Int. J. Sci. R. Tech., 2026 3(9)
A Multidisciplinary peer-reviewed Journal
www.ijsrtjournal.com [ISSN: 2394-7063]

Original Article

LST-Resnet: Optimization-Based Hybrid Network Architecture For An Effective Question And Answering System With Bar Chart Images

Neelu Verma*

Computer Science and Engineering, Indian Institute of Technology Jodhpur

ABSTRACT

In the modern period, people in various fields frequently use different types of charts to express their opinions and convey information by analyzing the data. Moreover, charts help to provide accurate answers about the data. Specifically, bar charts are an effective technique to convey numeric information, but today's algorithms have less potential to interpret such data. Due to the inability to process different types of charts at the same time, automatic data extraction from charts remains a challenging task. However, deep learning techniques are the one and only way to cope with such issues. Hence, this research presents a hybrid network architecture named LST-ResNet to achieve an effective Question Answering (QA) system using bar chart images. The hybrid network structure is derived by the integration of Deep Long Short Term Memory (Deep LSTM) and Deep Residual Network (Deep ResNet). Moreover, the residual layer of the hybrid network architecture is trained using the proposed Ant Lion Political Optimizer (ALPO), which is derived from the combination of Ant Lion Optimization (ALO) and Political Optimizer (PO), respectively. The proposed ALPO-hybrid network achieved higher performance in terms of accuracy, precision, recall, and F1-measure with the values of 87.231, 0.860, 0.826, and 0.866, respectively.

Keywords: Question Answering System (QA), Bar chart images, Ant Lion Optimization (ALO), Political optimizer (PO), Deep Residual Network (Deep ResNet).

INTRODUCTION

The general quote among business people is “Data everywhere, information nowhere”. It is commonly known to all people that all reports and presentations related to the company’s projects are in the form of folders, and they are usually structured with eye-popping charts that consist of images in the form of formidable data. However, achieving the relevant data from such charts still remains a challenging issue [18] [25]. In order to cope with such a limitation, an automatic system is developed to extract useful information from charts that could offer great advantages in the area of knowledge management happening inside the organization. This extracted knowledge can be merged with various datasets in order to improve the business value [3]. Question Answering (QA) is a usual task and is utilized to solve various issues. Such issues are very difficult to solve because of the presence of several complexities in natural language. However, a QA system utilizes knowledge that consists of different facts and captures the suitable answer from the knowledge. In most of the conventional QA models, the structured knowledge graph and raw text are employed as their knowledge. Unfortunately, the corpora of raw text are very difficult to understand for machines. In addition, most of the knowledge graphs consist of a higher amount of noise, and such graphs still require manual intervention [2]. Data visualizations, such as pie charts, bar charts, and plots, consist of abundant information in a precise manner. Such data visualizations are particularly developed to interrogate data for concerned people and are not established to be a machine-interpretable technique [30] [1]. Visualization techniques assist any individual to analyze the data, answer the questions, and describe how the relevant answer was achieved, which is considered a major step in many of the decision-making processes. However, analyzing the data with the use of visualization techniques is not preferable very much because of its complexities [17].

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Users frequently require answering compositional questions that need aggregation of various complex tasks, like reconstructing a value from the chart, determining extreme values, comparing the values and merging them, and computing sums and differences of values [24] [4].

Typically, a chart consists of a group of structural texts and data. In terms of bar charts, texts can be broadly categorized into four types, namely title, axis-tick or legend, axis-title, and data information, which are transformed into the height and width of the bar charts [16]. In order to extract the graphical and textual data obtained from bar charts, it is essential to determine the bounding boxes of bars and texts. In computer vision community, the detection of object is a major hurdle [3]. Statistical charts are specifically designed to facilitate the understanding of huge amounts of information and the relationships between various components of the data [19] [20]. The humans have high vision power can analyze the high amounts of data easily existed in a precise and rigid manner in the charts. Various applications are constructed according to the potential without considering the users. There exists a large amount of algorithms and tools that are capable of transforming text-to-speech, which helps individuals in understanding and reading the documents in most of the areas [7]. Charts and figures are efficient and broadly utilized to represent the data in documents in every field. Hence, it is significant to understand the documents by investigating secure technologies that can explain charts and figures and have the potential to answer different user queries. Due to the rapid advancement of deep learning technologies, it is easy to define the images, like charts that provide a high annotated corpus. However, it is a difficult task to collect massive data including of charts also with explanations mainly due to some reasons, such as cropping of legal and copyright problems during the period of extraction of charts from various documents and scrapping of charts from documents [31] [6]. In our previous research, we introduced a natural language query-to-chart image retrieval framework, entitled "Bridging Language to Visuals: Towards Natural Language Query-to-Chart Image Retrieval," which aimed to establish a connection between textual queries and visual chart representations. The proposed approach investigated the retrieval of relevant chart images in response to natural language queries, thereby providing an initial step toward facilitating efficient access to chart-based information. However, retrieving a relevant chart image does not directly provide the specific information requested by the user and still requires further interpretation of the retrieved visualization. This limitation motivates the development of a more comprehensive approach capable of understanding chart images and generating answers to natural language questions, which is the primary focus of the present research [32].

Deep learning has converted the computer vision and natural language processing landscapes and has become a widely utilized tool in their related applications [21]. The Convolutional Neural Networks (CNN) has become successful in the field of Image Net Classification task [1]. Moreover, Long Short-Term Memory (LSTM) has the potential to handle difficult process, such as machine passage comprehension [4], sentence summarization [10], and Neural Machine Translation [11]. The models of Neural Network are utilized as a preliminary attempt to model reasoning. An excessive amount of visual reasoning techniques [12], [13] have been generated to such potentials of neural networks. However, a visual question answering tasks need a collaboration of Natural Language Processing (NLP) [9], reasoning and Computer Vision Models. However, the model has the potential to obtain representations of the image and question rather than merging such representations in order to create an answer. This complex task assists the machines to perform visual signals and utilizes it to tackle multi-modal problems [8]. In addition, a deep learning solution is presented that automatically excerpts the data from bar and pie charts and significantly transforms a chart into a relational data table. Initially, this method detects the kind of a chart and then utilizes a single deep learning technique that excerpts all the relevant constituents and data [3].

The primary intention of this research is to design and develop hybrid network architecture for QA system from bar chart images. Conventional methods employed for question answering system faced many limitations and various types of charts cannot be processed by the same system. This method provides effective reading of answers from bar chart images. The proposed technique consists of two phases, namely training and testing phase. In training phase, the bar chart image and Question and Answer pair are considered as the inputs, whereas the testing phase considers the bar chart image and Questions as the inputs. Finally, the answers are predicted using residual layer of hybrid network architecture, where the network classifier is trained using the proposed algorithm named ALPO. However, the hybrid architecture is derived by the combination of Deep LSTM and Deep ResNet.

- **Proposed ALPO-based hybrid network:** An effective strategy is exploited for QA system to understand the answers from bar chart images using proposed ALPO-based hybrid architecture. However, the hybrid architecture is derived by the integration of Deep LSTM and Deep ResNet, which is specifically designed to increase the performance of QA system and to effectively predict the result.

The rest of the paper is structured as follows: Section 2 deliberates the literature review of QA systems along with their merits and demerits. Section 3 introduces newly developed hybrid architecture and the description of the structure is briefly explained. The proposed ALPO algorithm and its training procedure are detailed in section 4. Section 5 illustrates about the results and discussion of the proposed method. Finally, the research comes to a conclusion in section 6.

2. MOTIVATION

This section explains the literature review of various existing question answering system along with their advantages and demerits that motivates the researchers to design and develop an effective method for question answering system from bar chart images.

2.1 Literature Survey

Various traditional techniques related to question answering system are deliberated in this section. Kushal Kafle, *et al.* [1] developed DVQA for understanding the bar charts. Here, two DVQA algorithms were presented to handle chart-specific words in questions and answers. This model served as a significant proxy task for understanding visual attention, memory, and reasoning capabilities. However, even a slight variation was completely altered the chart's information is still remains as a challenging limitation. Hao Wang, *et al.* [2] modeled an efficient question answering system based on semi-structured tabular data. This system was mainly comprised with two parts that were utilized to choose the candidate tables and delivered answers. This system achieved higher performance due to the effect of attention layer. However, it failed to add MCQs with multiple correct choices to the dataset. In addition, the system also failed to concentrate on recall of table selection model. Xiaoyi Liu, *et al.* [3] devised a framework of single deep neural network that mainly includes text recognition, object detection, and object matching modules. The presented framework effectively controlled both bar and pie charts and there was a possibility to extend this framework to different types of charts by augmenting the training data. The method provided successful results for simulated bar charts. The major challenge lies in this technique is that it achieved poor performance while downloading the images from internet. Dae Hyun Kim, *et al.* [4], introduced an automatic chart question answering pipeline that creates visual explanations defining how the answer was obtained. The developed mechanism first extracted the data and visual encodings from an input vega-line chart. Then, it converted references into visual values. At last, it utilized a template-based model to describe by means of natural language in which how the answer was determined using the chart's visual features. The developed pipeline framework provided more transparency than the answers generated manually. The issues, such as slight variations, and low performance in fluency is also a major drawback of this system.

Kushal Kafle, *et al.* [5], developed a Chart Question Answering (CQA) algorithm called Parallel Recurrent Fusion of Image and Language (PReFIL) for fine grained-measurements and effectively handled the out-of-vocabulary words for both questions and answers. Here, PReFIL knows about the bio-modal embeddings by merging question and image features and after that collaborated such leaned embeddings to answer the questions. Although the method attained better performance in terms of reconstruction of information from charts, but it could not able to provide better performance utilizing dynamic encoding method. Ritwick Chaudhry, *et al.* [6], devised LEAF-QA along with an attention network to effectively deal with the limitations of multimodal QA system. Here, the data was augmented with a set of test that was built from unavailable data sources to test in order to generate the question answering models. This LEAF-QA approach attained better performance across various question types, but failed to provide higher performance for more complicated relational questions. Monika Sharma, *et al.* [7], developed an effective QA system called ChartNet-based on a MAC-Network to provide answers from a predefined vocabulary of generic answers. This approach was also employed to process visual question answering over statistical charts. The main specialty of this approach was that it predicted both vocabulary and out of vocabulary answers. However, it failed to generate a textual summary of the statistical charts. Revanth Reddy, *et al.* [8], presented a deep learning model, which was mainly designed for handling the issues of reasoning task of question-answering system on categorical plots. The presented model was tested on the FigureQA dataset that provides images for scientific representations, such as pie charts, and bar graphs. Moreover, this model targeted to handle the numeric and visual reasoning tasks by exploiting modular components. The major limitation exists on this method is that it delivered low performance while processing low-median and high-median questions.

2.2 Major Challenges

Some of the challenges confronted by the existing QA system are deliberated below:

- Though the DVQA dataset developed in [1], promoted the dynamic question encoding. The major complication in this dataset is that it had less capability to consider plots, pie-charts, and some other visualization since the dataset consists of only bar charts.
- The chart QA pipeline model provided a promising solution with reasonable accuracy. However, the system failed to detect the synonyms for visual features [4].
- To augment the training dataset through comprehensive simulation was a hassle free task as it considerably enhances the performance of Single Deep Neural Network introduced in [3]. However, this network was not enough capability to deal with small objects.
- An effective method modeled in [2], improved the performance of recalling of table selection technique. The major challenge is to provide accurate decision on choice selection.
- In [8], Deep Learning technique achieved significant improvements in training duration in comparison with Relational networks. In addition, QA tasks on real life scientific figures were not accomplished.

3. Hybrid architecture: LST-RES-NET

This research is focused to predict accurate answers from bar chart images by designing new hybrid network architecture named LST-ResNet for effective QA system. However, this hybrid network structure is derived by the integration of Deep LSTM and Deep ResNet. The proposed network consists of five layers, namely embedding layer, LSTM layer, residual layer, Attention layer, and Multi-layer perceptron (MLP) layer. The bar graph image, and Question and Answer pair are considered as an input data. The input data undergoes several processes and finally the answer is retrieved at the MLP layer. The schematic view of hybrid architecture of LST-ResNet is portrayed in figure 1. The layers in the hybrid LST-ResNet is described as follows:

3.1 Embedding layer

The first layer of the hybrid network architecture is embedding layer, in which the input data is acquired from the database. The input bar chart image B_i is applied to the embedding layer for further processing. The purpose of utilizing the embedding layer is that it completely transforms the bar graph images into vectors. It is essential to encode the input data into integer. Moreover, an embedding layer is utilized to convert syntactic words into real vectors by considering the semantic and syntactic relationships between words.

(i) Kernel: Kernel is flexible in specifying a definite positive kernel that is employed to specify any data representations, like matrices, vectors, sequences, and graphs. The kernel-guided embedding layer usually encodes a similarity metric that is customized for matching document and query. In addition to this, kernel is used to offer a versatile framework to process statistical learning with data representations [26].

(ii) Bias: Bias plays a major role as a threshold for meaningful activation of neurons. The result obtained from the embedding layer is further subjected to LSTM layer.

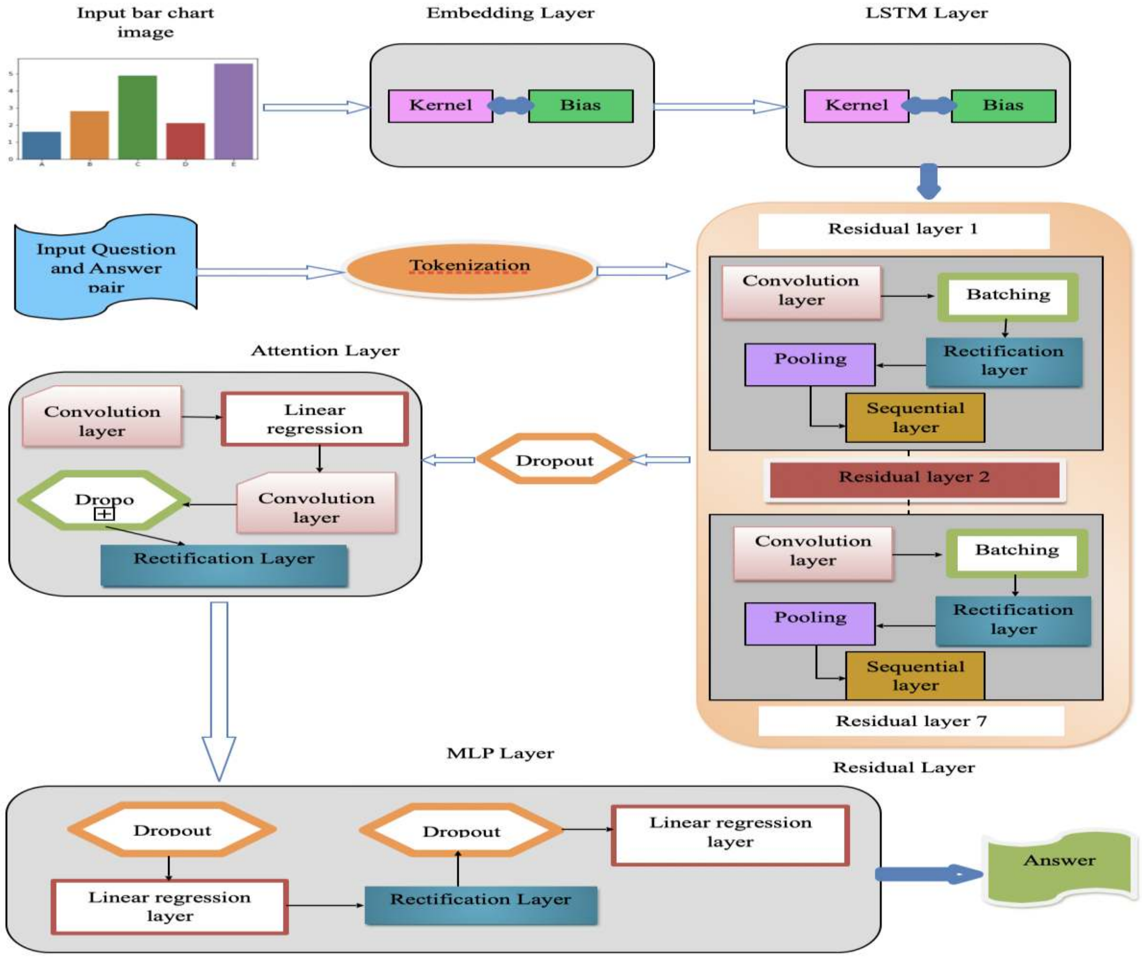


Figure 1. Proposed hybrid network architecture

3.2 LSTM layer

The LSTM layer is processed according to the selected input. However, not all input images are subjected to training process. Among all the input bar chart images, some specified and selected images are utilized to train the LSTM layer, thereby providing efficient information gain by effectively tuning the layer. With such inputs, LSTM layer observes the correlation and mapping between the input images and their projection. Typically, LSTM layer is mainly comprised with three elements, such as kernel, bias, and recurrent kernel.

(i) Kernel: The kernel is mainly utilized to filter small areas of an image in order to generate feature values of small areas. In some of the particular applications, it is significant to employ multiple convolution kernels, such that the convolutional kernels are considered to represent an image style.

(ii) Bias: For meaningful stimulation of neurons, these bias act as a significant threshold. Here, the kernel acquires the input of image along with the result of recurrent kernel in order to remember the former details based on the structure of image.

3.3 Residual layer

The third layer is the residual network layer, which is considered as an important layer in this architecture. Residual Network (ResNet) [18] is a kind of Convolutional Neural Network (CNN) structure that is generally utilized in the domain of ImageNet Large Scale Visual recognition challenge (ILSVRC). The benefit of ResNet is that it effectively trains the extreme deep neural networks by applying the principle of shortcut connections, such that one or more layers can be skipped. Moreover, it effectively addresses the issues of low accuracy without degrading the performance and solves the limitation of vanishing gradient [23]. Another advantage of ResNet is that it facilitates the training process due to the smooth travelling of gradients through the layers. Fast training process is a significant characteristic of ResNet because the residual connections do not hinder any gradients. There are n number of residual layers in the residual network. The input image is passed through a number of units including Convolutional layer, Batching layer, Rectification layer, Pooling layer, and Sequential layer.

(i) Convolution layer: Convolution layers are used to extract the effective features automatically from input feature maps by a group of small receptive fields, like kernels with sufficient depth.The mathematical process of conv2D is given by,

$$Conv\ 2D(z) = \sum_{W=0}^{p-1} \sum_{H=0}^{p-1} J_{W,H} V_{(e+W)(g+H)} \tag{1}$$

where, e and g are utilized to record the coordinates. Here, J denotes the $p \times p$ kernel matrix. The position index in the 2D matrix is denoted by $W \times H$.

(ii) Batching layer: In most of the deep learning techniques, the training set are partitioned into a group of small sets known as mini-batches. After that, the model is trained based on these mini-batches in order to achieve a better tradeoff between the computation complexity and the convergence speed. However, the mini-batch technique is highly vulnerable to limitation such as, internal covariate shift that may degrade the performance of the training speed and its stability. Therefore, Batch Normalization (BatchNorm) is a technique developed by Ioffe and Szegedy to decrease the issue of internal covariate shift by normalizing the input layers by means of scaling, thereby enhancing the training speed and also mitigates the overfitting problem.

(iii) Rectification layer: Rectification layer is a linear function that results the input directly if it is a positive one or else it results as zero**.**The raw dataset contains complex and non-linear features, such that the non-linearity of extracted features is increased using the non-linear activation function.

$$\mathrm{Re}\,LU(R) = \begin{cases} 0, & R < 0 \\ R, & R \geq 0 \end{cases} \tag{2}$$

where, R is the input feature.

(iv) Pooling layer: The pooling layer is generally impelled into the consecutive convolution layers is mainly employed to mitigate the spatial dimension of the feature map and effectively reduces overfitting. Moreover, it reduces the computational complexities. In fact, pooling layer effectively mitigates the dimensions of data when compared with the convolution layer.

$$W_{out} = \frac{W_{in} - K_W}{q} + 1 \tag{3}$$

$$H_{out} = \frac{H_{in} - K_H}{q} + 1 \tag{4}$$

Here, W_{in} and H_{in} specifies the width and height of 2D matrix inputs, respectively. Similarly, the outputs of 2D matrix is represented as W_{out} and H_{ou}. Moreover, the width and height of kernel size is denoted as K_W and K_H , respectively.

(v) Sequential layer: It is utilized to make decision functions. The feature maps of every layer are subjected to sequential layer without any variation. However, the organization of sequential layer is usually depends upon the type of application and data.

3.3.1 Dropout

Dropout is mainly employed to ignore vanishing of all useful information from the units. It is essential to consider both of the output of neuron and its internal structure to develop an efficient dropout model. Typically, dropout processes by decoupling the nodes from present layer to the next layer and it is fully a stochastic regularization technique specifically utilized for maintaining the fully connected layers.

3.4 Attention layer

Attention layer processes only essential information and also it enhances the performances of entire system. Attention layer gives more priority to the target value rather than focusing on all information. It has enough potential to prevent all useful information and has the ability to handle various types of images over the conventional techniques. This layer automatically chooses the significant relevant source data to generate suitable output in various fields, such as video or image captioning, machine translation, and visual question answering system. The layers impelled in this attention layer are convolution layer, Linear regression layer, dropout, and rectification layer.

3.5 Multi-Layer Perceptron

A multi layer perceptrons (MLP) is a finite acyclic graph in which the nodes are known as neurons connected with logistic activation. The features are combined to form a feature vector and subjected to the linear regression layer.

(i) Linear regression layer: In linear regression layer, the final output is predicted and finally the answer for input bar image is obtained.

On the other hand, Question and answer pair is considered as an input and is subjected to tokenization, where the significant keywords are extracted. The detailed explanation of this process is presented in section 4.

4. Proposed ALPO-based hybrid network

The main intention of this research is to design and develop hybrid architecture named LST-RES-NET to provide an effective Question and Answering system. The proposed approach called ALPO-based hybrid network plays a main role in providing effective reading of answers from bar chart images. The proposed scheme consists of two phases, namely training phase and testing phase. In training phase, the bar chart image and Question and Answer pair are considered as an input, whereas the testing phase considers the bar image and question as an input in order to retrieve the appropriate answer.

4.1 Training phase

The training phase consists of two inputs, such as bar chart image, and Question and Answer pair, which are acquired from the dataset specified in [29]. Initially, the input bar chart image is subjected to the embedding layer of newly developed hybrid architecture called LST-RES-NET. Then, the output of the embedding layer is passed through the residual layer of Deep ResNet for further processing. On the other hand, the Quest and Answer pair is considered as an input image, which is subjected to stop word removal and stemming process in order to remove the redundant and unnecessary words from the Question and Answer pair. Furthermore, the obtained result is subjected to perform tokenization, where the significant keywords are extracted. Finally, the extracted keywords are passed through a residual network to provide an appropriate answer. However, the residual layer of hybrid network is trained using the proposed ALPO algorithm. The proposed ALPO is derived by the integration of Ant Lion Optimization (ALO) [14], and Political Optimizer (PO) [15]. The schematic diagram of training phase of proposed ALPO-based hybrid network is represented in figure 2.

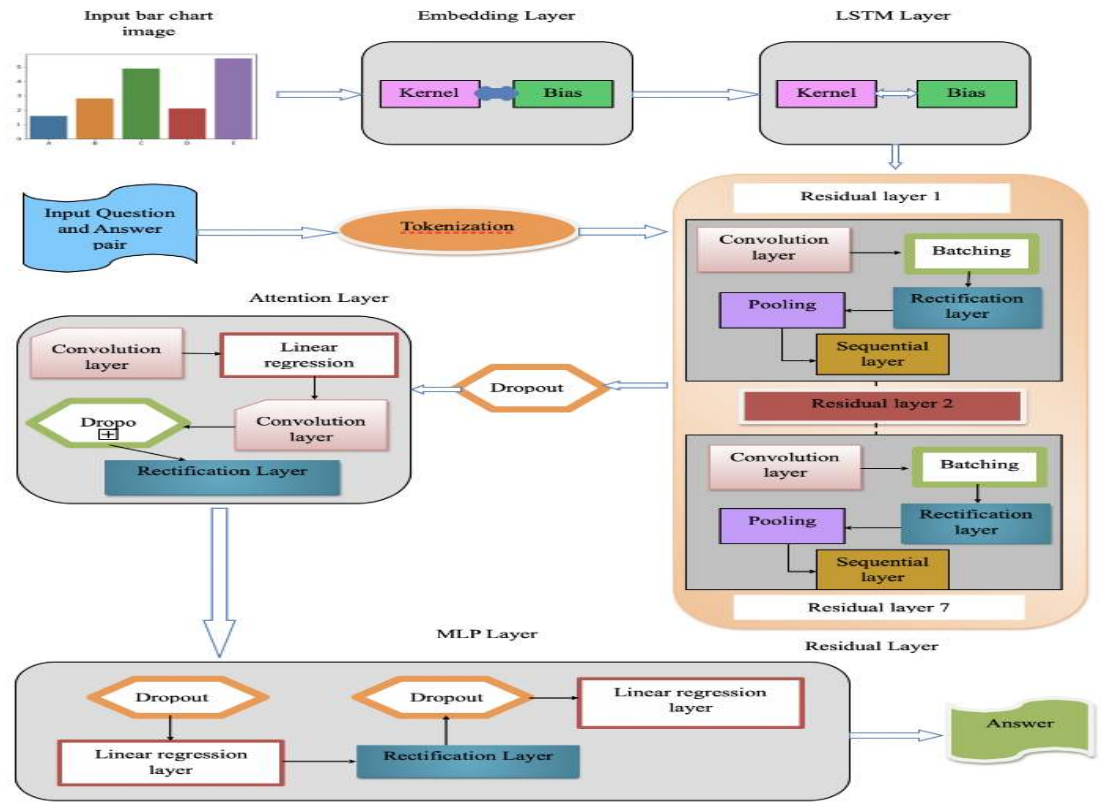


Figure 2. Training Phase of proposed ALPO-based hybrid network

4.1.1 Acquisition of Input bar chart image and Question and Answer pair

Let us consider D_1 be the training set of bar chart image and it is expressed as,

$$D_1 = \{B_1, B_2, ...B_i, ...B_n\} \tag{5}$$

where, B_n be the number of training samples and B_i represents the input bar chart image with a dimension of $[u \times v]$.

Similarly, D_2 be the training set of Question and Answer pair, which is represented as,

$$D_2 = \{Q_{a_1}, Q_{a_2, ...} Q_{a_h}, ...Q_{a_f}\} \tag{6}$$

where, Q_{a_h} is the input Question and Answer pair.

4.1.2 Embedding layer of LST-RES-NET

The input B_i is subjected to the embedding layer of hybrid LST-RES-NET, where the bar chart image is transformed into vectors. However, the embedding layer consists of kernel and bias and each has its own characteristics. The kernel assists to specify the data representations, such as bar chart, matrices, and sequences, whereas bias acts as a threshold activation function. The embedding layer of LST-RES-NET is elaborated in section 3.1.

4.1.3 Residual layer

The result obtained from embedding layer of LST-RES-NET is passed through the residual layer of Deep ResNet for further processing. The main purpose for selecting the residual network is that it provides fast training process and reduces the computational time. Moreover, it trains the deep neural networks effectively and also mitigates the problem of vanishing gradient. The residual layer in Deep ResNet consists of five units, including convolutional layer, batching, pooling, rectification and sequential layer, which are described in section 3.3.

4.1.4 Stop word removal

On the other hand, the input Q_{a_h} is subjected to process stop word removal and stemming technique, in which the unnecessary words are eliminated. In this process, the stop words, like am, is, are, etc., are eliminated according to the list of stop words prepared already. The existence of stop words overcomes several important words since they commonly occur in documents. Default words, such as of, is, and like are expelled from the features list as they do not represent any information. Typically, stop words are frequently happened at the time of sentence formation but does not have any significance to determine the subject of the document. The result obtained from stop word removal process is signified as S_r.

4.1.5 Stemming

Stemming is the process of reducing the term to its origin form by extracting the suffix and prefix of the word. Stemming is considered as a significant process. In this step, suffix list is utilized to eliminate the suffixes from words for generating appropriate stem word. Stemming process transforms the derivative words of document to its base root form and it is expressed as,

$$S_t = \{C_b, 1 \leq 0 \leq d\} \tag{7}$$

The result obtained from the stop word removal and stemming process is denoted as $S, S \in \{S_r, S_t\}$.

4.1.6 Tokenization

Tokenization is defined as the process of excerpting the suitable content from a document by breaking the content into small or chunks of text called as tokens [28]. The question entered into the tokenization process is transformed into single words or tokens and the process is mainly utilized to ease up the complex questions. It is used to extract the appropriate keywords from the given question.

$$K = \bigcup_{c=1}^{M} \text{Ext}\left(\text{POS-noun}(A_c)\right) \tag{8}$$

However, the query keyword [27] is compared with the documents and extracts the most appropriate document from the knowledge base using semantic similarity between them. Similarity computation between keywords in the document is given by,

$$Sim(x, y) = \frac{1}{3}\left(\frac{x}{l_1} + \frac{x}{l_2} + \frac{x-y}{x}\right) \tag{9}$$

where, x denotes the matching characters, n signifies the misplaced characters and length between two words is indicated as l_1, and l_2. The tokenized keyword extracted from tokenization is specified as T_k.

4.2 Testing phase

The second step in this approach is the testing phase, in which the bar chart image and a random question are considered as an input. The input bar chart image is subjected to the embedding layer of LST-RES-NET and then to the residual layer similar to that of the training phase. On the other hand, the input question is subjected to stop word removal and stemming process. Moreover, the next phase is keyword extraction, where the document is broken into number of chunks called tokens. The tokenized keyword is passed through the residual layer in order to retrieve an appropriate answer. Figure 3 portrays the block diagram of testing phase of proposed ALPO-based hybrid network.

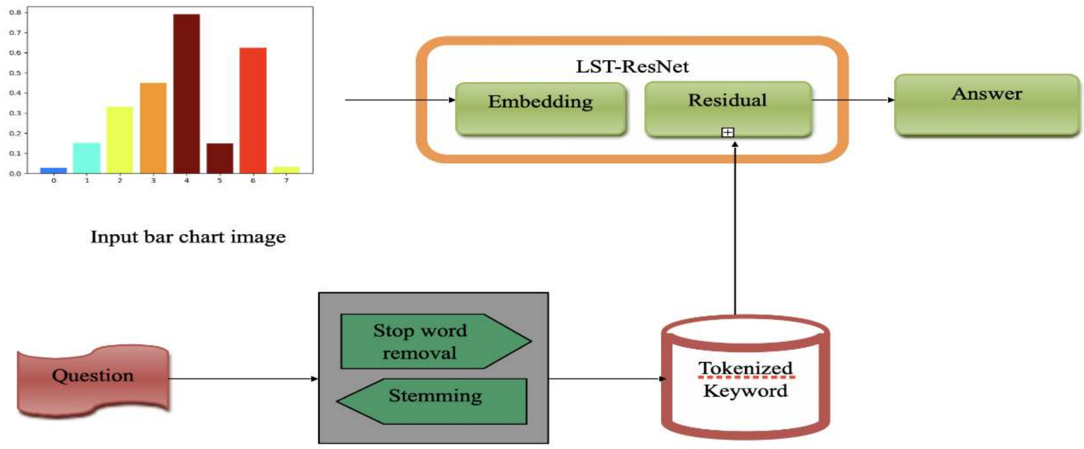


Figure 3. Testing phase of proposed ALPO-based hybrid network

4.3 Training model of proposed hybrid network

The main reason for selecting the residual network is that it helps to smooth out the training process in an easy manner due to the gradients and additional operators that bypass some layers. The architecture of proposed hybrid network is explained as follows.

4.3.1 Architecture of proposed hybrid network

The proposed hybrid network named LST-ResNet is derived by the combination of Deep LSTM and Deep ResNet. The hybrid architecture consists of five layers, namely embedding layer, LSTM layer, Residual layer, Attention layer, and Multi-layer perceptron. The embedding layer and LSTM layer is comprised with kernel and bias in order to transform the bar chart image into vector form. However, the residual layer consists of seven layers, in which each layer or block includes convolutional, batching, pooling, rectification, and sequential layer. The process of such layers are briefly explained in section 3.

4.3.2 Training residual layer of hybrid network using proposed ALPO

Politics is defined as the process of optimization from two different views, such as each individual discloses their goodwill in order to win the election and each party works well to increase their number of seats in the parliament to setup a new government. These reasons make the politics as a powerful inspiration for an optimization algorithm. However, each individual is assumed to be a candidate solution, and each individual's goodwill is assumed to be a position of the candidate solution. The objective function is evaluated by considering the election. PO is developed by inspiring from these four main aspects, like electoral process, intra-party collaboration, goodwill of the candidates, and cooperation of the winning candidates. In this research, a newly proposed method called ALPO is employed to train the proposed network. However, the proposed ALPO is derived by the incorporation of PO [15], and ALO [14]. The training procedure of PO is explained as follows:

Step 1: Initialization

Let us initialize the population G, which is divided into m number of political parties. Each party G_s includes m member or candidates. Each j^{th} member G_s^j is assumed to be a best solution.

$$G = \{G_1, G_2, G_3, \ldots\ldots, G_m\} \tag{10}$$

$$G_s = \{G_s^1, G_s^2, G_s^3, \ldots, G_s^m\} \tag{11}$$

$$G_s^j = \{G_{s,1}^j, G_{s,2}^j, G_{s,3}^j, \ldots\ldots, G_{s,w}^j\} \tag{12}$$

where, G_m denotes the number of political parties and G_s represents the s^{th} political party. Here, G_s^j specifies the j^{th} member of s^{th} political party and $G_{s,w}^j$ represents the w^{th} dimension of G_s^j. Correspondingly, there are m number of constituencies and s^{th} member of each party faces election from j^{th} constituency L_j and it is given by,

$$L = \{L_1, L_2, L_3, ..., L_m\} \tag{13}$$

$$L_j = \{G_1^j, G_2^j, G_3^j, ..., G_m^j\} \tag{14}$$

Step 2: Evaluate fitness function

The fitness function is utilized to compute the best solution of the optimization problem and it is expressed as,

$$\Delta_{fit} = \frac{1}{\gamma} \sum_{\delta}^{\gamma} [Z_\delta - P]^2 \tag{15}$$

where, Z_δ is the targeted output and the output obtained from the classifier is expressed as P.

Step 3: Determine the leader of each party and its constituency winner

The party leader is the best candidate in the party and the decision is released immediately after the common election. The Eq. (17) computed the selection process of party leader. G_s^* indicates the leader of s^{th} party and $F(G_s^j)$ determines the fitness of G_s^j.

$$\phi = \underset{1 \le j \le m}{\arg} \min F(G_s^j), \quad \forall s = \{1,2,..., m\} \tag{16}$$

$$G_s^* = G_s^\phi \tag{17}$$

Here, G^* indicates the group of all party leaders and the expression is given as,

$$G^* = \{G_1^*, G_2^*, G_3^*, ..., G_m^*\} \tag{18}$$

Once the election process is completed, the winner of all constituencies becomes the member of parliamentary. L^* specifies the group of all parliamentarians and L_j^* denotes the winner of j^{th} constituency.

$$L^* = \{L_1^*, L_2^*, L_3^*, ..., L_m^*\} \tag{19}$$

Step 4: Update the position of each party member

Election campaign phase is an important step, because it helps all the candidates to improve their performance during election stage. The whole process of election is presented using Eq. (20) and Eq. (21). According to the relationship of current fitness $F(G_s^j(\beta))$ of a candidate with its previous fitness $F(G_s^j(\beta-1))$. Based on such decisions, the position of the party member is updated either using Eq. (20) or Eq. (21). In case of improvements in the fitness value, the Eq. (20) is employed to update the member's position or else Eq. (21) is used. The variable N and X^* are utilized to develop the position of the party member, in which the variable N is randomly chosen between a limit of 0 and 1, and X^* keeps value of Y^{th} dimension of party leader $G_{s,Y}^*$, then the winner of the constituency is specified as $L_{j,Y}^*$.

There are three conditions to determine the best region.

$$G_{s,Y}^j(\beta+1) = \begin{cases} X^* + N\left(X^* - G_{s,Y}^j(\beta)\right), & if\ G_{s,Y}^j(\beta-1) \le G_{s,Y}^j(\beta) \le X^*\ or\ G_{s,Y}^j(\beta-1) \ge G_{s,Y}^j(\beta) \ge X^* \\ X^* + (2N-1)\left|X^* - G_{s,Y}^j(\beta)\right|, & if\ G_{s,Y}^j(\beta-1) \le X^* \le G_{s,Y}^j(tf)\ or\ G_{s,Y}^j(\beta-1) \ge X^* \ge G_{s,Y}^j(\beta) \\ X^* + (2N-1)\left|X^* - G_{s,Y}^j(\beta-1)\right|, & if\ X^* \le G_{s,Y}^j(\beta-1) \le G_{s,Y}^j(tf)\ or\ X^* \ge G_{s,Y}^j(\beta-1) \ge G_{s,Y}^j(\beta) \end{cases}$$

(20)

$$G_{s,Y}^{j}(\beta+1)=\begin{cases} X^{*}+(2N-1)\left|X^{*}-G_{s,Y}^{j}(\beta)\right|, if\ G_{s,Y}^{j}(\beta-1)\leq G_{s,Y}^{j}(\beta)\leq X^{*}\ or\ G_{s,Y}^{j}(\beta-1)\geq G_{s,Y}^{j}(\beta)\geq X^{*} \\ G_{s,Y}^{j}(\beta-1)+N\left(G_{s,y}^{j}(\beta)\right), if\ G_{s,Y}^{j}(\beta-1)\leq X^{*}\leq G_{s,Y}^{j}(\beta)\ or\ G_{s,Y}^{j}(\beta-1)\geq X^{*}\geq G_{s,Y}^{j}(\beta) \\ X^{*}+(2N-1)\left|X^{*}-G_{s,Y}^{j}(\beta-1)\right|, if\ X^{*}\leq G_{s,Y}^{j}(\beta-1)\leq G_{s,Y}^{j}(tf)\ or\ X^{*}\geq G_{s,Y}^{j}(\beta-1)\geq G_{s,Y}^{j}(\beta) \end{cases}$$

(21)

Let us consider the condition. (1) of Eq. (21),

$$G_{s,y}^{j}(\beta+1)=X^{*}+(2N-1)\left|X^{*}-G_{s,y}^{j}(\beta)\right| \tag{22}$$

Assume $X^{*}>G_{s,Y}^{j}(\beta)$,

$$G_{s,Y}^{j}(\beta+1)=X^{*}+(2N-1)X^{*}-(2N-1)G_{s,Y}^{j}(\beta) \tag{23}$$

$$G_{s,Y}^{j}(\beta+1)=X^{*}(1+2N-1)-(2N-1)G_{s,Y}^{j}(\beta) \tag{24}$$

$$G_{s,Y}^{j}(\beta+1)=2NX^{*}-(2N-1)G_{s,Y}^{j}(\beta) \tag{25}$$

Ant Lion Optimizer (ALO) [14] is a nature-inspired optimization algorithm that inspires about the hunting behavior of ant lions. The two interesting things that has been noticed in the hunting style of ant lions are level of hunger and shape of the moon. It traps the ants by digging a cone -shaped pit in sand. The edge of the pit easily trapped the ants as it is very sharp that grasps the ants to fall to the pit. In this step, ALO is incorporated with the PO to achieve optimal solution.

The random walk of ants are normalized using the below equation,

$$G_{s,Y}^{j}(\beta+1)=\frac{(G_{s,Y}^{j}(\beta)-U_{s})X(\alpha_{s}-\sigma_{s}^{bb})}{(\alpha_{s}^{bb}-U_{s})}+\sigma_{s} \tag{26}$$

where, U_s denotes the minimum of random walk of s^{th} variable, σ_s^{bb} signifies the minimum of s^{th} variable at bb^{th} iteration and the maximum of s^{th} variable at bb^{th} iteration is indicated as α_s^{bb}.

$$\frac{(G_{s,Y}^{j}(\beta)-U_{s})(\alpha_{s}-\sigma_{s}^{bb})}{\alpha_{s}^{bb}-U_{s}}=G_{s,y}^{j}(\beta+1)-\sigma_{s} \tag{27}$$

$$(G_{s,Y}^{j}(\beta)-U_{s})\left(\alpha_{s}-\sigma_{s}^{bb}\right)=(G_{s,Y}^{j}(\beta+1)-\sigma_{s})(\alpha_{s}^{bb}-U_{s}) \tag{28}$$

$$G_{s,y}^{j}(\beta)(\alpha_{s}-\sigma_{s}^{bb})-U_{s}(\alpha_{s}-\sigma_{s}^{bb})=(G_{s,Y}^{j}(\beta+1)-\sigma_{s})(\alpha_{s}^{bb}-U_{s}) \tag{29}$$

$$G_{s,Y}^{j}(\beta)(\alpha_{s}-\sigma_{s}^{bb})=(G_{s,Y}^{j}(\beta+1))(\alpha_{s}^{bb}-U_{s})+U_{s}(\alpha_{s}-\sigma_{s}^{bb}) \tag{30}$$

$$G_{s,y}^{j}(\beta)=\frac{(G_{s,Y}^{j}(\beta+1))(\alpha_{s}^{bb}-U_{s})+U_{s}(\alpha_{s}-\sigma_{s}^{bb})}{\alpha_{s}-\sigma_{s}^{bb}} \tag{31}$$

Substituting Eq. (31) in Eq. (25),

$$G_{s,Y}j = 2NX^* - (2N-1)\frac{(G_{s,Y}^j(\beta+1))(\alpha_s^{bb} - U_s) + U_s(\alpha_s - \sigma_s^{bb})}{\alpha_s - \sigma_s^{bb}} \tag{32}$$

$$G_{s,y}^j = 2NX^* - (2N-1)\frac{G_{s,Y}^j(\beta+1)(\alpha_s^{bb} - U_s) - \sigma_s(\alpha_s^{bb} - U_s) + U_s(\alpha_s - \sigma_s^{bb})}{\alpha_s - \sigma_s^{bb}} \tag{33}$$

$$G_{s,Y}^j(\beta+1) = 2NX^* - (2N-1)\frac{G_{s,Y}^j(\beta+1)(\alpha_s^{bb} - U_s)}{\alpha_s - \sigma_s^{bb}} - (2N-1)\frac{U_s(\alpha_s - \sigma_s^{bb}) - \sigma_s(\alpha_s^{bb} - U_s)}{\alpha_s - \sigma_s^{bb}} \tag{34}$$

$$G_{s,Y}^j(\beta+1) + \frac{2N-1}{\alpha_s - \sigma_s^{bb}}G_{s,Y}^j(\beta+1)(\alpha_s^{bb} - U_s) = 2NX^* - \frac{(2N-1)}{\alpha_s - \sigma_s^{bb}}(U_s(\alpha_s - \sigma_s^{bb}) - \sigma_s(\alpha_s^{bb} - U_s)) \tag{35}$$

$$G_{s,Y}^j(\beta+1)(1 - \frac{2N-1}{\alpha_s - \sigma_s^{bb}}(\alpha_s^{bb} - U_s)) = 2NX^* - \frac{(2N-1)}{\alpha_s - \sigma_s^{bb}}\left(U_s\left(\alpha_s - \sigma_s^{bb}\right) - \sigma_s\left(\alpha_s^{bb} - U_s\right)\right) \tag{36}$$

$$G_{s,Y}^j(\beta+1)\left(\frac{\alpha_s - \sigma_s^{bb} - (2N-1)(\alpha_s^{bb} - U_s)}{\alpha_s - \sigma_s^{bb}}\right) = 2NX^* - \frac{(2N-1)}{\alpha_s - \sigma_s^{bb}}\left(U_s\left(\alpha_s - \sigma_s^{bb}\right) - \sigma_s\left(\alpha_s^{bb} - U_s\right)\right) \tag{37}$$

$$G_{s,Y}^j(\beta+1)\left(\frac{\alpha_s - \sigma_s^{bb} - 2N\alpha_s^{bb} + 2NU_s + \alpha_s^{bb} - U_s}{\alpha_s - \sigma_s^{bb}}\right) = 2NX^* - \frac{(2N-1)}{\alpha_s - \sigma_s^{bb}}\left(U_s\left(\alpha_s - \sigma_s^{bb}\right) - \sigma_s\left(\alpha_s^{bb} - U_s\right)\right) \tag{38}$$

$$G_{s,Y}^j(\beta+1)\left(\frac{2\alpha_s^{bb}(1-N) - \sigma_s^{bb} + U_s(2N-1)}{\alpha_s - \sigma_s^{bb}}\right) = 2NX^* - \frac{(2N-1)}{\alpha_s - \sigma_s^{bb}}\left(U_s\left(\alpha_s - \sigma_s^{bb}\right) - \sigma_s\left(\alpha_s^{bb} - U_s\right)\right) \tag{39}$$

$$G_{s,Y}^j(\beta+1) = \frac{\alpha_s - \sigma_s^{bb}}{2\alpha_s^{bb}(1-N) - \sigma_s^{bb} + U_s(2N-1)}\left[2NX^* - \frac{(2N-1)}{\alpha_s - \sigma_s^{bb}}\left(U_s\left(\alpha_s - \sigma_s^{bb}\right) - \sigma_s\left(\alpha_s^{bb} - U_s\right)\right)\right] \tag{40}$$

Therefore, the update position of each party member becomes,

$$G_{s,Y}^j(\beta+1) = \begin{cases} \frac{\alpha_s - \sigma_s^{bb}}{2\alpha_s^{bb}(1-N) - \sigma_s^{bb} + U_s(2N-1)}\left[2NX^* - \frac{(2N-1)}{\alpha_s - \sigma_s^{bb}}\left(U_s\left(\alpha_s - \sigma_s^{bb}\right) - \sigma_s\left(\alpha_s^{bb} - U_s\right)\right)\right] \\ \qquad if\ G_{s,Y}^j(\beta-1) \le G_{s,Y}^j(\beta) \le X^*\ or\ G_{s,Y}^j(\beta-1) \ge G_{s,Y}^j(\beta) \ge X^* \\ G_{s,Y}^j(\beta-1) + N\left(G_{s,y}^j(\beta)\right), \quad if\ G_{s,Y}^j(\beta-1) \le X^* \le G_{s,Y}^j(\beta)\ or\ G_{s,Y}^j(\beta-1) \ge X^* \ge G_{s,Y}^j(\beta) \\ X^* + (2N-1)\left|X^* - G_{s,Y}^j(\beta-1)\right|, \quad if\ X^* \le G_{s,Y}^j(\beta-1) \le G_{s,Y}^j(tf)\ or\ X^* \ge G_{s,Y}^j(\beta-1) \ge G_{s,Y}^j(\beta) \end{cases}, \tag{41}$$

Step 5: Run party switching phase

This step is followed immediately after the election campaign phase. The party switching rate is denoted as η that starts from $\eta_{\max}$ and decreases to 0 over the course of iterations. Each member G_s^j is selected with a random probability of η is switched to randomly selected party G_O.

$$\phi = \arg\min_{1 \le j \le m} F\left(G_O^j\right) \tag{42}$$

Step 6: Run parliamentary affairs stage

A new government is setup after an interparty election. The constituency winners and party leaders are selected using Eq. (18) and Eq. (19). Each parliamentarian L_j^* updates their position with respect to the selected parliamentarian L_O^*.

Step 7: Termination

The above steps are repeated until the η is updated and reaches the best solution. Algorithm 1 portrays the pseudo code of proposed ALPO.

Sl. No	Pseudo code of proposed ALPO
1	**Input:** $m, \eta_{\max}$
2	**Output:** Final population ; $G_{s,Y}^j(\beta+1)$
3	Begin
4	Initialize the population G
5	Determine the fitness of each member G_s^j
6	Compute the group of party leader G^* using Eq. (18)
7	Determine the group of constituency winners L^* using Eq. (19)
8	$N=1$;
9	$G(N-1)=G;$
10	$F(G(N-1))=F(G);$
11	$\eta=\eta_{\max}$;
12	While $N \le \psi_{\max}$ *do*
13	*for each* $G_s \in G$ *do*
14	*for each* $G_s^j \in G$ *do*
15	$G_s^j = \left(G_s^j, G_s^j(N-1), G_s^*, L_j^*\right)$;

end
end
Run party switching phase (G,η);
Evaluate the fitness of each candidate G_s^j
Determine the group of party leaders G^*, using Eq. (18)
Compute the constituency winners L^* using Eq.(19)
Run parliamentary affairs (L^*, G);
$G(N-1) = G_{temp}$;
$F(G(N-1)) = F(G_{temp})$;
$\eta = \eta - \frac{\eta_{\max}}{\psi_{\max}}$;
$N = N+1$
end
Terminate

Algorithm 1. Pseudo code of proposed ALPO

5. RESULTS AND DISCUSSION

This section describes the results and discussion of proposed ALPO with respect to the evaluation metrics.

5.1 Experimental setup

The implementation of proposed ALPO-based hybrid network is carried out in a PC with intel core-i3 processor and the experimentation is done in the PYTHON tool using DVQA dataset specified in [29].

5.2 Dataset description

The corpus of DVQA dataset consists of images, metadata, and question and answer pairs. The images utilized in this dataset have the ability to expand up to 6.5 GB, whereas the question and answer pairs expand up to 750 MB. Moreover, this dataset provides elaborated annotations of each and every object that can act as either the source of extra supervision or additional evaluation of algorithm's performance.

5.3 Evaluation metrics

(i) Precision: precision is also called as positive predictive value and it is termed as the ratio of relevant contents to the reconstructed contents.

$$P_p = \frac{|\omega \cap \alpha|}{|\alpha|} \tag{43}$$

where, the relevant and retrieved document are denoted as ω and α, respectively.

(ii) Recall: Recall is also expressed as the ratio of relevant documents to the retrieved documents.

$$R_r = \frac{|\omega \cap \alpha|}{|\omega|} \quad (44)$$

(iii) F-measure: F-measure is typically utilized to test the accuracy. However, it is determined from precision and recall measure and it is termed as the twice of the product of recall and precision to the sum of the two measures and it is expressed as,

$$F_m = \frac{2.P_p.R_r}{P_p + R_r} \quad (45)$$

5.4 Performance analysis

This section explains the performance analysis of proposed ALPO-based hybrid network with respect to the evaluation metrics, such as precision, recall, and F-measure.

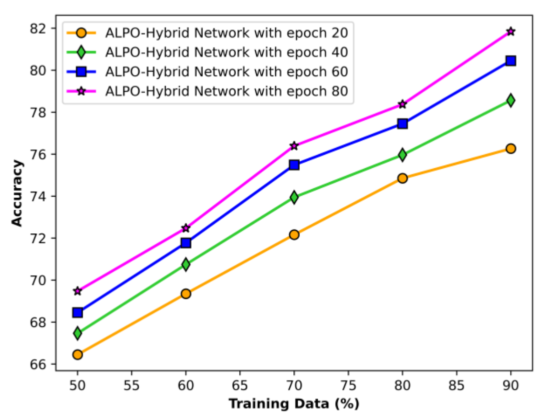
5.4.1 Analysis of proposed ALPO-hybrid network using easy type images

Figure 4 portrays the performance analysis of proposed ALPO-hybrid network with respect to the evaluation metrics by varying the percentage data for easy type images. Figure 4a) represents the analysis of proposed ALPO-hybrid network in terms of accuracy. When the training data=50%, the accuracy obtained by proposed ALPO-hybrid network with epoch 20=66.453, with epoch 40= 67.463, with epoch 60=68.453, and with epoch 80=69.475. Similarly, by varying the training data to 90%, the proposed ALPO-hybrid network attained the accuracy of 76.26 for epoch=20, 78.564 for epoch=40, 80.453 for epoch=60, and 81.846 for epoch=80

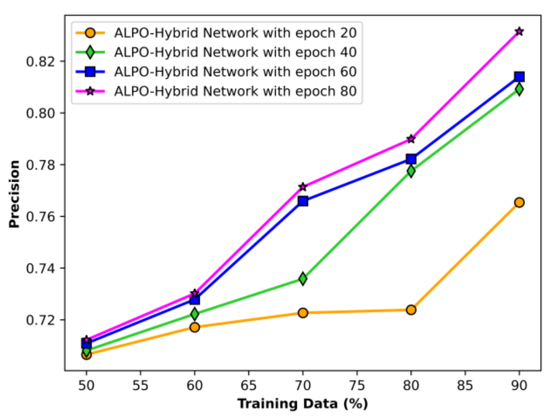
Figure 4b) represents the analysis of proposed ALPO-hybrid network in terms of precision. When the training data=50%, the precision obtained by proposed ALPO-hybrid network with epoch 20=0.707, with epoch 40= 0.708, with epoch 60=0.711, and with epoch 80=0.712. Similarly, by varying the training data to 90%, the proposed ALPO-hybrid network attained the precision of 0.765 for epoch=20, 0.809 for epoch=40, 0.814 for epoch=60, and 0.832 for epoch=80.

The analysis of proposed ALPO-hybrid network with respect to recall is depicted in figure 4c). If the training data=50%, the recall attained by the proposed ALPO-hybrid network with epoch 20, with epoch 40, with epoch 60, are 0.700, 0.716, 0.746, and with epoch 80 is 0.758, respectively. If the training data is increased to 90%, the recall obtained by the proposed approach with epoch 20 is 0.783, with epoch 40 is 0.796, with epoch 60 is 0.832, and with epoch 80 is 0.839.

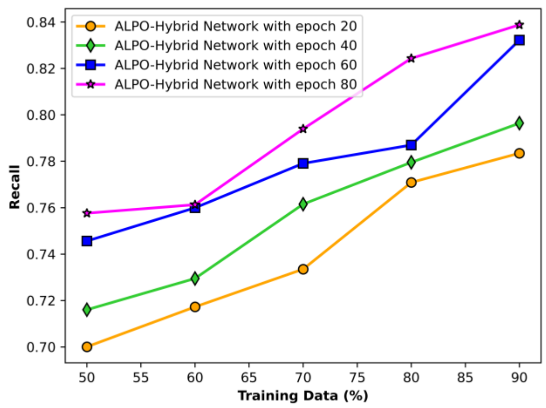
Figure 4d) illustrates the analysis of proposed ALPO-hybrid network in terms of F1-measure. When the training data is 90%, the F1-measure achieved by the proposed ALPO-hybrid network with epoch=20 is 0.774, with epoch=40 is 0.803, with epoch=60 is 0.823, and with epoch=80 is 0.835.



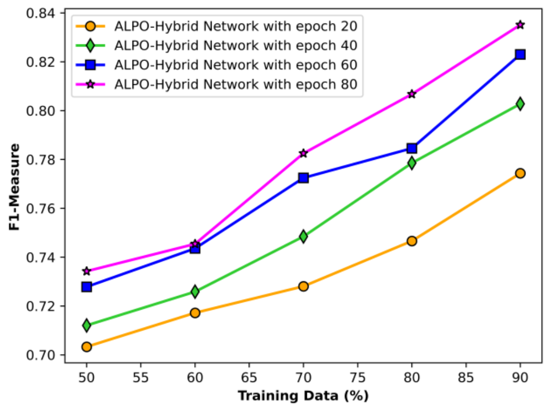
(a)



(b)



(c)



(d)

Figure 4. Analysis of proposed ALPO-hybrid network based using easy type images a)Accuracy b)Precision c) Recall d) F1-measure

5.4.2 Analysis of proposed ALPO using hard type images

Figure 5 represents the analysis of proposed ALPO-hybrid network with respect to the evaluation metrics for hard type images by varying the training data.

Figure 5a) depicts the analysis of developed scheme by considering accuracy. If the training data=50%, the accuracy achieved by the proposed ALPO-hybrid network with epoch=20 is 65.037, with epoch=40 is 65.786, with epoch=60 is 66.756, and with epoch=80 is 67.384.

Figure 5b) depicts the analysis of developed scheme by considering precision. If the training data=50%, the precision achieved by the proposed ALPO-hybrid network with epoch=20 is 0.705, with epoch=40 is 0.723, with epoch=60 is 0.733, and with epoch=80 is 0.760.

The analysis of proposed ALPO-hybrid network in terms of recall is portrayed in figure 5c). When the training data =50%, the recall attained by the developed ALPO-hybrid network with epoch 20 is 0.700, with epoch 40 is 0.701, with epoch 60 is 0.706, and with epoch 80 is 0.764. Correspondingly, if the training data is increased to 90%, the recall obtained by the proposed ALPO-hybrid network with epoch=20, with epoch=40, with epoch=60, and with epoch=80 is 0.751, 0.773, 0.794, and 0.807.

Figure 5d) depicts the analysis of proposed ALPO-hybrid network with respect to F1-measure. When the training data=90%, the F1-measure attained by the developed ALPO-hybrid network with epoch=20 is 0.768, with epoch 40 is 0.788, with epoch 60 is 0.805, and with epoch 80 is 0.812.

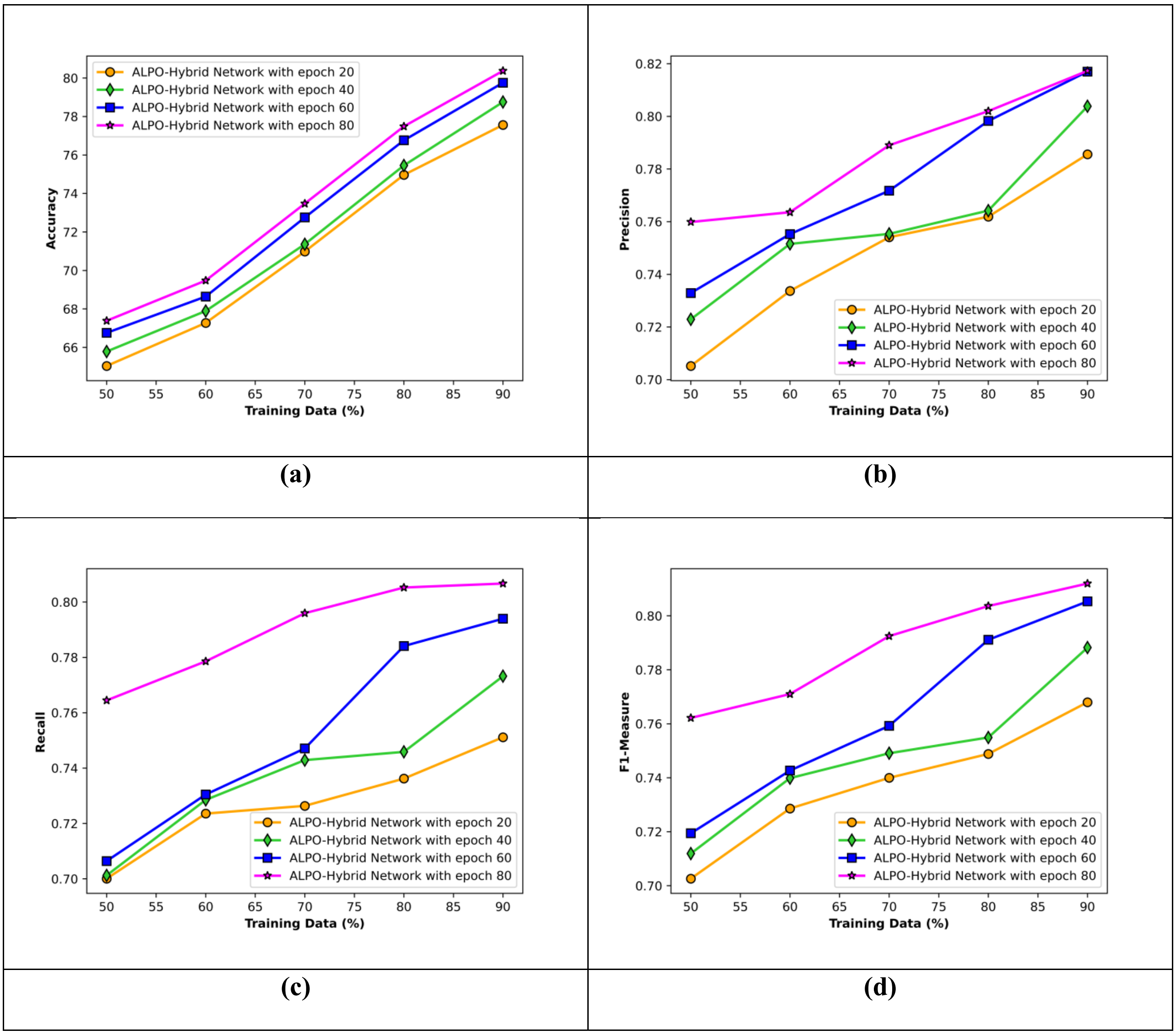


Figure 5. Analysis of proposed ALPO-hybrid network based using hard type images a)Accuracy b)Precision c) Recall d) F1-measure

5.5 Comparative methods

The performance improvement of the developed approach is analyzed by comparing it with the traditional schemes, such as MLP, ANN+ Attention [3], LSTM [9], and ResNet [18].

5.6 Comparative analysis

This section elaborates the comparative analysis made by the developed ALPO-hybrid network with respect to precision, recall, and F1-measure by varying training data.

5.6.1 Analysis using easy type images

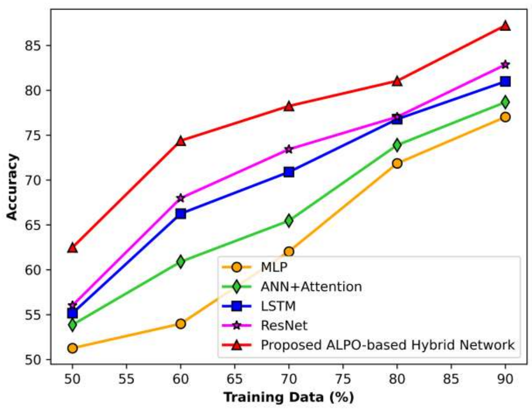
Figure 6 represents the comparative analysis of proposed ALPO-hybrid network with respect to the evaluation metrics by considering the easy type images.

The analysis of proposed scheme in terms of accuracy is illustrated in figure 6a). When the training data=90%, the accuracy obtained by the existing methods, like MLP is 77.023, ANN + Attention is 78.674, LSTM is 80.984, and ResNet is 82.856. However, the proposed ALPO-hybrid network achieved the accuracy of 87.231 that reveals the performance enhancement of proposed with that of the conventional approaches, such as, MLP is 11.7%, ANN + Attention is 9.8%, LSTM is 7.16%, and ResNet is 5.015%.

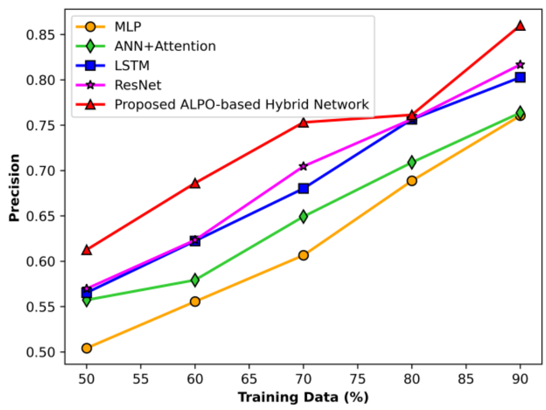
The analysis of proposed scheme in terms of precision is illustrated in figure 6b). When the training data=90%, the precision obtained by the existing methods, like MLP is 0.761, ANN + Attention is 0.764, LSTM is 0.803, and ResNet is 0.817. However, the proposed ALPO-hybrid network achieved the precision of 0.860 that reveals the performance enhancement of proposed with that of the conventional approaches, such as, MLP is 11.566%, ANN + Attention is 11.195%, LSTM is 6.661%, and ResNet is 5.027%.

Figure 6c) shows the comparative analysis of proposed scheme with respect to recall by varying the training data. If the training percentage=90%, the proposed ALPO-hybrid network attained the recall of 0.826, which shows the performance enhancement of proposed when compared with the traditional approaches, like MLP, ANN + Attention, LSTM, and ResNet is 12.988%, 6.343%, 5.976%, and 0.401%. However, the recall obtained by the existing methods, such as MLP is 0.719, ANN + Attention is 0.774, LSTM is 0.777, and ResNet is 0.823.

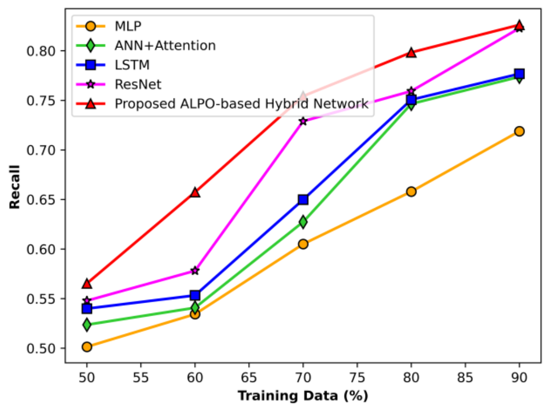
The comparative analysis of proposed ALPO-hybrid network by considering the F1-measure is depicted in figure 6d). By varying the training data to 90%, the traditional schemes achieved the recall of 0.766 for MLP, 0.781 for ANN + Attention, 0.804 for LSTM, and 0.821 for ResNet. However, the developed ALPO-hybrid network achieved the F1-measure of 0.866, which reveals the performance improvement of 11.559%, 9.782%, 7.075%, and 5.149%, respectively for MLP, ANN + Attention, LSTM, and ResNet.



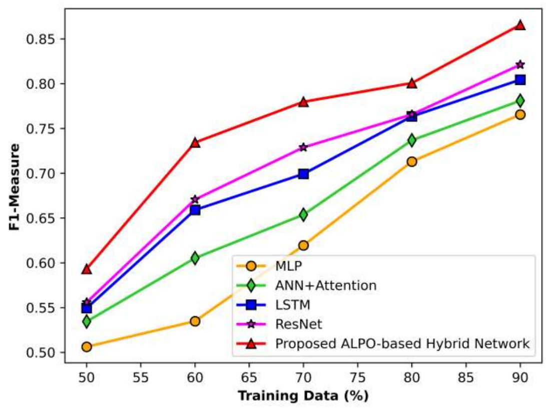
(a)



(b)



(c)



(d)

Figure 6. Comparative analysis of proposed ALPO-hybrid network based using easy type images a) Accuracy b) Precision c) Recall d) F1-measure

5.6.2 Analysis using hard type images

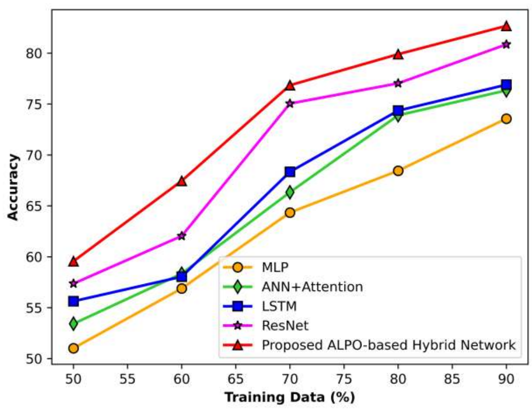
Figure 7 portrays the comparative analysis of proposed ALPO-hybrid network with respect to the evaluation metrics by considering the hard type images.

Figure 7a) represents the comparative analysis of proposed ALPO-hybrid network in terms of accuracy. If the training percentage=90%, the accuracy attained by the existing techniques, like MLP is 73.568, ANN + Attention is 76.342, LSTM is 76.894, and ResNet is 80.854, whereas the proposed ALPO-hybrid network achieved the accuracy of 82.674, which reveals the performance enhancement of developed with that of the conventional schemes, like MLP is 11.01%, ANN + Attention is 7.65%, LSTM is 7%, and ResNet is 4.100%.

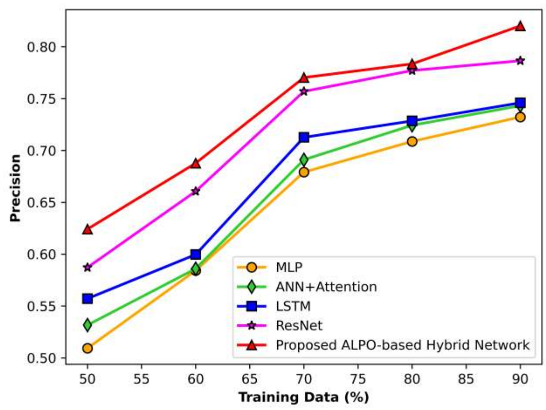
Figure 7b) represents the comparative analysis of proposed ALPO-hybrid network in terms of precision. If the training percentage=90%, the precision attained by the existing techniques, like MLP is 0.732, ANN + Attention is 0.743, LSTM is 0.746, and ResNet is 0.787, whereas the proposed ALPO-hybrid network achieved the precision of 0.820, which reveals the performance enhancement of developed with that of the conventional schemes, like MLP is 10.710%, ANN + Attention is 9.385%, LSTM is 9.044%, and ResNet is 4.100%.

Figure 7c) depicts the comparative analysis of developed scheme with respect to recall by varying the training data. If the training data=90%, the recall obtained by the proposed ALPO-hybrid network is 0.826 that shows the performance improvement when compared with the existing techniques, such as MLP, ANN + Attention, LSTM, and ResNet is 12.988%, 6.343%, 5.976%, and 0.401%.

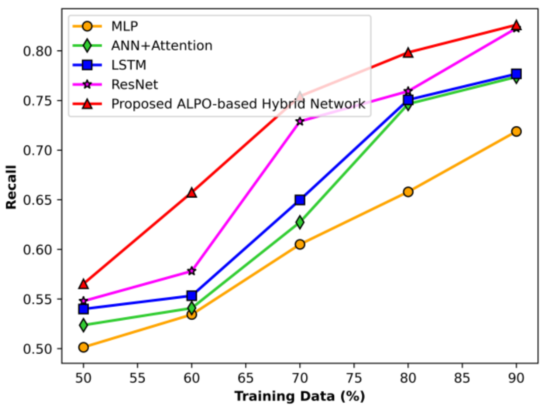
The comparative analysis of proposed ALPO-hybrid network in terms of F1-measure by varying the training data is illustrated in figure 7d). If the training data=90%, the proposed ALPO-hybrid network achieved the F1-measure of 0.823, whereas the existing methods, such as MLP, ANN + Attention, LSTM, and ResNet obtained the F1-measure of 0.726, 0.758, 0.761, and 0.804, respectively. However, the proposed ALPO-hybrid network shows the performance improvement of proposed with that of the traditional schemes, like MLP is 11.859%, ANN + Attention is 7.894%, LSTM is 7.541%, and ResNet is 2.292%.



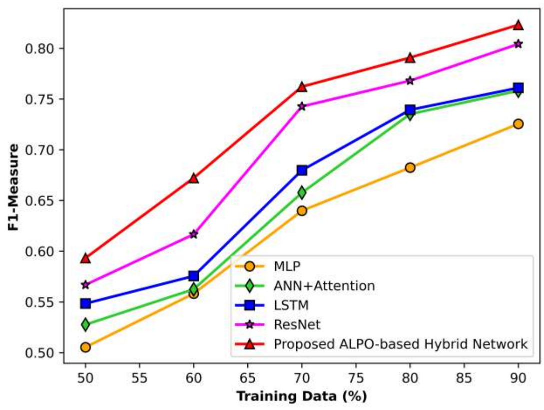
(a)



(b)



(c)



(d)

Figure 7. Comparative analysis of proposed ALPO-hybrid network based using hard type images a) accuracy b)Precision c) Recall d) F-measure

5.7 Comparative discussion

Table 1 portrays the comparative discussion of proposed ALPO-hybrid network. By considering the easy type images, the accuracy, precision, recall, and F1-measure obtained by the proposed ALPO-hybrid network is 87.231, 0.860, 0.826, and 0.866, respectively, which is maximum when compared with the hard type images. The precision obtained by the existing methods, such as MLP is 0.732, ANN + Attention is 0.743, LSTM is 0.746, and ResNet is 0.787. However, the proposed method attained the precision of 0.820 by considering the hard type images. The proposed ALPO-hybrid network reveals the superiority of the performance when considering the easy type images.

Image	Metrics	MLP	ANN + Attention	LSTM	ResNet	Proposed ALPO-hybrid network
Easy type image	*Accuracy*	77.02355	78.67454	80.98456	82.85675	87.23144
	Precision	0.761	0.764	0.803	0.817	**0.860**
	Recall	0.719	0.774	0.777	0.823	**0.826**
	F1-measure	0.766	0.781	0.804	0.821	**0.866**
Hard type image	*Accuracy*	73.56878	76.34216	76.89465	80.85467	82.6749
	Precision	0.732	0.743	0.746	0.787	**0.820**
	Recall	0.719	0.774	0.777	0.823	**0.826**
	F1-measure	0.726	0.758	0.761	0.804	**0.823**

Table 1. Comparative discussion

CONCLUSION

Recently, in many fields, such as business, academia, and industry are utilizing bar chart images to convey their information as it helps to express an individual opinions by means of analyzing the data. QA system is very effective in providing the accurate information about chart images. However, automatic extraction of data is a challenging issue in QA system. In order to address this issue, an effective strategy for QA system is introduced by developing hybrid architecture named LST-RES-NET, which is a combination of Deep LSTM and Deep ResNet. However, the Deep ResNet is trained using the proposed ALPO-hybrid network. The proposed approach consists of two phases, namely training phase, and testing phase. The bar chart and Question and Answer pair is considered as an input image in training phase, whereas the testing phase considers the bar chart image and question as an input. Finally, the results are predicted at the residual layer of Deep ResNet. Moreover, the proposed ALPO-hybrid network achieved higher performance in terms of accuracy of 87.231, precision of 0.860, maximum recall of 0.826, and maximum F1-measure of 0.866. The future dimension would be the inclusion of some other optimization algorithm in order to train the classifiers with efficient speed and also enhances the performance of the system.

REFERENCES

1. Kushal Kafle, Brian Price, Scott Cohen, and Christopher Kanan, "Dvqa: Understanding data visualizations via question answering", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5648-5656, 2018.
2. HaoWang, Xiaodong Zhang, Shuming Ma, Xu Sun, Houfeng Wang, and Mengxiang Wang, "A neural question answering model based on semi-structured tables", In Proceedings of the 27th International Conference on Computational Linguistics, pp. 1941-1951, August 2018.
3. Xiaoyi Liu, Diego Klabjan, and Patrick N Bless, "Data extraction from charts via single deep neural network", arXiv preprint arXiv:1906.11906, 2019.
4. Kim, D.H., Hoque, E. and Agrawala, M., "Answering questions about charts and generating visual explanations", In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1-13, April 2020.
5. Kafle, K., Shrestha, R., Cohen, S., Price, B. and Kanan, C., "Answering questions about data visualizations using efficient bimodal fusion", In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1498-1507, 2020.
6. Chaudhry, R., Shekhar, S., Gupta, U., Maneriker, P., Bansal, P. and Joshi, A., "Leaf-qa: Locate, encode & attend for figure question answering", In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3512-3521, 2020.
7. Sharma, M., Gupta, S., Chowdhury, A. and Vig, L., "ChartNet: Visual reasoning over statistical charts using MAC-Networks", In IEEE International Joint Conference on Neural Networks (IJCNN), pp. 1-7, July 2019.
8. Reddy, R., Ramesh, R., Deshpande, A. and Khapra, M.M., "Figurenet: A deep learning model for question-answering on scientific plots", In IEEE International Joint Conference on Neural Networks (IJCNN), pp. 1-82019, July 2019.
9. Hochreiter, S. and Schmidhuber, J., "Long short-term memory", Neural computation, vol.9, no.8, pp.1735-1780, 1997.
10. Rush, A.M., Chopra, S. and Weston, J., "A neural attention model for abstractive sentence summarization", arXiv preprint arXiv:1509.00685, 2015.
11. Bahdanau, D., Cho, K. and Bengio, Y., "Neural machine translation by jointly learning to align and translate", arXiv preprint arXiv:1409.0473, 2014.
12. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L., "Microsoft coco: Common objects in context", In European conference on computer vision, Springer, Cham, pp. 740-755, September 2014.
13. Johnson, J., Hariharan, B., Van Der Maaten, L., Fei-Fei, L., Lawrence Zitnick, C. and Girshick, R., "Clevr: A diagnostic dataset for compositional language and elementary visual reasoning", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2901-2910, 2017.

14. Mirjalili, S., "The ant lion optimizer", Advances in engineering software, vol.83, pp.80-98, 2015.
15. Askari, Q., Younas, I. and Saeed, M., "Political optimizer: A novel socio-inspired meta-heuristic for global optimization", Knowledge-Based Systems, 195, p.105709, 2020.
16. Al-Zaidy, R.A. and Giles, C.L., "Automatic extraction of data from bar charts", In Proceedings of the 8th international conference on knowledge capture, pp. 1-4, October 2015.
17. Andreas, J., Rohrbach, M., Darrell, T. and Klein, D., "Learning to compose neural networks for question answering", arXiv preprint arXiv:1601.01705, 2016.
18. Chen, Z., Chen, Y., Wu, L., Cheng, S. and Lin, P., "Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions", Energy Conversion and Management, vol.198, p.111793, 2019.
19. Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Zitnick, C.L. and Parikh, D., "Vqa: Visual question answering", In Proceedings of the IEEE international conference on computer vision, pp. 2425-2433, 2015.
20. Acharya, M., Jariwala, K. and Kanan, C., "VQD: Visual query detection in natural scenes", arXiv preprint arXiv:1904.02794, 2019.
21. Agrawal, A., Batra, D. and Parikh, D., "Analyzing the behavior of visual question answering models", arXiv preprint arXiv:1606.07356, 2016.
22. Zhong, Z., Li, J., Ma, L., Jiang, H. and Zhao, H., "Deep residual networks for hyperspectral image classification", In IEEE international geoscience and remote sensing symposium (IGARSS), pp. 1824-18272017, July 2017.
23. Ismael, S.A.A., Mohammed, A. and Hefny, H., "An enhanced deep learning approach for brain cancer MRI images classification using residual networks", Artificial intelligence in medicine, vol.102, p.101779, 2020.
24. Zhao, Y., Yang, R., Chevalier, G., Xu, X. and Zhang, Z., "Deep residual bidir-LSTM for human activity recognition using wearable sensors", Mathematical Problems in Engineering, 2018.
25. Zhu, J., Pande, A., Mohapatra, P. and Han, J.J., "Using deep learning for energy expenditure estimation with wearable sensors", In IEEE 17th International Conference on E-health Networking, Application & Services (HealthCom), pp. 501-506, October 2015.
26. Song, H., Thiagarajan, J.J., Sattigeri, P. and Spanias, A., "Optimizing kernel machines using deep learning", IEEE transactions on neural networks and learning systems, vol.29, no.11, pp.5528-5540, 2018.
27. Karpagam, K. and Saradha, A., "A Hybrid Optimization technique for effective document clustering in Question Answering system", ICTACT Journal on Soft Computing, vol.7, no.3, 2017.
28. Putra, S.J., Mantoro, T. and Gunawan, M.N., "Text mining for Indonesian translation of the Quran: A systematic review", In IEEE International Conference on Computing, Engineering, and Design (ICCED), pp. 1-5, November 2017.
29. "DVQA: Understanding Data Visualizations via Question Answering", accessed on May 2021.
30. Vidyadhari Ch, Sandhya N, and Premchand P., "A Semantic Word Processing Using Enhanced Cat Swarm Optimization Algorithm for Automatic Text Clustering", Multimedia Research, Vol.2,no.4, pp.23-32,2019.
31. Amolkumar Narayan Jadhav,Gomathi N, "DIGWO: Hybridization of Dragonfly Algorithm with Improved Grey Wolf Optimization Algorithm for Data Clustering", Multimedia Research, Vol.2, no.3, pp.1-11,2019.
32. Verma N, De A, and Mishra A., "Bridging language to visuals: towards natural language query-to-chart image retrieval", International Journal of Multimedia Information Retrieval, Vol.13, pp.32, 2024.

HOW TO CITE: Neelu Verma*, LST-Resnet: Optimization-Based Hybrid Network Architecture For An Effective Question And Answering System With Bar Chart Images, Int. J. Sci. R. Tech., 2026, 3 (9), 243-265. https://doi.org/10.5281/zenodo.22674352