Towards Assisting the Visually Impaired: A Review on Techniques for Decoding the Visual Data From Chart Images

K. C. SHAHIRA, (Graduate Student Member, IEEE), AND A. LIJIYA, (Member, IEEE)
Department of Computer Science and Engineering, National Institute of Technology Calicut, Calicut 673601, India
Corresponding author: K. C. Shahira (kcsahira@gmail.com)

ABSTRACT The textual data of a document is supplemented by the graphical information in it. To make communication easier, they contain tables, charts and images. However, it excludes a section of our population - the visually impaired. With technological advancements, the blind can access the documents through text to speech software solutions. In this method, even images can be conveyed by reading out the figure captions. However, charts and other statistical comparisons which involve critical information are difficult to be “read” out this way. Aim of this paper is to analyse various methods available to solve this vexatious issue. We survey the state-of-the-art works that do the exact opposite of graphing tools. In this paper, we explore the existing literature in understanding the graphs and extracting the visual encoding from them. We classify these approaches into modality-based approaches, conventional and deep-learning based methods. The survey also contains comparisons and analyses relevant study datasets. As an outcome of this survey, we observe that: (i) All existing works under each category need decoding in a variety of graphs. (ii) Among the approaches, deep learning performs remarkably well in localisation and classification. However, it needs further improvements in reasoning from chart images. (iii) Research works are still in progress to access data from vector images. Recreating data from the raster images has unresolved issues. Based on this study, we observe that: (i) All existing works under each category need decoding in a variety of graphs. (ii) Among the approaches, deep learning performs remarkably well in localisation and classification. However, it needs further improvements in reasoning from chart images. (iii) Research works are still in progress to access data from vector images. Recreating data from the raster images has unresolved issues. Based on this study, the various applications of decoding the graphs, challenges and future possibilities are also discussed. This paper explores current works in the extraction of chart data, which seek to enable researchers in Human Computer Interaction to achieve human-level perception of visual data by machines. In this era of visual summarisation of data, the AI approaches can automate the underlying data extraction and hence provide the natural language descriptions to support visually disabled users.

INDEX TERMS Accessibility, assistive technology, blind and visually impaired people, computer vision, data visualisation, document image analysis, image processing.

I. INTRODUCTION
The recent advancements in Assistive Technology (AT) has revolutionised the way cognitively limited users interact with the world. Visual impairment stands out as the most limiting amongst these disabilities. Assistive technology is defined as any technology that is built to help a person with disability. The advances in Artificial Intelligence is tremendous, paving the way to autonomous vehicles. We can make efficient use of these algorithms in Assistive Technology as well, to help the visually impaired ones to enhance in the field of education, navigation and also to improve social interaction. Visually disabled persons have access to information by touch or voice. Connier et al. [1] discusses the incorporation of different assistive technology devices into the Internet of Things (IoT). A new model is introduced, linking an AT device with Smart Objects and their cloud to enhance VIP’s perception.

According to World Health Organisation, 2.2 billion people in the world have a visual impairment, at least one billion people have a near or distance vision impairment that could have been prevented or has yet to be addressed. Population growth and ageing are the expected reasons to increase the risk of more people acquiring vision impairment.

An interaction with the teachers working in a special school indicated that the blind people opting science-related higher
studies and career is less common when compared to arts and teaching. Earlier days, braille helped them in reading text. Later, reading a text or e-book by the visually impaired is assisted by voice readers and text to speech converters. In some cases, when they are left with confusions in speech; they use a refreshable braille display [2]. This is a tactile system which changes the raised pins according to the cursor pointer by the user. Largely printed tactile graphs help them in map reading and give them a feeling of simple figures. Conveying the textual content have reliable solutions, where as the solutions for conveying the graphical contents, especially the chart contents in a pdf or other documents are yet to be found.

A. MOTIVATION
“A good graph forces the reader to see the information the designer wanted to convey” [3]. Studies are underway to derive the underlying information in the graphs as the brain understands it. This is an era of visual summarization of data such as the Covid-19 data. Human-level understanding of these curves by machine learning algorithms will eventually support the disabled people in understanding them.

B. CHART COMPREHENSION
The qualitative information in any document can be represented as charts. They can be part of textbooks, newspapers, scientific studies and other statistical studies. The problem of understanding charts by the visually impaired and the state-of-the-art methods in identifying the chart types is studied in [4]. The charts or any information is accessible to the visually disabled in the form of touch, audio or both. Document image analysis [5] deals with the processing of textual and graphical content in a document image. The graphical data can be natural scenes, tables, flow charts, lines and delimiters or charts. Substantial progress is seen in classifying the textual and non-textual information in a document, but the same cannot be said regarding graphical data, especially the chart information extraction. The textual content in images, whether handwritten or printed, can be converted into text with the help of OCR (Optical Character Recognition). However, this method does not work well for the chart images.

With the aim of bridging the gap between the biological and engineered vision in the field of document image understanding, the approaches can be divided into three classes: modality based approaches, traditional processing and deep learning-based approaches. Figure 1 gives a summary of this paper, and highlights its various classifications and applications. The modality based methods classify the related works in the perspective of mode in which the visually impaired perceive the information. In traditional methods, it is not easy to manually select the features and is time-consuming while deep learning counterparts automatically select the features. It is easier to process the enormous volume of data with better accuracy utilising the object detection algorithms. A document can contain textual and graphical contents, processing of which have to be dealt in two different ways: (i) textual processing and (ii) graphical processing

The textual processing has a growing demand due to applications like reading street board/ signs in autonomous driving, converting any language text to any other language text with smart phones (Neural Machine Translation). The textual content in any document needs to be localised first and then can be extracted by OCR like Google Tesseract [6] or using a python wrapper for Tesseract, pytesseract. Optical Character Recognition (OCR) is a solution to all the hurdles in text detection and recognition with 99% accuracy on documents. The textual elements can be part of the figures which can be captions, text embedded as product name or number plate. In such a case, the goal is to determine if the image has a text, present in it or not. If present, it needs to be detected, located and recognised. These text and graphics are logically or semantically related and this is identified by Huang et al. in [7]. Associating the textual content with its graphical parameters is an essential part in semantic labelling of graphics as put forth by Tombre and Lamirov [8]. They propose the use of vector-based signatures on the graphical part of the document for indexing it along with the text to help in retrieval and browsing. They discuss the various open challenges and the possibilities of the same. The textual areas in a document are considered as the horizontal grey-level variations which are evident in the histogram of the cumulative gradient matrix. The cumulative gradient is classified into two classes- textual and figures, by applying K-Means clustering. It is followed by a connected component analysis to find the text, in Doung et al. [9]. Though the text detection is an extension of document analysis, traditional text recovery could not solve all the challenges like video scene text [10].

In graphical processing, once we have a figure, it needs to be classified into different types. The graphical contents can be images with a detailed description or can be other scientific plots, tables and charts. Text recognition from a natural scene deals with recognising the text in an image and it has wide applications like road sign detection in autonomous cars. Table Question Answering [11] deals with understanding and parsing tables and making inferences from them. The analysis of any experiment can be in the form of a mixed style plot; it takes time for interpretation even by a human. The figures may have textual content around them. Clark et al. in [12] identifies the region as text, figures and captions around chart images by region identification. The captions and textual contents around figures are extracted by them as in [13]. Davila et al. [14] observe extraction of chart data into two steps: document segmentation and linking figures to captions.

The images and graphs should be translated into textual representations for helping the visually impaired. These texts are represented as alternate text (alt-text) which are read by screen readers like- Jaws in Windows and Voiceover in Mac OS. The true meaning or information in the image should reach the blind or the disabled correctly. Some browser extensions like [15] provide textual alternatives for graphical
information. This gives the textual description of graph elements but not based on their meanings, and therefore screen reader misses the significance of that graph and the data which it represents [2]. Besides, there are a variety of systems allowing the user to retrieve the data manually. This is typically the case with Graffit [16], that enables users to click on data points and labels to scale. These systems continue to be very time consuming, particularly for graphs which have a lot of information.

Another visual disability is colour blindness. Colour blindness is the inability to differentiate certain shades of colour. Accessibility of the colour blind ensures that only colour should not be a distinguishing feature for data perception. The works [17]–[20] focus on improving the accessibility of colour blind people. Wearables like the Enchroma glasses [21] support them to differentiate between different colours and such a wearable can assist them with a chart image too.

C. RESOLVING FOR VARIOUS GRAPH FORMS

Graphs share similar pieces of structure. The axes, labels, scales, title, X-axis title, Y-axis title, legends etc. describe or and provide guidance in measuring the data. A simple bar graph has L shaped axis, simple rectangular boxes or bars and not necessarily a legend and its corresponding symbols. A line plot/scatter plot also have L shaped axis, ticks and legends. A pie chart is slightly different from this as the perimeter of the arcs corresponds to the quantity of value it holds. To understand any graph, the following requirements need to be satisfied.

- Parsing of axes to understand position, label and scales.
- Comprehending the label which is associated with the axes. This gives the name of the axis (e.g., Year on X-axis, Price on Y-axis).
- Extracting the axis tick labels. The axis tick labels have the values, that may be linear or scaled.
- Extracting the textual content in the graph image. Specific pre-processing followed by an OCR can meet this.
- Systemic structural understanding like, total number of bars. How many bars are grouped in a multi-bar graph?
- Associating the legends with the same coloured bar/curve for data retrieval.
- Solving the variety of chart images. Graphs have variants among themselves. For example, a donut chart (pie chart), stacked bar chart, grouped bar chart and many others.
- Identifying the role of legends. This is a guide to symbols used when plotting multiple variables. There can be multiple ways in which legends can be placed and its format also can vary. The legend itself can be placed on anywhere on the centre, rightmost or at the bottom of the figure.

Though the graphs share similar structure, they all represent unique data/statistics. This necessitates an end-to-end pipeline which can understand all kind of chart images and extract the data from it. The data extracted can be used to summarise the figure-by-figure descriptions as alt-text, to support the low vision or blind people. The data extracted can also be used for visual reasoning and for summarising the statistical data represented by scientific charts. Baucom et al. in [22], provide a web interface that extracts the data from the scatter plot and produce a .csv file as output file or allow the user to modify the design of the chart. Cliche et al. [23] introduced a fully automated system using deep learning for data extraction from the scatter plots. The work [24] deals with recognising lines without markers while [25] and [26] extracts data from the line charts. A few others [27]–[31] deal with understanding pie charts.

The reconstruction of the bar in a picture was considered by a few works but the extraction of text and data remains untouched. Some of the related works listed in Table 1 gives an insight into the relevant works and the gaps identified. Microsoft’s FigureQA dataset [32] proposes a new analysis challenge for visual thinking, unique to graphic graphs and statistics. FigureQA can be considered as state-of-the-art in the Visual Question Answering in the bar graphs. This gives an answer to a set of questions as specified by a template and this need to do more reasoning to include the real data. The work [26] does a good amount of visual reasoning over the real data. In 2020 Kim et al. [33] introduced a pipeline for generating answers and visual explanations for an input chart and its corresponding question. Table 1 shows the possibility of further research in other types of graphs, including mixed/complex graphs.

D. CONTRIBUTIONS

The purpose of this paper is to perform a detailed survey of the current literature and state-of-the-art methods for graph comprehension. The accessible research has been minimal in the last few years, but there is a recent uptrend. By analysing the significant studies from 2016 to 2020,1 we find that the recent hike is attributed to deep learning. This work provides insight into the following:

- We categorise the decoding of graphs based on the approaches as traditional or deep learning-based.
- We also categorise based on the domain and discuss the key techniques used in each of these domains.
- This gives an insight into the available datasets for graph comprehension related research.
- We discuss and compare different approaches, challenges and specify future extensions.
- We discuss the possibility of AI in decoding the chart data to attain the human level perception.

E. BENEFITS

This review will support the research in automating the extraction of chart data. Although storytelling on geographic maps is popular like NewsViz [44], generating automatic

---

1 The review dataset is made available for further studies at: http://dx.doi.org/10.17632/whyk3bc9h6.1#folder-9d3c9262-13e3-4303-90c8-c80ce6391013
TABLE 1. Solutions introduced in the chart data understanding and future possibilities from few related works.

| Author, Year | Methodology | Merits | Demerits |
|--------------|-------------|--------|----------|
| Kim et al. [33], 2020 | Rule Based Approach. | A dependency parser for NLP to generate explanations that can convey how the pipeline obtained an answer to the chart specific questions. | Explanations to specific type of charts and lack fluency. |
| Methani et al. [34], 2020 | Graphical elements extraction using Faster-RCNN [35] and text extraction by OCR. Reasoning by Attention Networks. | A hybrid model that can do reasoning on complex questions. | More accurate visual content detection and reasoning required. |
| Charles et al. [36], 2019 | An encoder-decoder architecture for figure captioning | Generation of labels in figures | Relationship between labels is a challenge. |
| Liu et al. [29], 2019 | Data extraction from the bar chart and pie charts by the RCNN. | Simulated data gave better result ie, 79% for Bar charts and 88% for pie charts. | But for other dataset(FigureQA) and images from Internet, showed poor accuracy. |
| Choi et al. [37], 2019 | Object detection and text recognition using deep learning pipeline. | Automatically decode charts in webpage. | 49.7% text classification errors and text localization errors. |
| De et al. [38], 2018 | Data extraction from the pie chart is by extracting local gradients of the slice data. | Fails to extract when the adjacent slices have close gray values or fails to analyse the common boundary. | Only the pie chart is analysed. The textual data and the labels are not analysed, with accuracy of 76% for identifying chart regions using MobileNet. |
| Balaji et al. [39], 2018 | MobileNet for classifying the chart types and text extraction from charts by OCR. | Data Extraction specific for each type of charts, 78% accuracy. | Label analysis is not done. |
| Poco et al. [40], 2017 | Deep learning pipeline (CNN) in combination with traditional Connected Component Analysis for text localisation. | Bitmap images of the bar chart and pie charts are taken as input. | Legend Analysis not done. Not able to deduce scale mappings for visual channels such as color, shape and size. |
| Kahiou et al. [32], 2017 | Reasoning on fifteen templates. First model in recognising visual patterns of chart data. | Data Extraction and Question Answering based on fixed template. | Binary answers and no numerical output. |
| Siegel et al. [30], 2016 | Graph based reasoning using CNN. | Analyse figures by training each components independently. | Line charts only. Cannot handle labels in multiple lines. |
| Zaidy et al. [41], 2015 | Graphical component extraction by Connected Component Analysis. Text component identification by OCR. | Not detecting Y axis text, 87% of the time. | Data Extraction only for simple bar charts, with many heuristics. |
| Savva et al. [42], 2011 | Learning low level image features to classify charts using SVM. | Redesigning the graphs to improve perception. | Manual approach is used in annotating textual areas. Legend parsing is not done. |
| Zhou et al. [43], 2000 | Hough Transform detects the bars in images. | Detects even hand drawn bars. | Data extraction from the bar chart is not considered. Legends and text are not extracted. |

image descriptions for chart visualisation requires indispensable contribution from AI algorithms. The study described in Section 2C illustrates the diverse approaches of deep learning techniques and the problems faced by them. The extracted chart data can be provided as alt-text to the visually impaired people and enable reasoning on these charts. Today, machine learning algorithms generate visualisation of big data. In future, machines may interpret the data which was used to generate these visualisations.

The remainder of this paper is organised as follows: Section 2 reviews the various approaches in understanding the charts in traditional ways and deep learning methods, and modes by which the visually impaired people perceive the charts. Section 3 discusses the available datasets, Section 4 explains the commonly used evaluation metrics. Section 5 enlists the different applications, future work and finally, Section 7 concludes the paper.

II. METHODS FOR IDENTIFYING CHARTS

Extracting data from the chart images has been extensively studied recently. Some works concentrate on recognizing the figure in a document as a graphical image [45] or not, while others classify it into various types of charts [4], [28], [29], [39], whereas others concentrate on extracting the visual elements from the charts [26], [40], [46]. We categorise the methods in extracting the underlying data as modality based methods, traditional and deep learning-based methods. Figure 1 gives a summary of this paper, as well as the study’s various classifications and applications. We explore how the diagrams are made available to the blind and visually disabled people in modality based solutions. The traditional based methods deal with various image processing algorithms to segment the graphical data. These depend on human-defined values. Methods that aid in the automated learning of features are categorised as deep learning methods. Our taxonomy, however, spans both traditional and deep learning approaches and goes beyond addressing challenges and future enhancements.

A. MODALITY BASED APPROACHES

‘Seeing’ improves perception and the human visual system helps a sighted person to perceive the information given in any image. In Assistive Technology, blind people are open to information access with the help of other senses. Modality
refers to how something is experienced or expressed. Often the word modality is associated with a particular form of sensory perception, like vision or touch. A crucial Assistive Technology for Visually Impaired People is screen readers, which converts the text into audio or braille output. However, they do not work well for data-rich pictures. The screen readers read aloud the text (alt-text) given for image as their description. Many research works convert graphic data into a format that the blind can easily interpret: as haptics (tactile) or audio format.

1) HAPTIC OUTPUT
The tactile or haptic based systems deal with the perception through vibration or by force feedback from them. Yu Brewster et al. evaluates the comprehension of the graph by blind people in [47]. In this, the charts in the digital documents are converted into a haptic model using a SenSable Phantom and a Logitech WingMan force feedback Mouse. The pointer on the screen can be controlled by the user. They can also feel the force initiated by these devices. Moustakas et al. transform the 2D map image or 3D video of the maps into haptics, and audio output in [48]. The hurdles of cross-modal transformation of visual data into haptic is solved in this work. Kim and Lim [49] proposed an assistive device, Handscope, which can translate a statistical graph into tactile feedback. Another way to perceive chart data is through tactile charts itself. They have raised elements to perceive through touch. Different textures represent the colour information and labels. Engel and Weber [50] proposed improved guidelines for charts and studied how design can improve the readability of tactile charts. Research towards developing tactile graphs is advancing as in [50]–[52]. Further studies regarding the practical design of tactile charts are beyond the scope of this review.

2) AUDIO OUTPUT
The speech output or non-speech output like music falls in this category. Alty and Rigas [53] used an AUDIOGRAPH in which they had used music to communicate complex graphical images to the VIP. This paper discusses the various implications of using music for Aural interface design. A better understanding of the contents is possible by the audio representation. Microsoft Speech SDK 7.0 is used in [47] which produces both speech output and also uses the musical note in which high pitch corresponds to a higher data value and a low pitch amounts to lower data value. In [48] they use text to speech (TTS) converters. Audemes, a non-speech sound is used in [54], as a medium for learning by the visually impaired people. Use of different sound types-music and audeme together is observed to give better recognition in [55] than using a single or combination of the same sound type.

All of the above methods were successful in the case of line charts and simple bar charts. The methods were tested on blind people. They were asked to draw the graph on a paper to know if the mental sketches match with the graph. It was observed that when the number of lines increases in the chart, it was difficult to form an overview due to the narrow bandwidth on Phantom.

Also, imparting precise data values to them was challenging. However, combining auditory feedback with haptics was better than providing the output in one modality alone. Meanwhile, for the bar graphs, the overall trend of the data could be imparted. But when the bars are closely placed or when the height between the related bars are near, it is challenging. The multi-modal methods are not capable of solving complex and clustered graphs.

3) DISCUSSIONS
Haptic based approaches are based on friction keys. In haptic line graphs, the friction could hinder the movement of the line when the line has sharp bends or strong friction. This attributes to wrong judging of these areas as the edge of the line [47]. Perceiving information through any other mode is slower when compared to perception by vision. Incorporating audio feedback into the haptic interface is essential because missed information in one mode could be dealt with another mode. Too much information cannot be presented by haptics because the narrow bandwidth can be overloaded very fast. An enclosed rectangular area simulates a bar chart. This can be used to describe the overall trend of the data, like the highest and lowest bars. Audio seems to give effective results in this. Using the multimodal system is much better than any of the single-mode. A summary of the essential techniques used for the chart perception by the blind using modality based approaches is given in Table 2.

B. TRADITIONAL METHODS
In this section, we address the traditional methods known in the extraction of chart data. The visual content belongs to the image’s foreground, and the representative approaches used to segment these graphic elements—Connected Component Analysis and Hough transforms [56] are discussed.

A connected component analysis is used to separate the text and graphical contents in document image analysis. We categorise the Connected Component Analysis based approaches as traditional because they use the classical algorithms for image analysis. Connected Component Analysis is used for identifying different objects in a binary image based on the assumption that the objects or shapes belong to the foreground of the image. A group of pixels not connected by a boundary is said to be connected. Each maximal region of these connected pixels are called connected components [57]. It uses 4-way or 8-way connectivity; the common practice employed is 8-way connectivity for black pixels. Merging the connected components is done by taking the Euclidean distance between them. To establish the adjacent relation between the components, Minimum Spanning Tree or Delaunay Triangulation [58] can be used. This segments the foreground and the background regions, and further connected component labelling is used to label the pixels of the foreground regions. In this, the shape and position properties of each connected component are measured.
The chart image’s visual contents are of different shapes, and the works that segment these shapes, such as lines, bars, and pies are reviewed. In 2018, connected component analysis is used in [38] for detecting the pie charts. The total number of pixels after removing the boundary components and each connected components helps to find details of the pie chart. They make use of this technique to analyse if the chart is 2D or 3D.

There is no simple, robust method in place for detecting rectangles in an image in case of bar graphs. Al-Zaidy and Giles [41] propose a method for extracting data from bar charts. The bar chart image from a document is extracted by...
a connected component method, using LAB colour space to discriminate the components. The neighbouring pixels which differ in colour are labelled as different components. The name of the bar is the text below the bar and the y-axis holds the value. This method was successful in identifying and localising the bar charts except for the small bars. The OCR used was not able to detect the text, 80% of the time and this is open for future work.

Fuda et al. [24] recognise line graphs by tracing the connected components. The image is processed for the axes detection and is divided into interior and exterior of the graph region. The connected components inside the axes are extracted for the line even in cases when multiple line charts intersect. In this, the connected component trace method can identify four different types of line graphs. When the lines are overlapping or in contiguousness to other lines, it is a subject which requires further research.

Huang and Tan [60] recognise the shapes from raster document images. Directional Single Connected Chain (DSCC) Algorithm [61] is used to detect the small straight lines on the edge image. Further, Curve fitting is used to form the lines, curves and ellipses. Besides, the Curve fitting method is used by Huang et al. [59], in which they extract the graphical information in the documents by straight-line vectorisation and extend this to other graphical entities with curve fitting. They use a data structure to keep information about the connected chains. The curvature of the content in the data structure is evaluated and the chains are classified as lines or arcs. By this, they were able to extract the line, bar and pie charts in images. This work contributes to the understanding of scientific charts.

The CCA requires feature extraction followed by classification. Balaji et al. [39] labelled the regions by connected component analysis and used a minimum bounding rectangle to fit each of the labelled regions for identifying and extracting the bars from simple bar charts. They propose an automated chart image descriptor, in which they successfully classify the chart images with 99% accuracy using an Inception V2 model. For extracting the bars, binarise the input image and further apply morphological operations to remove the smaller components. Then apply Connected Component Analysis on these regions and fit them with minimum rectangular bounding boxes and extract different bars. To find the scale of bars in the chart from the pixel coordinates, they define the chart to pixel ratio as the ratio of two consecutive x or y values extracted using OCR and the corresponding distance between the centroids of their bounding boxes. They perceive pie charts by detecting the circle first and then finding each wedge’s angle in the clockwise direction. They followed the method as in [23]. Chester and Elzer [62] explain how image processing has been used to turn a piece of visual information into an XML representation that catches all aspects of the graphic that can be important to knowledge extraction.

Another approach in finding the graphical elements is using the Hough Transform. The classical Hough transform detects the lines and later this is extended to identify circles and curves. An initial work related to this is applied in chart images by Zhou and Tan [43]. With the assumption that the bars as a pair of parallel lines, traditional image segmentation by connected component analysis is done followed by Modified Probabilistic Hough Transform in order to reconstruct the bars in images. It is successful in identifying the presence of bar charts in documents and also in handwritten images. This method could also find skewed bars. The legends and text in the picture are left unextracted, and this does not deal with extracting data. The coordinate line detection is discussed in [63], [64]. The connected component analysis is used in [65] to separate text and graphical areas in preprocessing followed by new bar chart recognition algorithm, using syntactic segmentation based on Hough. In document images, it can recognise several bar charts like oblique bar charts. Table 3 summarises the traditional methods in chart understanding.

1) DISCUSSIONS
The traditional methods are based on human defined values and assumptions. Most of the methods mentioned above for solving graphs were successful in getting the structural property of a chart. But the extraction of underlying data is not considered, and the following impede this:

- Fixing the threshold for identifying the structures like bars and lines with a varying image quality.
- Extraction of text: It is difficult to distinguish the text from the saturated background color and also the text that is part of the chart.
- Distinguishing the clustered bars is difficult in the case of bar graphs, and almost all of these fail for the complex graphs which are a mixture of different graphs.
- How to semantically relate the text and the image?
- Solving for a real graph is more challenging than a synthetic one.

The classical methods alone cannot solve the problem discussed above. Use of learning based algorithms along with the traditional methods can find a better solution. Two different kinds of scientific chart images were identified with HMM (Hidden Markov Model) in [66]. A HMM is a sequence classifier and assign labels to each unit in a sequence. By this, they could classify the images into bar and line charts. The SVM also classifies the features extracted by traditional methods as in [12], [28], [66]. They syntactically analyse the structure of the chart and use the spatial relationship between the text and the graphical elements in chart, to find the meaning of the charts.

C. DEEP LEARNING BASED APPROACHES
Graph comprehension has been studied over the years. However, the key challenge here is in extracting underlying data and the corresponding legend entries. The conventional methods have been able to define the structural properties such as bar or line recognition, and OCR managed the extrac-
tion of text. However, the extraction of the underlying data from these informative graphics requires association of the legends and also associating the legends with the structural components and text extracted. Deep learning models can do this better because they usually outperform the traditional methods [67].

With the deep learning models’ progress, the roles of Computer Vision tasks involved in the future of Assistive Technology (AT) are beneficial. Leo et al. [68] categorise the Assistive Technology domains and the associated computer vision tasks in-depth in terms of consumer needs.

Deep learning models have the advantage of automatically learning the features, and the user no longer needs to feed the handcrafted features. Multi-Layer Perceptrons (MLP) are the quintessential deep learning model. An MLP has an input layer, a hidden layer and an output layer. Deep Neural Networks (DNN) is an MLP with more than one hidden layer. It learns features either by following supervised methods or unsupervised methods. In the supervised method, the labelled data is used for training and image classification problems. The unsupervised methods take in the input images, produce a compressed image with essential features and reconstruct the image from this compressed image as in Auto Encoders and Boltzmann machines.

Any deep learning-based object detection needs image pre-processing, feature extraction, classification or regression. Inspired by the reception of human visual cortex, Convolutional Neural Networks (CNN) [69] is designed. This has convolutional layers, pooling layers, and classification layer at the end. Firstly, images from the dataset is given to the model as input. Input images undergo pre-processing like resizing and remove noise (if any). We can augment the data by affine transformations or adding Gaussian noise. Data augmentation increase the training set. Generative adversarial networks (GAN) can be used for data augmentation as it can generate new images to enrich the diversity of inputs. Dataset is split in 80:20 ratio, where the training happens on the bigger portion and validation on the latter portion. Feature extraction is the key to detection. The convolution layers in CNN produces the feature map. The feature map is fed into a fully connected layer. The output layer at the end of the fully connected layer is responsible for proposing and refining the bounding box or producing the classification scores. Figure 2 illustrates the general architecture of a deep learning model similar to AlexNet [70].

The first step in extracting data from the given chart images given any document is extracting the figure, followed by identifying the type of chart. Extracting the chart data involves figure localisation, classification, text recognition and reasoning. The classification models can range from a neural network to deep learning model. Table 6 enlists the various deep learning models and traditional methods used for chart classification. Table 6 also compares few models deployed till date for the object detection, localisation, classification and reasoning in decoding the graphs. The various steps in data extraction can be generalised as in Figure 3.

At each step, different deep learning models can be used to make an end-to-end pipeline to extract the data, and produce figure descriptions and do reasoning to support the visually impaired people.

The Reverse Engineering aims to extract and understand the data, and this can be broken down into respective subsections, and we discuss deep learning approaches specific to each subsection.

1) LOCALISATION AND CHART TYPE IDENTIFICATION
In the context of this review, localisation is finding the chart image and taking it out from the document. Object localisation is a well-studied problem in computer vision. However, localising the figures in a document, specifically the chart images, is not well studied as others. Localisation deals not only with object detection but also include labelling it along with the bounding box around it. One of the most popular deep learning models in this is CNN with a deep CNN beating the state-of-art in ImageNet image classification [71]. A few pre-trained CNN on 1.2 million images from ImageNet [72] are AlexNet [70], VGG [73], Googlenet [74] and Resnet [75]. Any of them can be fine-tuned towards a specific type of classification task. Classifying the figures in documents into different chart types using CNN, which is similar in structure to an AlexNet [70] is used in [4]. Motivated by the positive results of a classification in natural images by CNN, Table 4 summarises the various network architectures designed and the pre-trained models used for recognising the charts images. The extracted image can be of different chart types and classification of these images by CNN is used in [45], [76].

Researchers are recently focusing on applications of chart images especially in the scholarly papers and use them in improving search engines [12], [13], [25], [77]–[79]. Siegel et al. [25] introduced a neural network for figure extraction called DeepFigures, in support of academic search engines. Figure localization [80] is a wide area of study in Computer Vision. Once the image is localized, it needs to be extracted.

Under the assumption that the image is extracted and classified as a chart image, the image analysis needs to be done further. For this, we have to detect all the elements in it, and this can be done by object detection [81] whose goal is to find a small set of object proposals that cover most of the objects of interest in a given image. The objects of interests here are all the visual elements in a chart image. Most recently, due to the availability of large training datasets and the advancement of high-performance GPUs, many object detection models outperform the state of art methods. A few of them are, Fast R-CNN [35], Faster-RCNN [82], YOLO [83], YOLO v2 [84], SSD [85].

You Only Look Once (YOLO) achieves real-time object detection by dividing the whole image into multiple bounding boxes and class probabilities for each bounding boxes. However, in this, it is difficult to detect small-sized objects. YOLO suffers from localisation errors which are handled by YOLO9000 [83] and YOLOv2 [84]. Choi et al. [86], uses
YOLOv2 to detect the bars in bar charts. This model identifies the bars from bar charts from left to right and predicts the top left and bottom right coordinate positions.

Single Shot Detector (SSD) [85], is proposed for improving the YOLO method, and it has achieved substantial improvement in performance. It detects smaller objects and increases the accuracy of localisation compared to YOLO. Methani et al. [26] in 2019, proposed a visual extraction module based on SSD.

Apart from the above object detection models, Detectron [87] is the Facebook AI Research software framework that implements the state-of-the-art object detection algorithms, including Mask R-CNN [88]. Methani et al. [34] used the Faster-RCNN along with Feature Pyramid Network [89] using the Detectron framework. It performs better to detect all the visual elements in a chart image.

2) LEGENDS AND TEXT EXTRACTION

Extracting a rich set of visual features based on the image content can be handled as explained in the above section. The visual contents in an image can be textual or graphical. Obtaining the visual elements in a graph can be solved by object detection followed by classification i.e., localisation. Data extraction deals with extracting the exact values from the charts. The labels in the chart are vital as they convey the semantics of the chart. The values along the X-axis and Y-axis holds the data, that can be actual data or scaled value. The title, X-axis name and Y-axis name add more details to the information. Hence, we should extract these texts and understand their role in chart understanding. Localising the texts in chart images and classifying their role is discussed in the following sections.

Associating the textual and graphical content in a graph is important to understand the scientific charts. Huang et al. [7] located the textual components in the image with the help of OCR. They consider the correlation of textual information with graphical information from both logical level and semantic level. The logical connection between text and graphics is obtained by analysing the spatial relationship between text blocks and chart elements. Text performs a variety of logical roles in a chart image and this can be summarised as: caption, X-axis title, X-axis label, Y-axis title, Y-axis label data value if in case the data is directly shown. The semantic association of text and graphics is challenging, and many heuristics are considered. Eventually, the data collected is expressed in XML.

Text recognition involves text detection, followed by identifying the detected text. Most of the text detection algorithms are deep learning-based. The text localisation and recognition are followed by OCR [6] for extracting the textual content. Text region is localised by PixelLink [93] in [86]. It uses the deep learning model VGG16 [73] as a feature extractor. The text recognition can be done by Tesseract as in [6] or by using Convolutional Recurrent Networks (CRNN) as in [94].
They found that CRNN easily recognised the Y-ticks than Tesseract.

Pocco and Heer [40] extracts the text from the documents and classifies the contents as the body of a document or figure text. The main focus is to extract chart figures from documents and apply several heuristics to identify the figure text, i.e., text which are part of the figure descriptions. Support vector machines (SVMs) are used to identify the text role (as in X-tick label, Y-tick, Y-label, caption) in the chart pictures. The bounding boxes and corresponding geometrical details define the feature vector for each text element.

3) REASONING

Getting bounding boxes around the text, retrieving the textual data in them and the recognition of the visual elements are discussed in previous sections. The reasoning at a higher level should be able to find all the visual relationships between chart elements, underlying data and text. Visual Question Answering (VQA) is successful in recognising patterns from visual representations of the data. Antol et al. in [100] proposed the VQA challenge in which a natural language question q for an image I, produces a natural language output or an answer, a. This requires vision, language and commonsense to answer correctly. Most of the works solving this challenge gave priority to linguistic priors. Kahou et al. [32] introduced a visual reasoning corpus, which includes questions from fifteen templates for reasoning on chart images. This question includes the maximum value and minimum value of the chart, and the area under the curve etc. The answers to these questions are binary (yes/no). But several studies on VQA show that a bias in the dataset can affect model performance evaluation by exploiting the statistical patterns [101], [102]. Handling the vocabulary out of a fixed dictionary and dynamic encoding is complicated in natural images. Experimentally, Kafle et al. [46] have shown that the VQA algorithms are capable of answering simple structure-based questions, and that their data recovery efficiency is low. They use a multi-output model which has a dual network architecture, with two subnetworks- one responsible for generic answers and the other responsible for chart specific answers. It is possible to use attention networks and relation networks for answering questions related to an image, and they are discussed as follows.

a: ATTENTION NETWORK

The existing model for VQA in charts which are discussed above, deals with answering the questions with a fixed vocabulary. The reasoning questions should be able to collect all the information regardless of the fixed template of questions. This deals with creating meaningful question-answer pair. These models use pre-trained CNN like VGG or Resnet or any other object detectors to extract the features and language encoders based on LSTM. The few models used in the upcoming researches are compared in Table 5. An IMG model passes the image through VGG19 to predict the answers from the vocabulary. The QUES model passes the text through LSTM to predict the answers from a fixed vocabulary. The IMG+QUES is a combination of LSTM and CNN followed by an MLP that produce language output. The VQA requires multiple steps of reasoning. These models also make use of attention mechanism. In attention networks, to formulate the answer from an image, first-locate the objects and give attention only to the indicative ones to infer the answer.

The SAN model is an autoencoder with stacked attention mechanism [90]. A multi-layer SAN learns natural language processing questions from images, and this is an extension to the attention mechanism. This involves CNN and LSTM, which use query representations to locate the relevant regions in the image and select answers from the vocabulary. The reasoning is done with relation networks in [32]. But they did only for Yes/No questions. The reasoning beyond the fixed vocabulary of words is possible by attention mechanism. The reasoning by attention mechanism used in chart related querying is by SAN, SAN-VOES and SAN-VQA. Table 6 highlights the attention network used for reasoning module in the recent works.

SAN or Stacked Attention Network calculates the weighted sum of the image vector for each image region and appends with the query vector. Compared to other models, attention mechanisms construct this vector by giving higher weights on visual regions that are more appropriate to the query. A variant of SAN [91] is SAN-VQA, used by [46] where a small upgradation to the image features produced a remarkable improvement in accuracy. The SAN-VQA performs well for the binary query and performs poor for the reasoning questions in case the answer is not from the fixed

### TABLE 4. Qualitative analysis of different model performances(accuracy) in reasoning.

| Author, Year       | Dataset | Type       | IMG     | QUES     | IMG+QUES | SAN     |
|--------------------|---------|------------|---------|----------|----------|---------|
| Methani et al. [26], 2020 | PlotQA  | BC, LC     | 14.84%  | 15.35%   | 46.54%   | 53.96%  |
| Kafle et al. [46], 2018    | DVQA    | BC         | 14.83%  | 21.06%   | 32.01%   | 36.04%  |
| Kahou et al. [32], 2017    | FigureQA| BC, LC, PC | 32.47%  | 50.01%   | 55-59%   | 72%     |

### TABLE 5. A comparison between various models used in decoding the data form chart images.

| Author, Year       | Text      | Image     | Text+Image | Reasoning               |
|--------------------|-----------|-----------|------------|-------------------------|
| Methani et al. [26], 2020 | LSTM     | VGG19     | CNN + LSTM | SAN [90]               |
| Kafle et al. [46], 2018    | LSTM     | Resnet    | CNN + LSTM + MLP | SAN-VQA [91]          |
| Kahou et al. [32], 2017    | LSTM     | CNN       | CNN + LSTM | Relation networks [92] |
vocabulary. SAN-VOES is another variant used by [34]. They propose VOES to answer open vocabulary questions. This VOES pipeline with SAN-VQA forms a hybrid model called SAN-VOES. The accuracy comparisons of all these versions of attention networks on the chart images are given in Table 5. This is a pipeline of visual perception and question answering modules. Interpreting information from the visual elements of the plot images is a difficult task, and more research is required.

b: RELATION NETWORK
A relation network is a powerful network for reasoning introduced by Santoro et al. [92]. The same network for

| Author et al. | Year | Models | Purpose | VBC | HBC | LC | SC | PC | FC | Dataset |
|---------------|------|--------|---------|-----|-----|----|----|----|----|---------|
| Kim et al.    | 2020 | Explainable AI, Rule based approach using dependency parsers | Answer the question about charts | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | NA      |
| Chaudhry et al. | 2020 | Mask R-CNN LSTM SAN | Classify Answer Encoding Reasoning | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | LeafQA  |
| Methani et al. | 2020 | Hybrid model | Visual and text extraction Structured information extraction | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | PlotQA  |
| Chen et al.   | 2019 | ResNet Attention Networks | Figure captioning Reasoning | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | FigureQA|
| Revanth et al. | 2019 | FigureNet CNN with two layer LSTM | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | FigureQA|
| Choi et al.   | 2019 | CNN YOLO PixelLink [93] | Classification Detection Text extraction | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | SynthText [97] FigureQA |
| Liu et al.    | 2019 | CNN Relation Networks CRNN | Classification Data Extraction Inferenceing | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | NS      |
| Sharma et al. | 2019 | LSTM CNN-LSTM Attention Networks | Question encoding Visual features extraction and encode questions Reasoning | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | Synthetic dataset made by the authors. |
| Kafle et al.  | 2018 | CNN Relation networks LSTM | Classification Reasoning Text extraction | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | DVQA    |
| Balaji et al. | 2018 | CNN MobileNet | Classification | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | Dataset created by authors using chart images from the web |
| Siegel et al. | 2018 | CNN-ResNet 101 | Figure Localization | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | Dataset created by authors from research papers |
| De et al.     | 2018 | Chart Identification (RCNN) | Data Extraction | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | NS      |
| Kabou et al.  | 2017 | CNN Relation Networks LSTM | Classification Reasoning Text Classification, Extraction | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | FigureQA |
| Siegel et al. | 2016 | CNN-ResNet50 | Localization Classification | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | FigureSee |
| Liu et al.    | 2016 | Faster RCNN Relation Networks CRNN | Classification Reasoning Text Extraction | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | FigureQA |
| Clark et al.  | 2016 | NA | Figure Extraction Table Extraction Caption Extraction | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | CS150    |
| Zaidy et al.  | 2015 | NA | Text Extraction Data Inference | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | NS      |
| Clark et al.  | 2015 | NA | Figure Extraction | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | CS150    |
| Elzer et al.  | 2007 | NA | Summarize result of research paper | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | NS      |
| Savva et al.  | 2011 | NA | Data Extraction Redesign Charts Classification (SVM) | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | NS      |
| Prasad et al. | 2007 | NA | classification (SVM) | ✔  | ✔  | ✔  | ✔  | ✔  | ✔  | CS150    |

Note: VBC: Vertical Bar Charts, HBC: Horizontal Bar Charts, LC: Line Charts, SC: Scatter Plots, PC: Pie charts, FC: Flow Charts, NS: Not Specified, NA: Not Applicable
finding the relation between the object representations is used by [29], [32].

Antol et al. [100] used the relational reasoning layer as a module in an end-to-end deep neural network and trained using traditional gradient descent optimisation methods. Inputs were a set of objects, and the relational layer learned the relationship between each pair of objects. Mathematically, the representations of relations can be expressed as in Equation 1 from [100].

\[
RN(O) = f_\phi \left( \sum_{i,j} g_\theta (o_i, o_j) \right),
\]

where \( f_\phi \) and \( g_\theta \) are implemented using multi-layer perceptrons and the parameters are the synaptic weights. \( O \) denote the list of input objects \( o_1, \ldots, o_n \). These object representations are provided by CNN as in [32], by dropping the dense layer at the end. The Relation Network then takes as input, all the pair of object representations concatenated with the question. LSTM does the question encoding. The object pair is individually processed by \( g_\theta \) to form the feature representation. Summation over all the features is processed by \( f_\phi \), to predict the output.

**D. DISCUSSIONS**

The primary focus of deep learning model is to recognise all the chart elements, both text and graphics and associate them towards extracting the data. The whole pipeline divides into object recognition, chart type identification, text identification and reasoning. We discuss the state-of-the-art in each section and the relevant works in chart images. The related recent works summarised in Table 5 to 7 show that it is a novel research area which opens the door for more improvement. This section briefly discusses the reasoning on images by the hot area in computer vision like attention mechanism and relation networks.

Before deep learning models, the features were hand-crafted. These features include frequency of edges in various regions of an image like SIFT, HOG etc.. Deep learning models automatically learn the features of the training data. The image features and the semantic vector related to the query is extracted and reasoning is done to produce any relevant chart question. Deep learning can be mixed with other approaches as well. For language encoding, bag-of-visual words or LSTM is used. CNN or any other object descriptors to extract image features and attention mechanisms for finding the relevant areas in an image and produce the relevant output. Despite the remarkable results by the deep learning models [103], it remains as a black box. Tuning the parameters of the network and choosing the features to be extracted are some of the challenges. Explainable AI (XAI) is an emerging area that includes inherently transparent algorithms into machine learning and provide more transparency to this black box.

**III. DATASETS**

Any significant advancements in the research require public datasets for training. The deep learning models require extensive datasets for training and for predicting the new unseen data. The visualisation is the only available data in the chart diagrams, and precise association of legends with axis values is very crucial in understanding this. The human brain can interpret this while the machine fails in reading this accurately. The time taken to interpret the data by a human depends on the simplicity of the chart image. We list out the available datasets comprising of chart images. These datasets have synthetic images or charts collected from the internet, and some have chart images collected from the scholarly documents. The available repositories with graph images which are in use for improving the readability of graphs, is given in Table 6.

A few datasets came up which treats this problem as a Visual Question Answering (VQA), ie, it deals with answering queries about images. FigureQA [32] is a visual reasoning corpus made of synthetic plots in five classes along with reasoning. It is made by using 15 question answer templates which are similar to CLEVR dataset [104] for visual reasoning by Stanford. This comprises of synthetic data made with Bokeh library [105]. Figure 2 gives sample figures from various datasets compiled into a single view. DVQA [46] is another synthetic dataset which includes three millions of figure and question-answer pair. This one is unique to bar charts. PlotQA [26] contains 28.9 million question-answer pairs with 224377 plots on data from real data sources. This involves line charts, bar charts, and real data scatter plots to bridge the gap between current datasets and actual data. In this they introduced a real dataset by crowdsourcing. The VQA models work poorly on such a dataset. Another dataset made with figures and charts of real world data released in 2020 is LEAFQA [95]. This includes 250,000 annotated images with two million question answer pairs.

**IV. EVALUATION METRICS**

Evaluation of the whole process is a challenging one as there is no common ground on which the different works can be compared quantitatively. This is because some work focus on localisation and extraction, while others do classification or data extraction, and a very few do reasoning on chart images. An altogether pipeline is very few, and if any, their results are on different datasets and use different models. Therefore, we present the various evaluation metrics for these different modules.

**A. CHART ELEMENTS IDENTIFICATION**

The deep learning models for detecting the visual element in the chart images use the object detection methods. The goal here is to find the object proposals by generating bounding boxes followed by inferring the objectness score of the selected bounding box. The metrics for evaluating object proposals are based on intersection over union (IOU) given
by Equation 2. IOU is based on Jaccard index between the proposal locations $B_p$ and the corresponding ground truth $B_{gt}$. An IoU of 1 implies that predicted bounding boxes and ground-truth overlap perfectly; that is, it gives the measure of how much the predicted boundary overlaps the ground truth.

$$\text{IOU} = \frac{\text{area}(B_p \cap B_{gt})}{\text{area}(B_p \cup B_{gt})}$$  \hspace{1cm} (2)

We can set a threshold for IOU to check if the detection is valid or not. The metrics make use of True positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). If IOU is greater than a threshold, it is True Positive and a False Positive, otherwise. False Negatives (FN) is when ground truth is present in the image, but not detected. True Negatives (TN) is the bounding boxes within the image which do not detect any image.

The performance is evaluated using precision and recall. The precision is the true positive value which measures the fraction of predictions that are true positives. The recall is the true positive rate, or it measures the fraction of positives that are correctly detected.

$$\text{Precision} = \frac{TP}{TP + FP}$$  \hspace{1cm} (3)

$$\text{Recall} = \frac{TP}{TP + FN}$$  \hspace{1cm} (4)

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$  \hspace{1cm} (5)

Plot the precision-recall curve, and the Average Precision (AP) is the area under this curve. Different competitors use different evaluation metrics. In PASCAL-VOC (2008) 11 point-interpolated AP is used. In COCO, Mean Average Precision (mAP) is used, which is a 101 point interpolated AP. ImageNet uses AUC (Area under the curve).

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$  \hspace{1cm} (6)

### B. CLASSIFICATION

Classification of the chart images can be evaluated by accuracy, confusion matrix, precision, recall, F1- Score, mean absolute error, mean squared error and logarithmic loss. It also gives the number of misclassifications and the diagonal gives the precise categorisation. Figure 4 provides a confusion matrix for binary classification. Based on this, we can calculate accuracy, precision, recall and F1 score as given by equation 3 to equation 6. The role of the text in the chart image can be classified into corresponding ticks, labels and title. The feature vector of text element is extracted and classified by SVM in [37]. The text role classification by [40] gives 98% to 100% F1 score.

Convolution Neural Networks (CNNs) are widely used for classification purposes. Different types of CNN are used in different works. The document figure classification using AlexNet gives an accuracy of 84% and ResNet-50 gives an accuracy of 86% in [30]. Chart images are classified by MobileNet in [39] which gives an accuracy of 99.72%. VGG16 gives an accuracy of 96.35% in [29]. Using ResNet, [37] achieves 97% accuracy.

### C. TEXT IDENTIFICATION

Text identification deals with extracting all the textual contents in a given figure. It deals with localisation and extraction...
using bounding boxes. Therefore this can be evaluated by accuracy, precision, recall and F1 score. The text identification result by OCR and by using text boxes is evaluated. The text recognition by Microsoft OCR in [30] is found to have an accuracy of 75.6% and an F1 score of 60.3%. Most of the authors use these metrics and the formula for computing. The evaluation of individual components like axis position, axis label, axis scale and data parsing gives better performance when compared to an overall accuracy of figure analysis in [25].

The performance of the axis parsing is measured by accuracy of the bounding box overlap criteria of object detection. The bounding box of prediction is correct if the intersection over union is higher than a threshold value. The bounding box of the predicted box $B_p$ and boundary box of the ground truth box $G_{th}$ obtained 95.9% accuracy for axis position and 91.6% for axis scale. Instead of taking each component level performance, when the overall figure analysis is considered, accuracy is as low as 17.3%. Word metric can be evaluated by Levenshtein distance ratio [106]. This is the ratio between extracted text and the ground truth text. In [39], the extracted text is true positive if Levenshtein distance ratio is greater than 0.80 and obtains a text accuracy of 82.4%. The accuracy of text retrieval can be manually done by collecting the number of text blocks identified among the total. The performance of OCR can be measured by using similarity functions for the textual strings. We can also validate the text localisation and extraction by accuracy of bounding box identification.

**D. REASONING**

The performance of each model on different question-answer pairs is evaluated during reasoning. The labels seen during training and testing are disjoint, and the labels in the test set are not available in dictionaries. Chart specific question-answering in [37] gives 44% accuracy. This work is looking forward to adding reasoning to incorporate out of vocabulary questions. Recent progress in attention mechanism focuses on inferring answers progressively from natural images, effective use of which can improve the existing accuracies on reasoning.

**V. APPLICATIONS AND FUTURE DIRECTIONS**

The application of deep learning methods using computer vision can solve many challenges faced by Assistive Technology. One of the issues is the graph comprehension by the visually impaired. Even though the works related to chart understanding by the visually impaired people are less in number, there are related works in chart image understanding from documents and web. Figure 5 shows related works over the years from 1995 to 2020. Figure 5 and figure 6 are generated in python-R. Lotka’s law [107] calculates the research performance, as shown in figure 6. The law gives an insight into the relationship between the number of authors and the articles. It is found that about 80% of authors published less than two articles over the given period. This shows that the number of articles published was less in the given area, but with the use of deep learning models, more research is in progress in 2020 [33], [108], [109].

The proliferation of deep learning approaches is clearing the way to access the data from vector images or recreate the data from raster images. The applications of data extraction from chart images have wide variety of applications, few of which are subgrouped as the following:

**A. IMAGE CAPTIONING**

Generating descriptions of an image is called image captioning. This is important for the automatic image indexing which can be applied in CBIR (Content Based Image Retrieval). The research papers have comparisons of results and analysis given as charts, which are embedded in pdf. The indexing of these scholarly documents for both document text and images improves searching. Al-Zaidy and Giles [41] discuss extraction of data from bar charts automatically to improve semantic labelling.

**B. DOCUMENT SUMMARIES**

Automatic image based abstracting is crucial for summarising. The prominent indices in academic writings, can make use of figures in semantic parsing and also in providing document summaries to the user. Clark and Divvala [12] identify the crucial role of tables and document figures and...
their difficulty indexing and provide a solution by decomposing the pdf documents into text, figures and captions. This extractor returns a set of figures for each document. And each figure has its page number and a bounding box for figure and its caption. Kita and Rekimoto [110] identified the increasing necessity of grouping scholarly papers and making the paper easier to read by summarising the figures. This can be further applied to enhance the information retrieval system.

C. VISUAL QUESTION ANSWERING

For a given query, learn the relevant visual and text representation to formulate the answer. VQA is at the early stage of development and reasoning is next higher level. Kafle et al. [46] understands the bar chart visualizations with the help of question answering. Answering the questions assuming a fixed vocabulary of words, for the chart images may not work all the time. CLEVR [104] dataset has reasoning questions about synthetic scenes but they perform poorly on chart datasets. FigureQA [32] is a synthetic corpus of chart images along with the question answer pair. However, it can only identify binary answers and no numerical answering is possible. These all require both visual and textual understanding. The works- [26], [34], [46] rely primarily on extracting image and question characteristics to learn how to embed their joint feature through attention mechanism [91]. Another synthetic dataset is [46] which is exclusively for bar charts. They have the metadata of all the information like position and appearances of the visual elements which can be used for ensuring “attention” to relevant region. This will not work properly with the real images. A real dataset comprising three different plot types which is able to answer simpler questions with numeric reasons is [34]. All these evidently point to the future of possibilities in extracting data from chart images.

D. ASSISTIVE TECHNOLOGY

In the last few years, the progress in Assistive Technology is supporting the Visually Impaired People. According to World health Organisation (WHO), at least 2.2 billion people have a vision impairment or blindness, of whom at least 1 billion have a vision impairment that could have been prevented or has yet to be addressed. Connier et al. in 2020 [1] discusses the various applications of smart objects for assisting the visually impaired people and they claim that though there is an increasing number of smart devices, little work is done to make use of these resources to the VIP. SeeingAI [111] is an app by Microsoft which helps them to read the bar code of products, identify currency and also describe a scene in front of him/her. The blind people depend on the screen readers, Braille and voice assistants for their education and mobility. The figures in a scene have a corresponding alt-text, for reading it loud. But they have limited chart access. The upcoming researches in extracting data will help to provide descriptions of chart images, and this can be represented as alt-text.

VI. DISCUSSIONS AND OPEN ISSUES

A chart image may highlight the comparison of results between different problems and solutions. We have reviewed the progress in graph comprehension and found that the recent trend is in deep learning approaches. The significant observations evolved from this survey are the following.

(i) The modality based procedures deal with different methods by which the visually impaired people perceive chart information. Perception by haptics alone is slow, and it depends on the feedback device used. Along with haptics feedback, the inclusion of audio is more effective. Only a limited amount of graph information can be conveyed by this.

(ii) Traditional methods include handcrafted features and most of the work solves structural property of graphs. Solving a real graph or raster images has a few challenges by this method, compared to synthetic vector chart images.

(iii) Several works are available for document images and textual extraction. However, more research is required for understanding chart images.

(iv) The charts be categorised into different types. For instance, a bar chart can be a simple bar chart, vertical, horizontal, grouped or multi-stacked. So different methods/models are required to understand each of them.

(v) An end-to-end data extraction and reasoning system is in the early stages of development.

(vi) In the case of image degradation and font variations in text, improvement can be brought in by including more invariant features.

(vii) Recent works deal with question answering the chart images, more effort is needed to incorporate open vocabulary questions in real world charts.

(viii) The performance of the models used are less when compared to the human brain, and there is wide scope for expansion.

VII. CONCLUSION

Our paper reviews the existing literature available in chart image understanding and underlying information extraction. We classify the related works based on the mode of perception of the information and as traditional and deep learning-based methods. The study focuses on extracting the chart data to aid the Visually Impaired for graph perception by reviewing the conventional methods and deep learning methods. Moreover, we enlist the various applications and challenges in reverse engineering of the chart image. In addition, we briefly outline the future research possibilities, and expect that this review will help the researchers in Assistive Technology for the Visually Impaired People, using Computer Vision.

The class agnostic classification of the visual elements and reinforcement learning for the chart data understanding are potential areas of future research. We expect more accurate studies on this in the near future. The review could start a new paradigm in enhancing the accessibility of documents and e-learning infrastructure.
REFERENCES

[1] J. Conner, H. Zhou, C. De Vaulx, J.-I. Li, H. Shi, P. Vaslin, and K. M. Hou, “Perception assistance for the visually impaired through smart objects: Concept, implementation, and experiment scenario,” IEEE Access, vol. 8, pp. 46931-46945, 2020.

[2] J. Lazar, A. Allen, J. Kleinman, and C. Malarkey, “What frustrates screen reader users on the Web: A study of 100 blind users,” Int. J. Hum. Comput. Interact., vol. 22, no. 3, pp. 247–269, May 2007.

[3] S. M. Kosslyn and S. M. Kosslyn, Graph Design for Eye Mind. New York, NY, USA: Oxford Univ. Press, 2006.

[4] K. C. Shahira and A. Lijiya, “Document image classification: Towards assisting visually impaired,” in Proc. IEEE Region 10 Conf. (TENCON), Oct. 2019, pp. 852–857.

[5] L. O’Gorman and R. Kasturi, Document Image Analysis, vol. 39, Los Alamitos, CA, USA: IEEE Computer Society Press, 1995.

[6] R. Smith, “An overview of the tesseract OCR engine,” in Proc. 9th Int. Conf. Document Anal. Recognit. (ICDAR), vol. 2, Sep. 2007, pp. 629–633.

[7] W. Huang, C. Lim Tan, and W. Kheng Leow, “Associating text and graphics for scientific chart understanding,” in Proc. 8th Int. Conf. Document Anal. Recognit. (ICDAR), 2005, pp. 580–584.

[8] K. Embrechts and M. Cliche, D. Rosenberg, D. Madeka, and C. Yee, “Scatteract: Automated restyling of scatterplots,” in Proc. 7th Int. Conf. Document Anal. Recognit., Aug, 2003, pp. 148–155.

[9] Y.-S. Chen, C.-Y. Zhou, and L.-Y. Li, “FigureSeer: Parsing result-figures in research papers,” in Proc. Eur. Conf. Comput. Vis. Amsterdam, The Netherlands: Springer, 2016, pp. 646–680.

[10] Q. Ye and D. Doermann, “Text detection and recognition in imagery: Towards assisting visually impaired,” in Proc. IEEE Region 10 Conf. (TENCON), Oct. 2019, pp. 852–857.

[11] O. Gambino, E. Minafo, R. Pirrone, and E. Ardizzone, “A tunable digital widget for improving chart images accessibility,” in Proc. 19th IEEE Int. Conf. Image Process., Sep. 2012, pp. 2865–2868.

[12] J. Lazar, A. Allen, J. Kleinman, and C. Malarkey, “What frustrates screen reader users on the Web: A study of 100 blind users,” Int. J. Hum. Comput. Interact., vol. 22, no. 3, pp. 247–269, May 2007.

[13] D. Cervone, P. Krautzberger, and V. Sorge, “Employing semantic analysis for enhanced accessibility features in MathJax,” in Proc. 13th IEEE Annu. Int. Conf. Comput. Anal. Recognit. (ICDAR), 2015, pp. 1–6.

[14] R. Girshick, “Fast R-CNN,” in Proc. 25th Int. Conf. World Wide Web, 2016, pp. 771–782.

[15] N. Methani, P. Ganguly, M. M. Khapra, and P. Kumar, “PlotQA: Reasoning over scientific plots,” in Proc. 18th ACM/IEEE Joint Conf. Digit. Libraries, May 2018, pp. 223–232.

[16] N. Methani, P. Ganguly, M. M. Khapra, and P. Kumar, “PlotQA: Reasoning over scientific plots,” 2019, arXiv:1909.00997. [Online]. Available: http://arxiv.org/abs/1909.00997

[17] S. Mirri, S. Peroni, P. Salomoni, F. Vitali, and V. Rubano, “Towards accessible graphs in HTML-based scientific articles,” in Proc. 14th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC), Jan. 2017, pp. 1067–1072.

[18] R. Smith, “An overview of the tesseract OCR engine,” in Proc. 9th Int. Conf. Document Anal. Recognit. (ICDAR), vol. 2, Sep. 2007, pp. 629–633.

[19] L. Gómez-Robledo, E. Valero, R. Huertas, M. Martínez-Domingo, and M. Stritzel, “Automatic extraction of data from computer generated charts,” in Proc. Int. Workshop Content-Based Multimedia Indexing, Jun. 2007, pp. 85–92.

[20] Y. Liu, D. Kljaban, and P. NBless, “Data extraction from charts via single deep neural network,” 2019, arXiv:1906.11906. [Online]. Available: http://arxiv.org/abs/1906.11906

[21] N. Siegel, Z. Horvitz, R. Levin, S. Divvala, and A. Farhadi, “FigureSeer: Parsing result-figures in research papers,” in Proc. Eur. Conf. Comput. Vis. Amsterdam, The Netherlands: Springer, 2016, pp. 646–680.

[22] J. Gao, Y. Zhou, and K. E. Barner, “View: Visual information extraction tool for improved charts images accessibility,” in Proc. 19th IEEE Int. Conf. Image Process., Sep. 2012, pp. 2865–2868.

[23] K. M. Hou, “Perception assistance for the visually impaired through chart image restyling,” in Proc. 24th Annu. ACM Symp. Universal Access Inf. Soc., Los Alamitos, CA, USA: IEEE Computer Society Press, 1995.

[24] L. O’Gorman and R. Kasturi, Document Image Analysis, vol. 39, Los Alamitos, CA, USA: IEEE Computer Society Press, 1995.

[25] R. Smith, “An overview of the tesseract OCR engine,” in Proc. 9th Int. Conf. Document Anal. Recognit. (ICDAR), vol. 2, Sep. 2007, pp. 629–633.

[26] N. Methani, P. Ganguly, M. M. Khapra, and P. Kumar, “PlotQA: Reasoning over scientific plots,” in Proc. IEEE Winter Conf. Comput. Vis. (WACV), Mar. 2020, pp. 1527–1536.

[27] D. H. Kim, E. Hoque, and M. Agrawala, “Answering questions about charts and generating visual explanations,” in Proc. CHI Conf. Hum. Factors Comput. Syst. Apr. 2021, pp. 1–13.

[28] J. Choi, S. Jung, D. G. Park, J. Choo, and N. Elmqvist, “Visualizing for the visually impaired,” in Proc. 24th Annu. ACM Symp. User Interface Softw. Technol., Jun. 2007, pp. 85–92.

[29] P. De, “Automatic data extraction from 2D and 3D pie chart images,” in Proc. IEEE 8th Int. Advance Comput. Conf. (IACC), Dec. 2018, pp. 20–25.

[30] A. Balaji, T. Ramanathan, and V. Sonathi, “Chart-text: A fully automated chart image descriptor,” 2018, arXiv:1812.10636. [Online]. Available: http://arxiv.org/abs/1812.10636

[31] J. G. J. Heer, and C. M. Bostock, “ReReverse-engineering visualizations: Recovering visual encodings from chart images,” in Computer Graphics Forum, vol. 36, no. 3. Hoboken, NJ, USA: Wiley, 2017, pp. 353–363.

[32] R. A. Al-Zaidy and C. L. Giles, “Automatic extraction of data from bar charts,” in Proc. 5th Int. Conf. Knowl. Capture, Oct. 2015, p. 30.

[33] M. Savva, N. Kong, A. Chhajta, L. Fei-Fei, M. Agrawala, and J. Heer, “ReVision: Automated classification, analysis and redesign of chart images,” in Proc. 24th Annu. ACM Symp. User Interface Softw. Technol., 2011, pp. 393–402.

[34] Y. Ping Zhou and C. Lim Tan, “Hough technique for bar charts detection and recognition in document images,” in Proc. Int. Conf. Image Process., vol. 2, Sep. 2000, pp. 605–608.

[35] T. Gao, J. R. Hullman, E. Adar, B. Hecht, and N. Diakopoulos, “NewsViews: An automated pipeline for creating custom geovisualizations for news,” in Proc. SIGCHI Conf. Hum. Factors Comput. Syst., 2014, pp. 3005–3014.

[36] I. Kavasidis, C. Pino, S. Palazzo, F. Rundo, D. Giordano, P. Messina, and C. Spampinato, “A saliency-based convolutional neural network for table and chart detection in digitized documents,” in Proc. Int. Conf. Image Anal. Process. Springer, 2019, pp. 292–302.

[37] K. Kafle, B. Price, B. Ju, A. Farhadi, and C. Kanam, “DVQA: Understanding data visualizations via question answering,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 5648–5656.

[38] W. Yu and S. Brewster, “Evaluation of multimodal graphs for blind people,” Universal Access Inf. Soc., vol. 2, no. 2, pp. 105–124, Jun. 2003.

[39] K. Moustakas, G. Nikolakis, K. Kostopoulos, D. Tzovaras, and M. Strintzis, “Haptic rendering of visual data for the visually impaired,” IEEE MultimediaMag., vol. 14, no. 1, pp. 62–72, Jan./Mar. 2007.
et al. [58] R. C. Gonzalez

K. Kise, “Page segmentation techniques in document analysis,” in K. Araki, T. Watanabe, and K. Minatani, “Development of tactile graph recognition,” vol. 9th Int. ACM SIGACCESS Conf. Comput. Accessibility (ASSETS), 2014, pp. 251–252.

[51] Y. Zhou and C. L. Tan, “Learning-based scientific chart recognition,” in Proc. ICDAR, vol. 1, Sep. 2015, pp. 47–50.

[52] S. Kataria, W. Brower, P. Mitra, and C. L. Giles, “Automatic extraction of figures from scholarly documents,” in Proc. ACM Symp. Document Eng., Sep. 2015, pp. 47–50.

[53] D.-J. Kim and Y.-K. Lim, “Handscope: Enabling blind people to experience 3D shapes,” in Proc. SIGCHI Conf. Hum. Factors Comput. Syst., 1998, pp. 574–581.

[54] S. Mannheim, M. Ferati, D. Huckleberry, and M. J. Palakal, “Using audemes as a learning medium for the visually impaired,” in Proc. HEALTHINF, 2009, pp. 175–180.

[55] M. Ferati, M. S. Pfaff, S. Mannheim, and D. Bolchini, “Audemes at work: Investigating features of non-speech sounds to maximize content recognition,” Int. J. Hum.-Comput. Stud., vol. 70, no. 12, pp. 936–966, Dec. 2012.

[56] D. H. Ballard, “Generalizing the Hough transform to detect arbitrary shapes,” in Readings in Computer Vision, Amsterdam, The Netherlands: Elsevier, 1987, pp. 714–725.

[57] R. C. Gonzalez et al., Digital Image Processing. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[58] J. Amara, P. Kaur, M. Owonubi, and B. Bouaziz, “Convolutional neural network based chart image classification,” in Proc. 25th Int. Conf. Elec. Comput. Graph., Vis. Comput., 2017.

[59] T. Abekawa and A. Aizawa, “SideNote: Scholarly paper browsing system based on PDF restructuring and text annotation,” in Proc. COLING 26th Int. Conf. Comput. Linguistics: Syst. Demonstrations, 2016, pp. 136–140.

[60] T. Desealets, B. Alexe, and V. Ferrari, “Localizing objects while learning their appearance,” in Proc. Eur. Conf. Comput. Vis. Marseille, France: Springer, 2010, pp. 452–466.

[61] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu, “Advanced deep-learning techniques for salient and category-specific object detection: A survey,” IEEE Signal Process. Mag., vol. 35, no. 1, pp. 84–100, Jan. 2018.

[62] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in Proc. Adv. Neural Inf. Process. Syst., 2015, pp. 91–99.

[63] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 779–788.

[64] J. Zhang, M. Huang, X. Jin, and X. Li, “A real-time Chinese traffic sign detection algorithm based on modified YOLOv2,” Algorithms, vol. 10, no. 4, pp. 127, Nov. 2017.

[65] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “SSD: Single shot multibox detector,” in Proc. Eur. Conf. Comput. Vis., Amsterdam, The Netherlands: Springer, 2016, pp. 21–37.

[66] Z. Yang and T. C. Lim, “Coordinate systems reconstruction for graph diagrams,” in Proc. 9th Int. Conf. Document Anal. Recognit. (ICDAR), vol. 1, Sep. 2007, pp. 521–525.

[67] W. Huang and C. L. Tan, “A system for understanding imaged information and its applications,” in Proc. ACM Symp. Document Eng. (DocEng), 2007, pp. 9–18.

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

[69] Y. LeCun et al., “Convolutional networks for images, speech, and time series,” in The Handbook of Brain Theory and Neural Networks, vol. 3361. Cambridge, MA, USA: MIT Press, 2002.

[70] Y. LeCun et al., “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, Jun. 2015.

[71] Y. LeCun et al., “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, Jun. 2015.
[97] A. Gupta, A. Vedaldi, and A. Zisserman, “Synthetic data for text localisation in natural images,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2315–2324.

[98] M. Sharma, S. Gupta, A. Chowdhury, and L. Vig, “ChartNet: Visual reasoning over statistical charts using MAC-networks,” in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2019, pp. 1–7.

[99] S. Elzer, E. Schwartz, S. Carberry, D. Chester, S. Demir, and P. Wu, “A Browser extension for providing visually impaired users access to the content of bar charts on the Web,” in Proc. WEBIST, vol. 2, 2007, pp. 59–66.

[100] A. Gupte, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. L. Zitnick, and D. Parikh, “VQA: Visual question answering,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 2425–2433.

[101] S. Elzer, E. Schwartz, S. Carberry, D. Chester, S. Demir, and P. Wu, “A Browser extension for providing visually impaired users access to the content of bar charts on the Web,” in Proc. WEBIST, vol. 2, 2007, pp. 59–66.

[102] A. Jabri, A. Joulin, and L. Van Der Maaten, “Revisiting visual question answering baselines,” in Proc. Eur. Conf. Comput. Vis. Amsterdam, The Netherlands: Springer, 2016, pp. 727–739.

[103] P. Angelov and A. Sperduti, “Challenges in deep learning,” in Proc. ESANN, 2016, pp. 1–8.

[104] T. Kafle and C. Kanan, “An analysis of visual question answering algorithms,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 1965–1973.

[105] S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. L. Zitnick, and D. Parikh, “VQA: Visual question answering,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 1965–1973.

[106] L. Yujian and L. Bo, “A normalized levenshtein distance metric,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 6, pp. 1091–1095, Jun. 2007.

[107] A. J. Lotka, “The frequency distribution of scientific productivity,” J. Washington Acad. Sci., vol. 16, no. 12, pp. 317–323, 1926.

[108] H. Singh and S. Shekhar, “STL-CQA: Structure-based transformers with localization and encoding for chart question answering,” in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2020, pp. 3275–3284.

[109] P. Ganguly, N. Methani, M. M. Khapra, and P. Kumar, “A systematic evaluation of object detection networks for scientific plots,” 2020, arxiv:2007.02240. [Online]. Available: http://arxiv.org/abs/2007.02240

[110] Y. Kita and J. Rekimoto, “Prediction of importance of figures in scholarly papers,” in Proc. 12th Int. Conf. Digit. Inf. Manage. (ICDIM), Sep. 2017, pp. 46–53.

[111] S. Kelley. (2018). Seeing AI: Artificial Intelligence for Blind and Visually Impaired Users. [Online]. Available: https://www.microsoft.com/en-us/ai/seeing-ai

K. C. SHAHIRA (Graduate Student Member, IEEE) received the B.Tech. degree in computer science and engineering from the NSS College of Engineering Palakkad, Calicut University, in 2009, and the Master of Engineering degree in multimedia technology from the College of Engineering Guindy, Anna University, India, in 2011. She was a part of Teaching Faculty at the National Institute of Technology Calicut for three and a half years. She is currently a Research Scholar with NIT Calicut. Her research interests include image processing, assistive technology, human–computer interaction, computer vision, and pattern recognition. She has a few conference papers to her credit.

A. LIJIYA (Member, IEEE) received the Ph.D. degree from NIT Calicut for her thesis titled “Automatic Karyotyping: Novel Approaches for Performance Enhancement.” She is currently an Assistant Professor with the Department of Computer Science and Engineering, NIT Calicut. She is also a part of Teaching Faculty at NIT Calicut for more than 13 years. She has authored seven journal articles and 12 international conference papers. She has reviewed many papers for International Journals and conferences. She had co-ordinated two Faculty Development Programmes. Her research interests include image processing, machine learning, and artificial intelligence. She was the Co-Chair of IEEE SPICES 2015 and has been a part of the Organizing Team of ISB 2011. She was the Co-Chair of the Track on Emerging Technologies in Policing, in second Police Science Congress, in 2019.