Journal of Theoretical and Applied Information Technology
15th July 2022. Vol.100. No 13
© 2022 Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

RECENT ADVANCES AND APPLICATIONS OF DEEP LEARNING TECHNIQUE

ADITYA DUBEY[1], AKHTAR RASOOL[2]

[1, 2] Department of Computer Science and Engineering, Maulana Azad National Institute of Technology, Bhopal, India

E-mail: [1]dubeyaditya65@gmail.com, [2]akki262@gmail.com

ABSTRACT

Deep learning is a predominant branch in machine learning, which is inspired by the operation of the human biological brain in processing information and capturing insights. Deep learning uses several layers of neurons; each layer of neurons is connected to the successive layer, which helps to provide better accuracy for complicated tasks. Machine learning evolved to deep learning, which helps to reduce the involvement of an expert. In machine learning, the performance depends on what the expert extracts manner features, but deep neural networks are self-capable for extracting features. Deep learning performs well with a large amount of data than traditional machine learning algorithms, and also deep neural networks can give better results with different kinds of unstructured data. Due to these advantages, deep learning algorithms are applied to a variety of complex tasks. With the help of deep learning, the tasks that had been said as unachievable can be solved. Now deep learning is an inevitable approach in real-world applications such as computer vision where information from the visual world is extracted, in the field of natural language processing involving analyzing and understanding human languages in its meaningful way, in the medical area for diagnosing and detection, in the forecasting of weather and other natural processes, in field of cybersecurity to provide a continuous functioning for computer systems and network from attack or harm, in field of navigation and so on. This paper describes the brief study of the real-world application problems domain with deep learning solutions.

Keywords: *Conceptual based Information Retrieval, Ontology, Semantic Search, Convolutional Neural Network, Deep Learning*

1. INTRODUCTION

Deep learning helps us in day-to-day activities like shopping, language translation, voice recognition, etc., by imitating human skills [1]. Deep learning emerged in the 1950s; deep learning did not get appreciated because of the lack of required processing power [2]. With the help of high-power graphical processing units and massive data, deep learning has begun to flourish. Deep learning uses deep neural networks having more than one hidden layer between the input and output layer. These networks are trained in a supervised or unsupervised manner [3], [4]. In supervised learning, the process network produces an output to given input data; this output is compared with actual output; hence, an error is obtained; the learning objective is to reduce this error using an algorithm and update the parameters accordingly in each iteration. Learning is continued until the model reaches the required accuracy [5]. Unsupervised learning focuses on identifying the unexplored pattern in the data when the expected outcome is unknown. Because of this, unsupervised learning is not directly suitable for regression and classification tasks [6].

For various applications, different types of deep learning models are used [7], [8]. Based on the application, a major model can be chosen such as multilayer perceptron, convolutional neural network (CNN), recurrent neural networks (RNN), Boltzmann machine (BM), and auto-encoders. The multilayer perceptron is a classical network with multiple hidden layers. It can be trained with various kinds of datasets and can be applied for classification and regression problems. CNN can learn many filters instead of handcrafting the filter, which helps identify the essential features in the data. CNN is mainly applied for image problems. RNN owns a property of maintaining an internal state or memory from previous inputs used to calculate the output from present input considering previous output. This recurrent behavior aid the network in applications involving capturing sequence information. They are

Journal of Theoretical and Applied Information Technology
15th July 2022. Vol.100. No 13
© 2022 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

widely used in language modeling and generating text. BM is a build-up of nodes connecting each other, which can cause stochastic conclusions regarding whether actuate or not. They can be used to find the solution for combinatorial optimization problems. Auto-encoders consist of the encoder, code, and decoder. The encoder is fed with input that compresses it and produces a code; later, the encoder is decompressed or reconstructed by the encoder. Auto-encoders help in dimensionality reduction by producing a compressed representation through unsupervised learning. Auto-encoders are used in data compression and retrieval task.

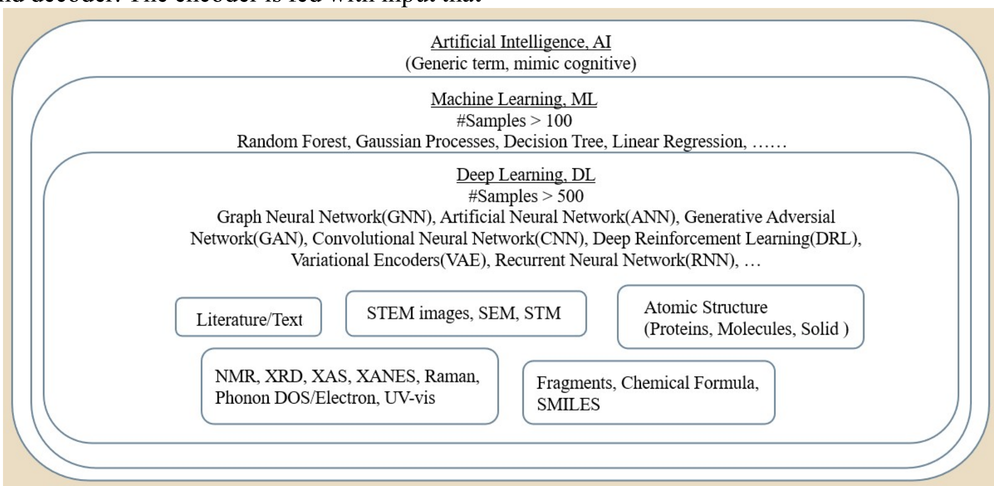

Figure 1: Applications of AI, ML, and Deep Learning in Material Science

Figure1 represents the hierarchy of AI, ML and deep learning in material science field. Here, the term SMILES represents simplified molecular input line entry system. Some image modalities include scanning transmission electron microscopy (STEM), scanning probe tunneling microscopy (STM), and scanning electron microscopy (SEM). In figure, X-ray Diffraction (XRD), X-ray Absorption Near Edge Spectroscopy (XANES), and X-ray Absorption Spectroscopy (XAS) some spectroscopy models Applications of deep learning extended to diverse fields. Computer vision is a branch of computer science that helps machines understand visual data such as video and images. Deep learning emerges as a vital tool for computer vision problems such as image classification, object detection, image reconstruction, face recognition, image synthesis, etc. [9]. Deep learning networks such as convolutional neural networks can better predict performance using adequate computing resources and data. LeNet-5, AlexNet, VGGNet, GoogLeNet, and ResNet are competing for CNN architectures or computer vision problems [10]. Signal processing is a sub-branch of electrical engineering that involves synthesizing, altering, and modifying signals obtained from a transducer. CNN architectures and recurrent network such as LSTM provides better results than traditional signal processing methods [11]. Deep learning made a significant impact on the healthcare and medical field. Because of its remarkable performance in analyzing images such as X-ray or MRI, it can provide excellent assistance to healthcare experts [12].

Many kinds of research are going presently on pandemic covid-19 in vaccine development and diagnosis with deep learning. The process of forming future predictions by evaluating available previous and present data is known as forecasting. Deep learning enables a better platform for forecasting, which helps humans in planning and decision-making. The traditional techniques in time series forecasting are based on the linear method, but deep learning methods can be trained to develop a complex mapping between input and output. CNN and LSTM play a significant role in time series forecasting. Navigation is a field of study that consists of controlling and auditing the motion of an aircraft or automobile from its starting position to the destination. Self-driving cars are getting a lot of attention, showing increased demand in this field [13]. In this digital world, person faces many cyber-attacks and cybercrimes; cybersecurity is a field that consists of technologies that protect computers, networks, and valuable data from unauthorized use and damage [14]. Deep learning-based malware detection, spam detection, network traffic analysis shows excellent results. The remaining part of this paper consists of a summary of recent deep learning approaches used to solve real-world problems. In each approach, we point out the challenge intended to be solved, the methodology followed with input and output. Figure 2 represents different applications in today's era of IOT and big data.

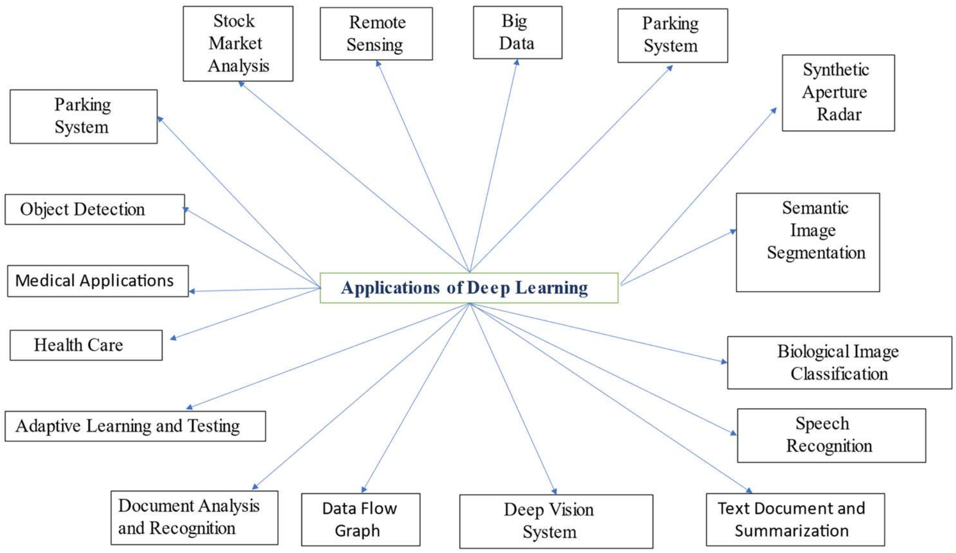

Figure 2: Different Applications of Deep Learning

2. CONTRIBUTIONS OF DEEP LEARNING

2.1 Medical

Conventional disease detection approaches are not convenient in the real-time analysis due to their complex procedures. Bhaskar et al. (2019) proposed a new method for kidney disease detection in the article "A Deep Learning-based System for Automated Sensing of Chronic Kidney Disease" incorporating CNN and SVM classifier [15]. For detection, saliva samples are taken instead of blood because a disease-identifying particle can also be found in saliva of that person and their easy availability. The levels of urea and creatinine in serum can be measured to monitor kidney functioning. Recent studies had revealed that the level of urea in saliva and serum are related [16]. An increased urea level in saliva can disclose the abnormal operation of the kidney. Here urease enzymes are used to hydrolysis of urea. This enzymatic reaction results in the production of ammonia gas; then, the level of ammonia is measured using a semiconductor-based sensor. The hardware setup of the sensing module consists of an Arduino board, an MQ series ammonia gas sensor, and a crafted sensing chamber that enclose the sensor. During testing, the sample is dropped to the input opening to the gas sensing chamber. The electrical conductivity of the sensor varies with the level of produced ammonia gas. This variation in conductivity is provided in terms of analog voltage to the output of the sensor unit. A CNN- SVM-based algorithm is used to extract the features from the output of the sensor unit automatically. Here the CNN is modified to process 1D signal. Using the kernel, convolution is performed on the sensor output signal. The resulting feature map is described in equation (1)

$$ci(n) = \sum_{m=-p}^{p} x(m+1)k(q-m+1) \qquad (1)$$

x is the input signal having length p, and k is the kernel function having length q, convolution operation followed by pooling function which down-samples the feature map. Convolution and pooling are performed repeatedly to capture the reduced feature map. Here a five-layered CNN is adopted, and Gaussian kernel is used for convolution. Instead of a fully connected neural network classifier, an SVM classifier is proposed in this work. RBF (Radius Basis Function) kernel-based SVM classifier is used for the classification of features. This method can classify the samples with 98.04% accuracy.

Hookworms are blood-consuming parasites commonly seen inside the human intestine. If it has gone undiagnosed, it can seriously impact health, such as anemia, leading to heart failure. Reports show that it has infected more than 600 million people worldwide [17]. He et al., (2018) proposed the first deep hookworm detection framework in the article "Hookworm Detection in Wireless Capsule Endoscopy Images with Deep Learning" for wireless capsule endoscopy images [18]. The framework consists of two CNN networks; one is an edge extraction network, and the other is for hookworm classification. The intestinal images are collected using wireless capsule endoscopy, which the patient has swallowed. The device is capable of taking around 50,000 images of the intestine. It is tedious for a specialist to analyze the affected area, which can take hours manually; these images also affect poor resolution, light conditions, and orientation. The edge pooling layer is the central part of this network with a two-level fusion architecture. To identify the tubular structure of hookworm, an edge map produced from an edge extraction network is applied to a multi-scale dual matched filter (MDMF). The edge map generation holistically nested edge detection (HED) method shows remarkable performance in edge detection. The HED consist of a cascade of deep neural network with multiple side outputs. Five layers of side output are selected here. The side output at lower layers holds more edge details, so 2nd and 3rd layers side output are collected for edge detection.

MDMF is Gaussian shaped template to identify the tubular region of hookworm because hookworm has a cross-section of Gaussian shape. Hence convolving the image with filter results in a higher response. So, the MDMF focuses on the edge of the hookworm and suppresses the noise. Using edge pooling and regularized edge pooling, the tubular maps are integrated into the classification network. Nine sparsely connected inception models are implemented in the hookworm classification network to account for overfitting and computational resources. Max pooling layers are used before edge pooling operation to match the spatial dimension of the tubular region map and hookworm classification feature map. Above it, the average pooling layer is inserted for parameter reduction, further connected to a fully connected and soft-max network. Experiments are conducted on the WCE image dataset having 440K images show the potency of this framework in clinical application.

2.2 Forecasting

Household load forecasting is an essential step of planning. The uncertainty of load profiles and their unstable behavior load forecasting is a challenging task. Traditional methods such as load aggregation, customer classification, and spectral analysis eliminate these uncertainties. For the first time, Shi et al., (2018) in their article “Deep Learning for Household Load Forecasting A Novel Pooling Deep RNN,” solve this problem [19]. The method consists of 2 steps-initial steps for load profile pooling, and the final step is short STFL with deep learning. The pooling strategy was implemented to account for the significant challenge in STLF, such as the inherent high uncertainty and the overfitting. Overfitting is caused due to the abundant layers of neural networks and when the load profile data inadequate. The pooling step boosts the data volume and reduces overfitting.

By pooling the customers' load profiles jointly can improve the load dataset diversity. Daily load profiles are collected for every 30 minutes from smart meters in 48 value points. Long vector is used in pooling which has merged load profiles of continuous dates. Load profile pooling consists of 3 steps. Initially, a dummy variable is assigned to the demand data to identify its customer id, then training and testing sets are formed by splitting the demand data of each customer. Finally, training and testing pools for each customer are formed by batching together the training and testing set. Later training and testing are conducted. For training, DRNN is initiated firstly and trained the network till load loss gets converged. At each training epoch, a batch with size B is fed with dimension B x (sequence size), and the output obtained is B x (output sequence size). Batch size B is increased for higher epochs so that the optimum point can be achieved rapidly in early-stage and better performance at higher epochs. The loss function is framed as (2) where y_p is predicted output during training and y_a is the actual output.

$$Loss = \sqrt[2]{\frac{1}{O}\frac{1}{B}\sum_{I=1}^{O}\sum_{J=1}^{B}(y_p - y_a)^2} \quad (2)$$

The deep learning program implementation is done in TensorFlow. The stages involved in program implementation are data cleaning and preprocessing, data pooling and sampling, and finally, training and benchmarking. This method performs better than state-of-the-art techniques such as ARIMA, SVR, and RNN.

2.3 Navigation

Due to its unpredictable and dynamic environment, acoustic target recognition in the undersea is a troublesome task. Traditional machine learning methods involve a large human endeavor from feature extraction and selection to classifier building. Traditional methods show poor generalization when handling complex data in massive amounts. For navigation of underwater vehicles, a new approach is proposed in “A New Cooperative Deep Learning Method for Underwater Acoustic Target Recognition” [14]. Here a cooperative deep learning model is used to classify ship radiated noise. This method combines DLSTM (Deep Long Short-Term Memory) and DAE (Deep Auto Encoder) networks. The processing of the time-series signal is done with the help of the LSTM network, and the DAE contributes to the input data compression. This method consists of two stages: first, LSTM based DAE network is created with an encoder and decoder.

The encoder is employed to compress the data to the reduced dimension, which has one or more layers of LSTM. The decoder's reconstruction of input data consists of at least one layer of LSTM. LSTM based DAE network is utilized to pre-train the DLSTM model in the DAE network. The network learns by effectively reconstructing the data close to input data through unsupervised learning. This network enhances the process of extracting the inherent characteristics of input data by removing redundant, irrelevant information. A cooperative DLSTM network is constructed in the next stage with the previous pre-trained encoder layer and a softmax layer. The new collaborative network learns the variability and fundamental characters with the help of LSTM. Experiments are done to evaluate the proposed method with the datasets obtained from ship radiated noise data and Oceans Network Canada Observatory. The background noise data are collected when ships are absent within a 5 Km of radius from the hydrophone. The frequency of sampling the noise data is 32 kHz. The total duration of the dataset used is 21.7 hours which splits into 13.7 hours for training and 8 hours for testing. This method has 90% accuracy, which is higher than the DAE network and DLSTM network, which are taken for comparison.

Landing is the most challenging phase of a flight. Human have witnessed a lot of landing accidents due to bad weather conditions or any hardware malfunction. Unmanned aerial vehicles (UAV) got popular in surveying, military applications, urban planning, inspection, and rescue. UAVs are automatically controlled by a processing system that analyses the aerial image obtained from onboard cameras. For safe landing, vision-based accurate runway detection systems should be implemented. Akbar et al., (2019) proposed runway detection and localization methods in their article [20]. In the first stage, a classifier is built to detect the runway in the aerial image; then, the detected runways are localized with the help of deep learning and conventional approaches. It is sufficient to detect binary runway classifications, but it is not a good approach, so a remote sensing dataset of multiple classes is used in this paper. Input images are resized to 224×224 and mean normalized. For feature extraction, three classifiers are built from VGGNet, ResNet, DenseNet with a softmax layer at the output. Based on the performance, ResNet is selected as the best among them. CNN algorithms and line detection algorithms such as Hough transform and linear segment detector are used for runway localization. For segmentation purposes masked R-CNN model, which is pre-trained in the coco dataset, is fine-tuned with the runway dataset is utilized. Datasets used are taken from Label Me, which are 700 images of the class runway. These runways in the images are labeled by filling polygons. For experimentations, self-made 100 images from google earth are taken in to train validation and test. This method has achieved a reasonable IOU of 0.8 and uses deep learning methods to eliminate the need for handcrafted features.

2.4 Natural Language Processing

A huge amount of text data is available on the web, so identifying the useful or relevant text data from the huge document is very important. Parvathi & Jyothis (2018) had proposed a method using CNN to identify the relevant text from the text document [21]. The data is first given to preprocessing stage, which involves steaming and spam detection. Natural Language Toolkit (NLTK) is used tokenization and stemming. In the next step, feature extraction takes place. Obtained relevant data is classified using a neural network classifier to get the result. For the classification of relevant text documents, 20 documents and four classes (education, goodbye, sandwich, and greetings) are chosen. The proposed approach is proving an approximate accuracy near to unity.

Bai, (2018) proposed an improved method for text classification which utilizes convolution and combing long short-term memory (LSTM) with an attention mechanism [22]. In the first stage, using convolutional layers, preliminary features are extracted. Different convolutional filters are used to learn local information from the text data. The three gates present in LSTM cells enables it to learn long time-dependent sequence. The attention mechanism produces attention probability distribution which consists of deep features. These deep features obtained from LSTM and attention network is given to the softmax layer to perform classification. Figure 3 shows the steps proposed. The data set taken for simulation is the Chinese Opinion Analysis Evaluation Microblog data set from which 9423 microblogs are used for training and 2426 microblogs are taken for testing. They have verified the performance of the proposed algorithm by comparing it with CNN, RNN, and LSTM. It shows that LSTM performance is better. The LSTM is combined with an attention mechanism to prioritize the critical input from the data.

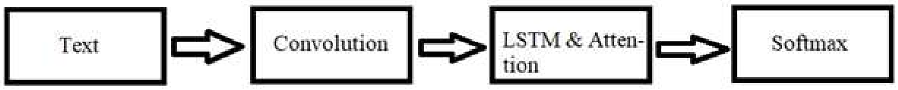

Figure 3: Text Classification Block Diagram

Li et al., (2018) studies text classification by using the Bi-LSTM-CNN method [23]. The model comprises a Bi-LSTM layer consisting of a word vector with right and left context, A local and global feature, and finally, a softmax layer. Word2Vec is used to train the word vector. Continuous bag of words adopted for training the word vector. The bidirectional LSTM layer considers the forward's semantics and also reverse order semantics. Convolution is used to combine right and left contexts to obtain an expression for the first context word. This new expression, tanh activation function, is applied then a Max pooling layer is used to convert the text to fixed-length vectors. The data set for training and testing is a subset of THUCNews. 65000 corpora are present in this corpus from which 50000 is used for training, 5000 is taken for verification, and 10000 is used for testing. This method is compared with other models such as SVM, CNN and LSTM. The proposed model has improved the efficiency of classification by 0.84%.

2.5 Signal Processing

Electrocardiogram (ECG) consists of noise signals, so it is essential to effectively de-noise for processing ECG. Arsene et al., (2019) had presented two deep learning-based de-noising models with a standard wavelet technique [24]. The first deep learning model is CNN-based, and the other is LSTM based model compared with a wavelet technique. The CNN model investigated here comprises six 2 D convolutional layers, each consist of 36 filters having kernel dimension 19x1 with input layer dimension 30000x1x1. After convolutional layers, a batch normalization layer is used; its output is fed into a rectified linear unit (RELU), and the average pooling layer reduces dimensionality. A fully connected layer is used before the output layer. The regression layer is used as the output layer where the ECG de-noised signals are obtained. The second model is made of LSTM layers; it comprises two layers of LSTM with 140 hidden units in each layer with sequence input layer with dimension 30000x1. RELU activation function is used in these hidden units. A fully connected layer and a regression output layer are placed at the output. The last model for de-noising is wavelet-based on the empirical Bayesian method for comparison; as per the results, the CNN-based model can eliminate high-level noise from the ECG. In comparison, CNN provides high quality and time-saving results than the LSTM based model.

Pak et al., (2019) proposed a method for calculating accurate noise evaluation units to compensate for the annoyance of aircraft noise using the CNN model [25]. The dataset consists of manually labeled 100,000 mp3 files of noise from five different noise monitors placed on the roof of buildings near Jeju International airport in Korea. Mel-frequency cepstral coefficients (MFCC) extract feature vectors from the noise signal. 2 types of data are required: the obtained feature vector, and the other is the mel-spectrogram image. Librosa module in python is used for converting the sound signal to mel-spectrogram. After conversion, the data is split into train and test in ratio 7:3. The CNN model consists of 5 convolutional layers, three pooling layers, and a fully connected layer. Convolutional layer has RELU activation function with 3x3 sized kernel, 1st dense layer after flatten layer having an activation of tanh with a dropout of 0.25 and the final layer having softmax activation function. This method is cost-effective when the MFCC feature vectors are applied for input; the model gives an accuracy of 99.84% with an FP rate of 0.16 and an FN rate of 0%.

2.6 Autonomous Vehicle

For the safe functioning of autonomous vehicles, fault detection is essential. Failure to accurately and immediately detect faults may lead to the breakdown of the vehicle or even accidents. Ren et al., (2019) proposed a general framework is developed for Fault detection in the system dynamics model [26]. To build a system model, they have collected input-output data from a multi-wheeled autonomous vehicle. To generate the dynamic system model, they have used the MATLAB System identification toolbox. The proposed framework consists of a system model, continuous wavelet transforms (CWT), and deep neural network. The input to the continuous wavelet transform is generated from the system model by supplying input. The CWT produces two-dimensional images from one-dimensional signals. Fig 4 shows the fault detection framework. These images having size (512×512) are given input to the ten layered deep neural networks. They have tested the proposed deep neural network, and wavelet transform technique giving separate 500 defected and normal images and established this method can effectively use for fault detection.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

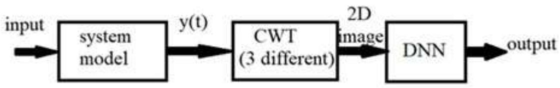

Figure 4: Fault Detection Framework

Sanil et al., (2020) had identified obstacles and their avoidance using CNN [27]. The miniature self-driving car consists of 2 DC motors which the L293d motor driver controls. They perform acceleration and steering movements of wheels. For controlling the car movement and processing the information, Raspberry pi 3B is used. Figure 5 shows the operational framework of designing a self-driving car, in which a camera connected in front of the car is responsible for capturing images for training. Input images and direction commands are collected and trained in the neural network model. The CNN model consists of 2 convolutional layers. 1st convolutional layer contains 32 filters of dimension 3×3, then a dropout layer of the dropout rate of 20% is applied. Next convolutional layer of 64 filters having dimension 3×3. Then a max-pooling layer of size 2×2 is applied with a dropout layer of 20%. The network is flattened to 1dimensional and applied to a dense layer having 128 neurons; the activation function used in this layer is RELU. The final output dense layer consists of 4 units with softmax activation. The output indicates the vehicle's direction of steering and movement (forward, stop, left, and right). During training, an accuracy of 86.6% is achieved. Car driving pattern successfully detects and avoid obstruction, and it requires more data and training for new scenarios.

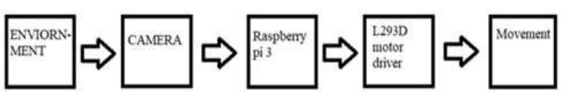

Figure 5: Operational Framework of Self-Driving Car

2.7 Computer Vision

Facial expression is the natural way of communicating emotions. Through facial expression recognition (FER), the deep feelings of a person or a crowd can be understood. Zhang et al., (2019) proposed a method for facial expression recognition using a hybrid deep learning method on videos [28]. Figure 6, shows the FER block diagram. Inputs to the CNN are fixed-sized, so the video sample is divided into segments. Two stream CNN are used for special temporal Feature extraction. Inputs to the temporal CNN are optical flow images constructed from consecutive frames. Input into the special CNN is cropped spatial image. Both CNN in the network are fine-tuned from the pre-trained VGG16 network, and they are applied to the deep belief network. A two-step strategy is adopted to train the RBM model. The method is experimented with using three public facial expression data sets. The output results show that the method outperforms the state of the arts on the experimental data sets.

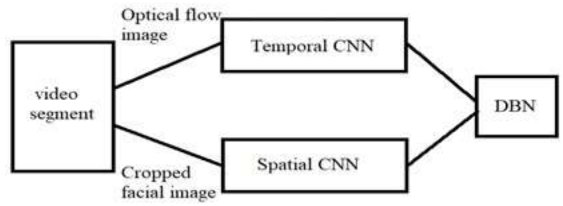

Figure 6: Block Diagram of FER Method

Image feature matching is a primary feature used in computer vision for image retrieval and object tracking. Irrespective of geometric transformation or illumination, the image should be appropriately matched. Liu et al., (2018) adopted CNN based deep learning model for image feature point matching [29]. The adopted CNN structure has seven convolutional layers. The final layer having a kernel size of 8x8, and the remaining kernel size are 3x3. The batch normalization layer separates them. RELU activation function is used in convolutional layers. Batch normalization is used for accelerating convergence. A dropout layer is placed on the sixth layer of RELU to prevent overfitting. The model is trained using the triplet loss function. Feature-based matching consists of two parts which are feature detection and matching. SIFT feature points are chosen to get local feature points because SIFT Features are intolerant to variation in Rotation scale or brightness. For achieving feature point description, a deep convolutional neural network model is used. The image patch gives us input to the trained model and obtained 128-dimensional feature descriptions, representing the description of the image feature points. The information obtained from the feature description is used to build KD-tree. KD-tree is used to find corresponding feature points. Finally, from comparison results with other methods shows the proposed method is intolerant to geometry, illumination, and appearance.

2.8 Speech Recognition

Speech recognition is very troublesome under noisy environments, and visual speech recognition can improve the quality of speech recognition technologies. Observing and understanding lips, eyebrows, and face movement can pave the way to recognizing the words. Mudaliar et al., (2020) proposed a deep learning approach for visual speech recognition [30]. Figure 7 depicts the operational framework of speech recognition. The model consists of 3D convolutional layers of ResNet architecture as encoders and Gated Recurrent Unit (GRU) for decoding. Input is a video signal, so the

Journal of Theoretical and Applied Information Technology
15th July 2022. Vol.100. No 13
© 2022 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Resnet input layer is modified by adding a 3D convolutional layer, and the remaining layers are kept as 2D convolutional layers. This part is the encoding part, where the features from the video are extracted. The Gated Recurrent Unit decodes the features, and each frame is extracted from video to find the ROI or the region of the lip using OpenCV, then the frames containing the ROI are added to an array. This array becomes the input to the model. The encoder and decoder were trained separately and gave a training accuracy of 99% and validation accuracy of 90%. The model performance on frames with facial hair features is not good, but this limitation can be eliminated by modifying the dataset.

Figure 7: Block Diagram of Speech Recognition System

3. CONCLUSION & FUTURE WORK

Applications of deep learning methods are growing to wider fields of human problems [31], [32], [33], [34], [35]. CNN and RNN are widely popular and have been used in most of the recent deep learning applications. Automatic feature extraction from input data by different CNN models has given many advantages for developing artificial intelligence. Many pre-trained CNN architectures are available, so these networks are fine-tuned in the particular dataset for each of the specific uses. CNN can understand the pattern in spatial and pixel information present in images, so these are widely used in image and video datasets. The recurrent-based neural network helps to understand the pattern in time-varying data such as audio signals. Recurrent neural networks having a recurrent weight that helps retain previous information in the network give a memory effect. Due to the vanishing gradient problem, LSTM and GRUs are considered for long sequences. Gates's presence in them enables them to learn information selectively. Recurrent networks are used for de-noising of signals which outperform traditional methods. The major limitation of deep learning supervised methods is the lack of sufficient labeled data. Many of the literature have used custom-made datasets by manually labeling them.

Each real-life application dataset has distinct properties, regardless of this, deep learning performance is more accurate. Besides of designing a new algorithm, some changes in the steps of deep learning may result in the improved performance. Selection of optimal parameters is also a major concern for deep learning. Time and space complexity are also major issues, while using the deep learning approach. Properly selecting the hierarchy of layers and supervising the learning process are most significant components in developing a successful deep learning model. However, the improved performance of deep learning can bear this limitation.

REFERENCES:

[1] Y. Jiang, "Semantically-enhanced information retrieval using multiple knowledge sources", *Cluster Computing*, Vol. 23, No. 5, 2020, pp. 1-20.

[2] A. Dubey, and A. Rasool, "Data Mining based Handling Missing Data", *Third International conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud) (I-SMAC), Palladam,* India, 2019, pp. 483-489.

[3] A. Dubey and A. Rasool, "Time series missing value prediction: Algorithms and applications", *International Conference on Information, Communication and Computing Technology*, 2020, pp. 21–36.

[4] A. Dubey and A. Rasool, "Clustering-based hybrid approach for multivariate missing data imputation", *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 11, No.11, 2020, pp. 710–714.

[5] A. Yadav, A. Dubey, A. Rasool and N. Khare, "Data Mining Based Imputation Techniques to Handle Missing Values in Gene Expressed Dataset", *International Journal of Engineering Trends and Technology*, Vol. 69, No. 9, 2021, pp. 242-250.

[6] V. K. Gond, A. Dubey and A. Rasool, "A Survey of Machine Learning-Based Approaches for Missing Value Imputation", *Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2021, pp. 841-846.

[7] A. Dubey and A. Rasool, "Efficient technique of microarray missing data imputation using clustering and weighted nearest neighbour", *Scientific Reports*, Vol. 11, 2021, pp. 1-12.

[8] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN", *International Conference on Computer Vision*, 2017, pp. 2980-2988.

[9] N. O. Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, "Deep Learning vs. Traditional Computer Vision", *Computer Vision Conference (CVC)*, 2019, pp. 128-144.

[10] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, “A Survey of the Recent Architectures of Deep Convolutional Neural Networks”, *Artificial Intelligence Review*, 2020, pp. 5455-5516.

[11] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S. Y. Chang, and T. Sainath, “Deep Learning for Audio Signal Processing”, *IEEE Journal of Selected Topics in Signal Processing*, Vol. 13, No. 2, 2019, pp. 206-219.

[12] A. A. Aiad, R. Duwairi, and M. Fraiha, “Survey: Deep Learning Concepts and Techniques for Electronic Health Record”, *IEEE/ACS International Conference on Computer Systems and Applications (AICCSA)*, 2018, pp. 1-5.

[13] S. Kuutti, R. Bowden, Y. Jin, P. Barber, and S. Fallah, “A Survey of Deep Learning Applications to Autonomous Vehicle Control”, *IEEE Transactions on Intelligent Transportation Systems*, 2019, pp. 1-22.

[14] H. Yang, G. Xu, S. Yi, and Y. Li, “A New Cooperative Deep Learning Method for Underwater Acoustic Target Recognition”, *OCEANS 2019 - Marseille*, 2019, pp. 1-4.

[15] N. Bhaskar, and M. Suchetha, “A Deep Learning-based System for Automated Sensing of Chronic Kidney”, *IEEE Sensors Letters*, Vol. 3, No. 10, 2019, pp. 1-4.

[16] P. Celec, L. Tothova, K. Sebekova, L. Podracka, and P. Boor, “Salivary markers of kidney function-potentials and limitations”, *Clinica Chimica Acta*, 453, 2016, pp. 28-37.

[17] A. Fenwick, “The global burden of neglected tropical diseases”, *Public health*, Vol. 126, No. 3, 2012, pp. 233–-236.

[18] J. Y. He, X. Wu, Y. G. Jiang, Q. Peng, and R. Jain, “Hookworm Detection in Wireless Capsule Endoscopy Images with Deep Learning”, *IEEE Transactions on image processing*, Vol. 27, No. 5, 2018, pp. 2379-2392.

[19] H. Shi, M. Xu, and R. Li, “Deep Learning for Household Load Forecasting—A Novel Pooling Deep RNN”, *IEEE Transactions on Smart Grid*, Vol. 9, No. 5, 2018, pp. 5271-5280.

[20] J. Akbar, M. Shahzad, M. I. Malik, A. Ul-Hasan, and F. Shafait, “Runway Detection and Localization in Aerial Images Using Deep Learning”, *Digital Image Computing: Techniques and Applications (DICTA)*, 2019, pp. 1-19.

[21] P. Parvathi, and T. S. Jyothis, “Identifying Relevant Text from Text Document Using Deep Learning”, *International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET)*, 2018, pp. 1-4.

[22] X. Bai, “Text classification based on LSTM and attention”, *International Conference on Digital Information Management (ICDIM)*, 2018, pp. 29-32.

[23] C. Li, G. Zhan, and Z. Li, “News Text Classification Based on Improved Bi-LSTM-CNN”, *International Conference on Information Technology in Medicine and Education*, 2018, pp. 890-893.

[24] C. T. C. Arsene, R. Hankins, and H. Yin, “Deep Learning Models for Denoising ECG Signals”, *European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1-5.

[25] J. W. Pak, and M. K. Kim, “Convolutional Neural Network Approach for Aircraft Noise Detection”, *International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2019, pp. 430-434.

[26] J. Ren, R. Ren, M. Green, and X. Huang, “A Deep Learning Method for Fault Detection of Autonomous Vehicles”, *International Conference on Computer Science & Education (ICCSE)*, 2019, pp. 749-754.

[27] N. Sanil, P. A. N. Venkat, V. Rakesh, R. Mallapur, and M. R. Ahmed, “Deep Learning Techniques for Obstacle Detection and Avoidance in Driverless Cars”, *International Conference on Artificial Intelligence and Signal Processing (AISP)*, 2020, pp. 1-4.

[28] S. Zhang, X. Pan, Y. Cui, X. Zhao, and L. Liu, “Learning Affective Video Features for Facial Expression Recognition via Hybrid Deep Learning”, *IEEE Access*, Vol. 7, 2019, pp. 32297-32304.

[29] Y. Liu, and X. Xu, “Image Feature Matching Based on Deep Learning”, *International Conference on Computer and Communications*, 2018, pp. 1752-1756.

[30] N. K. Mudaliar, K. Hegde, A. Ramesh, and V. Patil, “Visual Speech Recognition: A Deep Learning Approach”, *International Conference* on Communication and Electronics Systems (ICCES 2020), 2020, pp. 1218-1221.

[31] M. K. Sharma, P. Kumar, A. Rasool, A. Dubey, and V. K. Mahto, "Classification of Actual and Fake News in Pandemic", *International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2021, pp. 1168-1174.

© 2022 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[32] P. Vyas, F. Sharma, A. Rasool, and A. Dubey, "Supervised Multimodal Emotion Analysis of Violence on Doctors Tweets", *International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2021, pp. 962-967.

[33] A. M. A, A. Rasool, A. Dubey, and B. N. Roy, "Optimized Weighted Samples Based Semi-Supervised Learning", *International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2021, pp. 1311-1318.

[34] V. S. Charan, A. Rasool, and A. Dubey, "Stock Closing Price Forecasting using Machine Learning Models", *International Conference for Advancement in Technology (ICONAT)*, 2022, pp. 1-7.

[35] A. Soni, A. Rasool, A. Dubey, and N. Khare, "Data Mining based Dimensionality Reduction Techniques", *International Conference for Advancement in Technology (ICONAT)*, 2022, pp. 1-8.