities rather than relying on human encoded heuristics. The need for generating faster and effective decisions about systems, processes, and applications gave rise to many artificial intelligences motivated approaches such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), fuzzy analytics, etc. Deep learning deploys diverse multiple layers of cascaded processing elements to enable features extraction and transformations. These deep learning approaches conduct multiple levels of depiction corresponding to distinct abstraction levels. There are several applications of deep learning algorithms including weather forecasting, object recognition, stock market performance forecasts, medical diagnosis, and emergency warning systems. This paper investigates the performance of the deep learning approach on the basis of processing components, data representation, and data types. To achieve this, a deep learning algorithm based on a long short-term memory-recurrent neural network (LSTM-RNN) was utilized to learn hidden patterns and features in the textual and image datasets respectively. The outcomes reveal that the performance of the image-based deep learning model was better in terms of speed due to well-defined patterns of data representation against the data with sentiments-based deep learning by 3.49 mins. to 18.25 mins. While the LSTM-RNN with images offered better classification accuracy by 96.50% to 85.69% due to complex network architecture, processing elements, and features of the underlying datasets.
Povzetek: .

1   Introduction

The concept of transfer learning or knowledge transfer is a machine technique that allows the reuse of a pre-trained model in the source domain as a point of reference in a dissimilar but associated target domain. Machine learning makes use of an algorithm to perform specific functions or tasks but, this can be extended to fresh tasks by means of transfer learning. Deep learning and machine learning are data-driven models which can be trained on an end-to-end basis for the purpose of accomplishing tasks such as feature extraction, feature selection, malignancy forecasts and optimization tasks [1].

CNNs is a feature depiction model that makes use of a collection of convolutional kernels in its convolution layer to extract features in high-dimensional structural data especially multi-channel inputs such as colour images. Many variants of CNNs structure have evolved over the years including: ResNet, VGG, GoogleNet, AlexNet and DenseNet for computer vision and image processing tasks [2]. CNN models perform computations by utilizing numerous processing layers to generate features inside raw data at multilevel depiction and hierarchical abstraction [1]. The effectiveness of CNN depends on the exploitation of domain knowledge concerning features invariances inside its structure which has been deployed on diverse recognition tasks such as image processing. This puts CNN over the standard fully connected deep neural networks (DNNs) [3].

Content-based learning methods have been applied to textual information for purpose of revealing hidden structures and meanings represented in linguistic/writing styles. Natural language processing enables crossed multiple levels of language represented in different hand-crafted features to be extracted before classification operations on their sentences, phrases and other syntactic features. Machine learning and deep learning techniques have been utilized to extracting features [4] within news content in order to classify fake news and rhetoric in case of political bias. SVM, random forest, vector space model, and bag-of-words features. CNNs, RNNs have recently been adopted for extracting the explicit and latent multi-

modal features with news contents [5]. The concept of fake news gave rise to machine learning based on enormous annotated data for the English language and few others. The deep learning models effectiveness in two cases of data representation, that is, images and textual with LSTM-RNNs classifications are investigated in this article.

2   Literature review

2.1   Overview of Convolutional Neural Network

The CNN has its origin in the biological science domain for purpose classification tasks [6]. It is considered to be a special neural network because of its kind of arrangement in terms of full weight sharing. In this CNN arrangement, the foremost layer comprises of various feature maps (convolution layer) [3,7]. The convolution layer consists of neurons for accepting input emanating from a local receptive field matching features of a limited frequency range. The sets of neurons belonging to the similar feature map share the same weights known as kernels (or filters), which accept distinct frequency-shifted inputs. Consequently, the convolution layer performs convolution of the kernels using lower layer activations.

The pooling layer is located on top of the convolutional layer for the purpose of computing a lower resolution depiction of the convolution layer activations by means of sub-sampling. The pooling function generates certain statistics of the activation which is applied to the neurons using a frequency band window realized from the same feature map in the convolutional layer. Also, the max pooling function is used to compute the feature with maximum value on the matching frequency bands. In particular, max-pooling and weight sharing are significant for attaining invariance to the small frequency shifts, and reduction of over-fitting caused by reduced amounts of trainable parameters.

The higher level of features can be realized by combining or stacking pairs of convolution-pooling. Then, standard fully connected layers are introduced to aggregate the diverse bands of features.

In image processing, CNN is considered to be potent because its weight sharing approach enables similar image patterns can be depicted at any point in the image. This makes the limited weight sharing approach most appropriate in that different (or unshared) kernel is applied at the different frequency window inside the convolution layer. It processes only a limited range of input bands and produces an output band in the pooling layer. The basic architecture of CNN is depicted in Figure 1 [8,9].

CNNs are categorized to match the number of layers in their architectures including: shallow layer CNN, deep layer CNN, extract features from pre-trained CNN, Fine-tune pre-trained CNN which are implemented for many image classifications tasks [9].

2.2   Model of CNNs

The CNN model is similar to the structure of artificial neural network (ANN) in that they are consist of neurons that improve the results through learning. In ANN and CNN, the neuron will receive the input to perform an operation to rectify the weights assigned to each connection between neurons. The loss functions that associated with the last layer in the ANN are the same in CNN. CNN is often used in the pattern recognition field in images. The features of image are encoded into network architecture to make the focus more accurate and to reduce the parameters required for the model set up step [10].

The basic structure of neuron model that consists the ANN and CNN model can be [11]:

a. Single input neuron
In this model as shown in Figure 2, the scalar input p and the scalar weight w are multiplied to produce the wp, where this result is sent to the summation function. The Figure 1 is the other input and the bias b are multiplied and passed to the summation function. The transfer function receives the output of summation function n to produce the scalar output of neuron.

b. Multiple input neuron
The neuron in this model has more than one input as shown in Figure 2. A weight element is assign to each individual input in the network. The weighted matrix is start with $w_{1,1}$, $w_{1,2}$, ….$w_{1,R}$. The bias within neuron is summed with the value of the weighted inputs to produce the net input as in equation (1):

$$n = w_{1,1}P_1 + w_{1,2}P_2 + \cdots + w_{1,R}P_{3R} + b \quad (1)$$

The output is calculated as calculated in equation (2):

$$a = f(Wp + b) \quad (2)$$

Recurrent Neural Network (RNN) is a class of powerful deep neural network using its internal memory with loops to deal with sequence data. The architecture of RNNs, which also is the basic structure of LSTMs [12]. For a hidden layer in RNN, it receives an input vector, and generates the output vector. RNNs exhibit the superior capability of adapting themselves to predict nonlinear time
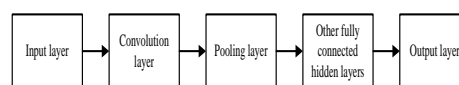

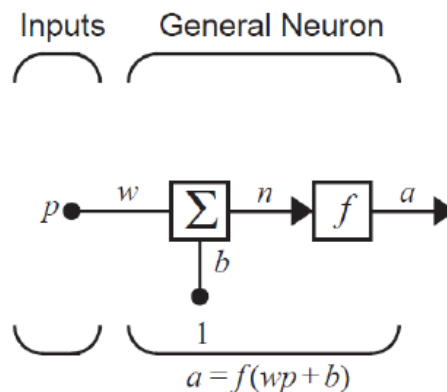Figure 1: The basic structure of CNN.

Figure 2: Single input neuron.

series problems. Though, certain RNNs are bound to reach the vanishing with the Backpropagation coefficient learning, thereby making them unsuitable for long period lags learning nor accounting for long-term dependencies. These short-lived the widespread usage of RNNs, which gave rise to more improved approaches including the Long short-term memory (LSTM) and Gated Recurrent Unit (GRU) architectures. In recent applications, the LSTMs have shown promise on sequence-based computations with long-term dependencies. Though, GRU is an abridged LSTM architecture, which is a relative innovation in machine translation tasks such as SemSeq4FD [13].

2.3　The Real-valued CNNs

In computer vision, CNNs are considered to be the most effective models in performing tasks. After the huge success of LeNet in the domain of digit recognition, AlexNet was developed as the foremost deep CNN. It highly outperforms previous image classification modeling tasks [14]. Thereafter, several models having complicated and deep structures were developed including VGG [15] and ResNet [16], which have performed incredibly well in ILSVRC. Though, the CNN models have been extended to low-level vision tasks as in the case of SRCNN for image super-resolution. Aside from this, CNNs have been applied for denoising and in painting with encouraging outcomes. Recently, there are a number of efforts in extending real-valued neural networks to several number-related fields [1]. In the main time, complex-valued neural networks have been developed and proved to take advantage of generalization capability [17] not without the ease of optimization. In particular, audio signals have complex number representations which make complex CNNs idea for these tasks rather than real-valued CNNs. Nevertheless, there evidence that deep complex networks offer improved outcomes when compared with real-valued models in terms of audio-based tasks [18]. Another deep quaternion network is put forward by [19] in which its quaternion convolution basically substitutes the real multiplications, and its quaternion kernel is not parameterized further.

2.4　Deep Sentiment with CNNs

The concept of sentiment analysis became an interesting area of research after the breakout applications of social networks and e-commerce websites. The foremost meaning of sentiment is to understand the public views and opinions concerning a particular topic, trends or product by means of tweets or reviews. Later, it was extended to the political and social trends extraction through annotation of opinions. Recently, machine learning algorithms were deployed to automatically generate annotations from diverse points of views to form a real-world knowledge base to map various scenarios to associated emotions for the purpose of extracting an overall sentiment in a statement. The advent of computer vision and natural language processing offered the opportunity to combine enormous labeled datasets which make deep learning models suitable. There is a shift from

the traditional approach of extracting hand-crafted features from images and texts and supplying them to the classification model. At present, it is possible to conduct end-to-end models in order to simultaneously lean the feature depiction and classify accordingly. Deep learning inspired models are capable of undertaking many tasks including machine translation, object detection, word embedding, sentiment analysis, name entity in NLP, question answering, image classification, image generation, and computer vision's unsupervised feature learning. The process of conducting sentiment analysis involves progressive training on certain labeled data known as transfer learning such as CNNs, and LSTM [20].

3　Related works

3.1　Image classification

A hybrid transfer learning approach based on CNNs' AlexNet and VGG-16 net were developed for the purpose of detecting cervical cancer. The technique is based on the reduction of filters needed for undertaking the tasks especially the feature selection process by means of convolutional layers of pre-trained networks considered [9]. A fresh image dataset together with expert annotated diagnosis for performing image-based cervical disease classification tasks was proposed by [21]. A fine-tuned CNN model for extracting the several features in the images before classifying with classifiers such as ensemble-tree models, and multi-layer perceptron models. The effectiveness of randomized and optimized weights of CNN model was investigated by [22]. These models were applied to linear and non-linear data types with more desirable outcomes for CNN with optimized weights. In a separate work, the effectiveness of CNN was tested in classifying habitus images for the purpose of recognizing insects, which is at 74.9% according to [4]. To ascertain optimal deep learning schemes for point-of-care ultrasound (POCUS) cardiac image classification tasks in [23]. Six pretrained CNNs including AlexNet, Inception-v4, VGG-16, VGG-19, ResNet50, and DenseNet201 were trained with five samples of POCUS images which generated accuracy between 96%-85.6%. VGG-16 was the best while was the least performing classifier. Multi-Object-based CNN (OCNN) was proposed DenseNet201 by [24] to improve the performance of large-area classification across diverse landscapes, which is composed of image segmentation, skeleton-driven algorithm object analysis, CNN final classification. The overall results showed 87.2% for sensed images.

The peculiarity of Hyperspectral image (HSI) makes it challenging for classification tasks due to enormous bands with close correlation in spatial and spectral domains. The problem of smaller training samples is even more difficult which led to use of deep CNN for extracting spatial features in HSI. In addition, AdaMAx, RMSprop, SGD, Adadelta, and Adagrad classifiers were used to optimize the deep CNN model. The outcomes relieved that deep CNN with Adam classifier produced the best performance at $98.97 \pm 0.81\%$ according to [25]. Another similar work on HSI classification utilized CNN with

Markov Random Field (MRF) in order to enhance effectiveness [26]. Again, In [27], the authors investigated a combination of CNNs for learning joint spatial-spectral features effectively across multiple scale in HSI which produced 66.73% accuracy. The adoption of machine learning technique for detecting early stages and severity of Diabetic Retinopathy based on colour fundus image. Hyperparameter tuning Inception-v4 (HPTT-v4) model was utilized by [28] for detection and classification tasks with 99.49% performance accuracy. In [29], the authors proposed enhanced feature selection-based model for medical image classification. Opposition-based Crow Search algorithm was used to optimize the DL model with 95.22% accuracy.

3.2   Textual classification

The ensemble model founded on CNN and LSTM for extracting temporal information and local structure of textual datasets was developed by [20]. The performance of the proposed model outperforms the individual model and previous studies. Fake news classification model based on reinforcement learning of textual content was proposed by [30]. It is made up of an annotator, reinforced selector and fake news detector for automatically generating labels of unlabeled news items, selecting high-quality samples for training, and detection of fake or genuine news items. A fusion method of multi-modal features extraction with CNN, Inception, Multilingual-BERT, XLNet, XLM-Roberta and VGG16 models for the image and textual dataset was used to investigate social media memes by [31]. Sentiment analysis of textual items based on lexicon integrated two-channel CNN-LSTM family models was proposed by [32]. It is a deep learning approach that combined CNN and LSTM/BiLSTM channels in a concurrent manner whose results are encouraging for different datasets considered. Another study on a multimodal fusion of linguistic and speech in the detection of depression using Gated CNN, LSTM, and CNN was experimented by [33]. The hybrid model of audio features extraction of the GCNN-LSTM model and text features acquired from the CNN-LSTM architecture was better in both CCC and RMSE parameters for the development and the test datasets. An automatic approach for labeling on-topic social media posts composed of visual-textual features was experimented by [7]. The extractions of textual and visual features were achieved by word embedded CNN and Inception-V3 CNN. The classification process utilized concatenated features for effective and timely disaster reporting and mitigation. A model based on deep CNN and LSTM was developed to enhance the accuracy of image capturing [34]. The model based on gated CNN and BERT-CNN multimodal approach was used to determine depression detection in the work by [33].

A texture recognition task based on optimized CNN was developed by [6]. It optimized filter values, and weights, and biased values in the convolution layer and fully-connected layer by means of a whale optimization algorithm. The speech modality features were extracted from the trained VGG-16 network then Gated CNN and
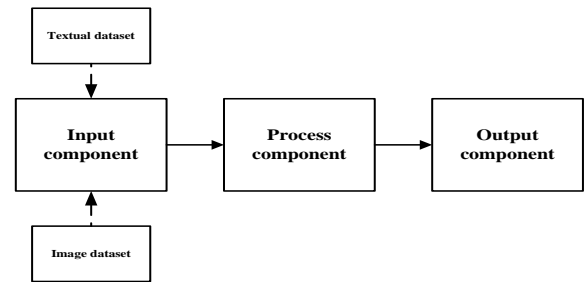

Figure 3: Conceptual structure of proposed deep learning models.

LSTM layer. The textual features embedding, BERT features from textual were extracted, then CNN and LSTM layer. These provided Patient Health Questionnaire score estimates. Hierarchical Graph Attention Network (HGAT) makes use of hierarchical attention architecture that is schema-level and node-level attention to recognize fake news of online news articles. According to [35], this is effective for heterogeneous networks and no previous knowledge is required. A deep CNN known as FNDNet was developed to identify fake news across online social media by automatically learning the discriminatory features. This model extracts diverse features of news articles at each layer with superior outcomes over available models according to [36].

4   Research methodology

This section presents the conceptual models for image and textual data feature extractions and classification. The effectiveness of the models is validated accordingly.

4.1   Description of the proposed model

The complete structure of the proposed model is composed of the input, process and output components respectively as shown in Figure 3.

The classification model utilizes image and a textual dataset which are received at the input component of the model. The distinct datasets are preprocessed through the removal of noise and redundancy for the purpose of obtaining enhanced input. In case of the image dataset, the preprocessing and augmentation involve whitening of image samples, after their up sampling, its 32 by 32 crop size was selected [37]. While, the textual dataset during preprocessing were grouped into similar clusters [38], and conversion textual information into vectors or numeric values, and removal of infrequent classes using semantic and syntactic association of words by mean of natural language processing (NLP) [39]. The process component undertakes feature extraction, feature selection and classification after training procedures. These are achieved with deep learning algorithms of LSTM-RNN situated within the process component of the model. The output component provides the results of the evaluation carried out by the distinct deep classification models of LSTM-RNN using a test dataset, that is, a portion of the input dataset. The results are expressed in error rates and percentage of accuracy.

4.2   Simulation settings

The minimum specifications for the experimenting concept of textual and image-based classifications of LSTM-RNN model are presented in Table 1.

Hyperparameters training for the RNN-LSTM for textual: The procedure of choosing hyperparameters is the major aspect of the most of deep learning approaches, which can be achieved in a manual or automatic way. The goal is to minimize the cost and memory of execution. The learning algorithm makes use of the hyperparameter settings for purpose of training datasets the context-specific dataset on the model as provided in Table 2.

Hyperparameters training for LSTM-RNN for images: The complete values of the hyperparameters for the deep learning approach which are manually carried out in order to reduce memory and cost of execution. The learning algorithm make use of the hyperparameter settings for purpose of training datasets the context-specific dataset on the classifier is given in Table 3.

System Requirement	Property
Hardware	
Hard Disk Drive	68.35 GB
RAM	12.69 GB
Software	
Operating system	Python 3 Google Compute Engine backend (GPU)
Simulator	Google Collaboratory (Colab)
Classification algorithm	LSTM-RNN
Input types	Text, Image
Browser type	Google Chrome Version 91.0.4472.124 (Official Build) (64-bit)

Table 1: Simulation settings.

Hyperparameter	Value
Network model	Keras-LSTM
Number of layers	32
Embedding dimension	64
Number of dense layer(s)	1
Max number of Epochs	10
Gradient Threshold	1
Dropout	0.5
Activation	relu
Optimizers	Adam
Metrics	Accuracy
Loss function	BinaryCrossentropy
Input type	Textual

Table 2: Minimal parameters for LSTM-RNN model for text.

Hyperparameter	Value
Number of inputs	28
Number of steps	28
Classifier	RNN-LSTM
Activation function	Relu
Loss function	CrossEntropyLoss
Number of neurons	150
Number of Epochs	10
Number of outputs	10
Optimizer	adam
Learning rate	0.001
Metrics	Accuracy
Input	Grayscale images

Table 3: Minimal parameters for LSTM-RNN for images.

The steps for proof-of-concept in both RNNs-based classification cases are presented as follows:
i.   The data input is collected from the standard textual and image repository (TensorFlow).
ii. The text and image datasets are preprocessed by removing noise and irrelevant items.
iii. The minimal experimental parameters are used to construct the two RNN models.
iv. The textual and image features are extracted using the deep neural networks built in the previous step (iii) to generate the training sets.
v.   The textual and visual features are used to train the RNN classifiers in both cases.
vi. The outcomes of classification are generated and analyzed.

4.3   Dataset Description

The image dataset is a large set of Fashion MNIST images comprising of 221.83 MiB size, 70000 items with nine labels namely: T-shirt/top, Trouser, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag, and Ankle boot were downloaded from TensorFlow image catalogue. Similar dataset was utilized in the work by [40].

Again, the textual dataset is a large IMDB movie reviews collected from TensorFlow composed of customers' reviews of positive and negative sentiments, which is 32.06 MiB size. It is made up of 10000 buffer size and 64 batch size categorized as texts and labels. Again, the training and testing components were 70% to 30% for training and testing respectively. Similar textual dataset was used to measure the performance of models in previous studies [20,36].

In both cases, the datasets were split into 70% and 30% for training and testing of classification models respectively.

Evaluation parameters: The performances of the two classification approaches are computed using accuracy metric as represented in Equations 3 and 4 [29]:

$$Accuracy = \frac{TN + TP}{FN + FP + TN + TP} \quad (3)$$

$$Elaspsed\ Time = ET - ST \qquad (4)$$

Where, TP, FP, TN, FN, ET and ST depict the following:

True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN), Execution end time, and ST is execution start time.

4.4   Results and Discussion

The progress of training and validation accuracy for the long short-term memory (LSTM) network structure of RNN are shown in Figure 4.

In Figure 4, there are strong correlations in the accuracy and validation accuracy during the training process of the dataset especially at epoch 1 to 10. Similarly, the loss distributions over the textual data deep learning are relatively diminished for loss and validation loss from epoch 1 to 10 respectively. The summary of LSTM-RNN deep leaning outcomes after training and validation of textual data are presented in Table 4.

From Table 2, the LSTM classified 1000 words into 5 classes after training and identifying the hidden structures with considerable effectiveness at 85.69% accuracy. However, the time for training and validation of the deep learning model provided by RNN is relatively ineffective at 18.25 mins.

Similarly, the LSTM-RNN deep learning structure of the CNN model is used for training the image features selected are contained in Table 5.

From Table 3, the process of classifying 60000 image features with 9 classes provided 88.46% accuracy over the period of 3.49 mins. This model performs a complete training and validation procedure in relatively faster elapsed time due to improved feature selection and well-defined hidden data patterns in the images. The comparison with existing deep learning models for textual and image classifications is presented in Table 6

From Table 6, the dataset nature and network architecture strongly impact on the deep learning classification accuracy and speed. Therefore, deep learning offered by LSTM-RNN for images is better for speed and accuracy of classification outcomes against textual due to the complicated structure of network and data formation in spectral and spatial domains. When compared to existing approaches using the same datasets, majorly text has higher accuracy than image classified due to domain of application, complexity of hidden patterns in media, and preprocessing performed prior to classification tasks.

5   Conclusions and future works

This paper demonstrated the effectiveness of deep learning approach of LSTM-RNN, which is a derivative of convolutional neural networks for image and textual datasets classification tasks. In both cases, LSTM-RNN was used to extract information concerning the local structure of the data by means of the application of multiple filters with distinct dimensions in the images and textual before classifying. Though, the process of
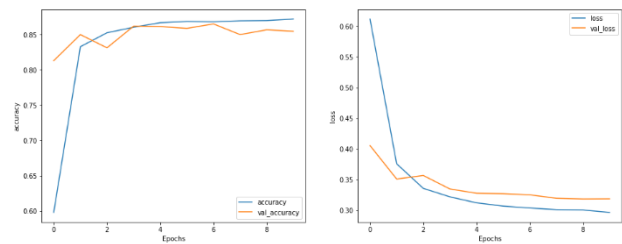

Figure 4: The training process for accuracy and loss functions of textual learning.

S/N.	Parameter	Value
1.	Dimension of embedding	171
2.	Number of classes	5
3.	Number of vocabularies	1000
4.	Accuracy	0.8569
5.	Loss	0.3169
6.	Elapsed Time	18.25 mins

Table 4: Textual deep learning outcomes.

S/N.	Parameter	Value
1.	Number of image features	60000
2.	Number of classes	9
3.	Accuracy	0.9650
4.	Loss	0.1250
5.	Elapsed Time	3.49 mins

Table 5: .

Classification model	Accuracy	Elapsed time	Dataset
LSTM-RNN	96.50%	3.49 mins	TensorFlow
LSTM-RNN	85.10%	18.25 mins	TensorFlow
Modified CNN	74.6%	Unspecified	TensorFlow
CNN	99.98%	Unspecified	TensorFlow
LSTM	80.5%	Unspecified	Kaggle
Deep CNN (FNDNet)	98.36%	Unspecified	Kaggle

Table 6: Comparisons of classification models of textual and image.

appropriately extracting the temporal correlation of data and dependencies in the text snippet through the word encoding approach is different from the images. The performance of image classification tasks outperformed textual classification in terms of accuracy due to quicker way of learning the patterns, knowledge, and information contained in the dataset by 96.50% to 85.69%. More so, LSTM-RNN for images performed better in terms of speed of classification by 3.49 mins to 18.25 mins. In future work, there is a need to utilize more classification models to demonstrate the concept understudied in this article. Also, there is need to effectively classify various data representations in textual and other media at higher levels of accuracy and speed. This work can be extended to other high-performance deep learning algorithms, and datasets (such as video, audio and non-English textual).

References

[1] X. Zhu, Y. Xu, H. Xu, and C. Chen, "Quaternion convolutional neural networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 631–647.

[2] L. Zou, S. Yu, T. Meng, Z. Zhang, X. Liang, and Y. Xie, "A technical review of convolutional neural network-based mammographic breast cancer diagnosis," *Comput. Math. Methods Med.*, vol. 2019, 2019.

[3] O. Abdel-Hamid, L. Deng, and D. Yu, "Exploring convolutional neural network structures and optimization techniques for speech recognition.," in *Interspeech*, 2013, vol. 11, pp. 73–75.

[4] O. L. P. Hansen *et al.*, "Species-level image classification with convolutional neural network enables insect identification from habitus images," *Ecol. Evol.*, vol. 10, no. 2, pp. 737–747, 2020.

[5] V. Suryanarayanan, B. Patra, P. Bhattacharya, C. Fufa, and C. Lee, "ScopeIt: Scoping Task Relevant Sentences in Documents," *arXiv Prepr. arXiv2003.04988*, 2020.

[6] U. Dixit, A. Mishra, A. Shukla, and R. Tiwari, "Texture classification using convolutional neural network optimized with whale optimization algorithm," *SN Appl. Sci.*, vol. 1, no. 6, pp. 1–11, 2019.

[7] X. Huang, Z. Li, C. Wang, and H. Ning, "Identifying disaster related social media for rapid response: a visual-textual fused CNN architecture," *Int. J. Digit. Earth*, 2019.

[8] F. E. F. Junior and G. G. Yen, "Particle swarm optimization of deep neural networks architectures for image classification," *Swarm Evol. Comput.*, vol. 49, pp. 62–74, 2019.

[9] V. Kudva, K. Prasad, and S. Guruvare, "Hybrid transfer learning for classification of uterine cervix images for cervical cancer screening," *J. Digit. Imaging*, vol. 33, no. 3, pp. 619–631, 2020.

[10] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv Prepr. arXiv1511.08458*, 2015.

[11] A. Hamoud and A. Humadi, "Student's success prediction model based on artificial neural networks (ANN) and a combination of feature selection methods," *J. Southwest Jiaotong Univ.*, vol. 54, no. 3, 2019.

[12] S. Sokolov, S. Vlaev, and M. Chalashkanov, "Technique for storing and automated processing of weather station data in cloud platforms," in *IOP Conference Series: Materials Science and Engineering*, 2021, vol. 1032, no. 1, p. 12021.

[13] Y. Wang, L. Wang, Y. Yang, and T. Lian, "SemSeq4FD: Integrating global semantic relationship and local sequential order to enhance text representation for fake news detection," *Expert Syst. Appl.*, vol. 166, p. 114090, 2021.

[14] M. Thoma, "Analysis and optimization of convolutional neural network architectures," *arXiv Prepr. arXiv1707.09725*, 2017.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv Prepr. arXiv1409.1556*, 2014.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[17] A. Hirose and S. Yoshida, "Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 23, no. 4, pp. 541–551, 2012.

[18] C. Trabelsi *et al.*, "Deep complex networks," *arXiv Prepr. arXiv1705.09792*, 2017.

[19] C. J. Gaudet and A. S. Maida, "Deep quaternion networks," in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–8.

[20] S. Minaee, E. Azimi, and A. Abdolrashidi, "Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models," *arXiv Prepr. arXiv1904.04206*, 2019.

[21] T. Xu *et al.*, "Multi-feature based benchmark for cervical dysplasia classification evaluation," *Pattern Recognit.*, vol. 63, pp. 468–475, 2017.

[22] W. Yu and M. Pacheco, "Impact of random weights on nonlinear system identification using convolutional neural networks," *Inf. Sci. (Ny).*, vol. 477, pp. 1–14, 2019.

[23] M. Blaivas and L. Blaivas, "Are all deep learning architectures alike for point-of-care ultrasound?: evidence from a cardiac image classification model suggests otherwise," *J. Ultrasound Med.*, vol. 39, no. 6, pp. 1187–1194, 2020.

[24] V. S. Martins, A. L. Kaleita, B. K. Gelder, H. L. F. da Silveira, and C. A. Abe, "Exploring multiscale object-based convolutional neural network (multi-OCNN) for remote sensing image classification at high spatial resolution," *ISPRS J. Photogramm. Remote Sens.*, vol. 168, pp. 56–73, 2020.

[25] S. Bera and V. K. Shrivastava, "Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification," *Int. J. Remote Sens.*, vol. 41, no. 7, pp. 2664–2683, 2020.

[26] X. Cao, J. Yao, Z. Xu, and D. Meng, "Hyperspectral image classification with convolutional neural network and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4604–4616, 2020.

[27] K. Safari, S. Prasad, and D. Labate, "A multiscale deep learning approach for high-resolution hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 1, pp. 167–171, 2020.

[28] K. Shankar, Y. Zhang, Y. Liu, L. Wu, and C.-H. Chen, "Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification," *IEEE Access*, vol. 8, pp. 118164–118173, 2020.

[29] R. J. S. Raj, S. J. Shobana, I. V. Pustokhina, D. A. Pustokhin, D. Gupta, and K. Shankar, "Optimal feature selection-based medical image classification using deep learning model in internet of medical things," *IEEE Access*, vol. 8, pp. 58006–58017, 2020.

[30] Y. Wang *et al.*, "Weak supervision for fake news detection via reinforcement learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, no. 01, pp. 516–523.

[31] E. Hossain, O. Sharif, and M. M. Hoque, "NLP-CUET@ DravidianLangTech-EACL2021: Investigating Visual and Textual Features to Identify Trolls from Multimodal Social Media Memes," *arXiv Prepr. arXiv2103.00466*, 2021.

[32] W. Li, L. Zhu, Y. Shi, K. Guo, and E. Cambria, "User reviews: Sentiment analysis using lexicon integrated two-channel CNN–LSTM family models," *Appl. Soft Comput.*, vol. 94, p. 106435, 2020.

[33] M. Rodrigues Makiuchi, T. Warnita, K. Uto, and K. Shinoda, "Multimodal fusion of bert-cnn and gated cnn representations for depression detection," in *Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop*, 2019, pp. 55–63.

[34] N. Gupta and A. S. Jalal, "Integration of textual cues for fine-grained image captioning using deep CNN and LSTM," *Neural Comput. Appl.*, vol. 32, no. 24, pp. 17899–17908, 2020.

[35] Y. Ren and J. Zhang, "Fake news detection on news-oriented heterogeneous information networks through hierarchical graph attention," *arXiv Prepr. arXiv2002.04397*, 2020.

[36] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "FNDNet–a deep convolutional neural network for fake news detection," *Cogn. Syst. Res.*, vol. 61, pp. 32–44, 2020.

[37] B. Zoph and Q. V Le, "Neural architecture search with reinforcement learning," *arXiv Prepr. arXiv1611.01578*, 2016.

[38] F. A. Ozbay and B. Alatas, "Fake news detection within online social media using supervised artificial intelligence algorithms," *Phys. A Stat. Mech. its Appl.*, vol. 540, p. 123174, 2020.

[39] Q. Umer, H. Liu, and I. Illahi, "CNN-based automatic prioritization of bug reports," *IEEE Trans. Reliab.*, vol. 69, no. 4, pp. 1341–1354, 2019.

[40] P. Yuan and R. Huang, "Integrating the device-to-device communication technology into edge computing: A case study," *Peer-to-Peer Netw. Appl.*, vol. 14, no. 2, pp. 599–608, 2021.