Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 14 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i14.58675

PAPER

Making STEM Subjects Graphs Accessible for Blind and Visually Impaired Students Using Document Understanding Transformer (DONUT) Model

Syed Muhammad Hassan Zaidi(✉), Abdul Hafeez Khan, Sarmad Ahmed Shaikh, Imtiaz Hussain

Sindh Madressatul Islam University, Karachi, Pakistan

m.hassan@smiu.edu.pk

ABSTRACT

Most STEM (Science, Technology, Engineering, and Mathematics) subjects rely heavily on graphs and charts, which remain largely inaccessible to students who are blind or visually impaired. While text can often be made accessible through screen readers, complex visual structures such as charts are much harder to interpret non-visually. This study presents a proof-of-concept system that applies the DONUT (Document Understanding Transformer) model to STEM charts. The model was trained and evaluated on the Benetech STEM dataset and tested on multiple images, demonstrating promising results in extracting key information such as chart type, chart ID, and x–y coordinate values. Although no user-centered trials or formal educational studies have yet been conducted, this work establishes an initial technical foundation for converting chart data into accessible formats. By enabling interpretation of chart types and data trends, the proposed system has the potential to improve accessibility in STEM education for blind and visually impaired learners, pending further validation and integration with assistive technologies.

KEYWORDS

education, HCI, AI, STEM (Science, Technology, Engineering, and Mathematics) subjects, visually impaired, graphs, DONUT (Document Understanding Transformer), transformer, deep learning

1 INTRODUCTION

Vision is a critical human sense that enables individuals to carry out everyday tasks effectively [1], [2]. In the digital age, the rapid growth of electronically generated and stored information has further emphasized the importance of visual accessibility [3]. According to the Orbis Government Association [4], approximately 43 million people are completely blind, while 295 million live with moderate to severe visual impairment. These conditions are commonly assessed using World

Zaidi, S. M. H., Khan, A. H., Shaikh, S. A., Hussain, I. (2025). Making STEM Subjects Graphs Accessible for Blind and Visually Impaired Students Using Document Understanding Transformer (DONUT) Model. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(14), pp. 4–19. https://doi.org/10.3991/ijoe.v21i14.58675

Article submitted 2025-08-20. Revision uploaded 2025-10-04. Final acceptance 2025-10-08.

Health Organization (WHO) standards, such as visual acuity tests (e.g., Snellen chart) and visual field evaluations. In the context of education, particularly in STEM (Science, Technology, Engineering, and Mathematics) disciplines, fostering accessibility supports a more inclusive and forward-looking approach to learning [5]. Although data visualizations serve as powerful tools for communicating information [6], they remain largely inaccessible to individuals who are blind or have low vision (BLV) when not adapted into non-visual modalities [7]. Furthermore, as the global population continues to expand and age, the number of people affected by vision loss due to chronic conditions, injuries, or optic nerve damage is expected to increase [8].

Individuals who are blind or visually impaired (BLV) may experience partial sight, color blindness, or complete loss of vision [9], [10]. Such conditions often limit their ability to interpret and analyze insights typically conveyed through visual data representations, including political outcomes, health statistics, weather forecasts, and financial indicators. Data visualizations—such as scatter plots, bar charts, and line graphs—are widely employed in STEM fields to present quantitative information efficiently. However, in many regions of the Global South, particularly across Africa, South Asia, and Latin America, students with BLV between the ages of 10 and 11 are frequently excluded from meaningful participation in STEM education due to the inaccessibility of instructional materials [11]. This exclusion does not reflect a lack of capability or interest on the part of these students but stems from the fact that visual elements, such as charts and graphs in STEM curricula [12], remain inaccessible without supplementary verbal explanations or descriptions provided by teachers or sighted facilitators. In contrast, sighted learners can independently engage with and interpret such materials, underscoring the inequities in access to STEM education.

Visually presented information is a critical component of STEM classroom instruction. Science subjects such as low-level microbiology curriculum courses include elements that require students to understand the core concept by interpreting data using trends and charts to address the hypothesis [13], which is very difficult for students with visual disorders. The visual impairment research study aims to accomplish visual improvement, vision substitution, or vision replacement, as first coordinated in 1981 [14], for completely and partially blind students. Research is needed to get some information about their experience report [15], which states that numerous teachers "neglected to give guidelines in classes that included students with [visual impairments]." Frequently, this was the case since educators did not adjust study hall guidelines or forgot to furnish them with open assets [16]. Subsequently, just a tiny part of the instructional materials is accessible for students with this learning distinction, except if artificial intelligence could assist with overcoming that issue.

Due to the development of semantics-oriented multimedia information retrieval techniques, medical analysis, document image understanding, and the development of intelligent systems all require image understanding. To empower BLV clients to get to perceptions by utilizing tactile modalities beyond vision, such as sound, surface, or text [17], [18], [19]. Mobile handheld devices are widely accessible and adaptable to assistive technologies. Numerous assistive innovations have been created, such as Artificial Intelligence of Things (AIoT), which is one of the significant patterns for the future, as many brilliant gadgets are interconnected and profoundly applied in human society [20], [21].

The primary obstacle facing an image understanding system is the absence of a comprehensive, general, and adaptable understanding of the real world. Machine vision frameworks develop inner models of the handled scene and check and update them. Picture understanding is accomplished by assuming the interior model matches reality [22]. Numerous scientists have examined alternative methods of showing graphical data for individuals with visual impairments. They are severely constrained because they focus on rendering graphical elements in an alternative medium. Consequently, the significant objective is to foster a framework that construes the planned message in a chart and gives an underlying outline that incorporates the message alongside eminent elements of the diagram [22]. In scholarly papers, outlines can be utilized as strong synopsis devices [23], which permit scientists to explore rapidly through results and understand them. Graphs ordinarily supplement the realities depicted in the fundamental text of a record [24].

In this research work, the model is train and test on Benetech stem subject chart images data using DONUT (Document Understanding Transformer) high-level model intended for grasping archives by joining two significant parts: a picture Transformer Encoder and a Text Transformer Decoder [25] to analyze and interpret Chart ID, Chart Types, and Data Trends to store them. The other sections are a literature review followed by our contribution in the methodology section, along with data size and description, with parameter tuning and configuration through experiments of interpretation and extraction from STEM graph images.

2 LITERATURE REVIEW

The power of decision-making towards good quality health, financial management, and other everyday activities [26], as well as education [27] and employment opportunities [28], can be affected due to the inability to access information. Countless graphs encoded in bitmap pictures with information locked away for everyone to access BLV individuals more than sighted users [29]. Vision hindrance is a typical handicap with various degrees of seriousness. Hwang, Jumi, et al. [30] investigated new mechanical open doors involving licenses in assistive innovation for blind and outwardly disabled individuals. Computer-based intelligence gadgets can assist outwardly impeded individuals with understanding the text, walking in the city, and even making art. Most researchers in assisted reading concentrate on producing sounds that visually impaired individuals can understand.

Deep learning is broadly utilized in clinical image [31] examinations since it can learn deep associations between data and information. Assistive technologies have provided visual alternatives for BLV individuals through products, gadgets, software, or frameworks [32–33]. Such degrees of progress are being utilized by experts to continually redesign people's very own fulfillment, particularly for those with disabilities or chronic medical conditions [34]. Gadgets, present services, or projects known as assistive technology (AT) are utilized to improve the functional abilities of people with disabilities [35]. The level of assistive innovation research studies integrates hearing block, visual inadequacy, and mental shortcomings, among others [36–37]. Perception is an integral asset that assists us with understanding complex data by changing it into visual structures [38]. Nevertheless, it has fundamentally revolved around the visual viewpoint, which can be risky for apparently obstructed individuals. One of the promising technologies is deep learning, which enhances and improves the lives of BLV people and addresses the issue [39] with a voice-based development to address and depict pictures embedded in printed texts.

Ganesan, Jothi, et al. [39] identify image captioning, hidden features, and converting written text to speech by using convolutional neural network (CNN) and long short-term memory (LSTM). Tasnim et al. [40] assisted people with visual impairments and presented an automated method for using a convolutional neural network to identify Bangladeshi banknotes. The examination demonstrated effectiveness, as shown by the framework was 92% exact in determining the notes and could give composed and sound results. Mukhiddinov et al. [41] designed an innovative glass framework for visually impaired individuals with visual weakness using computer vision and deep learning algorithms. This proposed approach combines four unquestionable modules: improvement of low-light pictures, location of items, sound feedback, and making of material illustrations. Mishra et al. [42] made ChartVi, an automated framework for summing up outlines that produces a summary from various graph pictures, including lines, pie, bars, etc. In this technique, the CNN-VGG16 network model is utilized to distinguish the classifications of outline pictures. From that point forward, highlight extraction strategies are utilized to naturally isolate text-based data from graphical data.

Outwardly impeded individuals commonly utilize screen readers to peruse the web. Screen-reader gadgets, for example, JAWS [43] and NVDA [44], support individuals with visual hindrances in exploring site pages and reading texts through keyboard shortcuts. As a result of the varieties in the format of the diagrams, the performance of these frameworks can be improved with human direction, Jung, Daekyoung, et al. [45] as the ChartSense tool demonstrates. Software packages exist that generate data from specific diagrams [46, 47]. Méndez et al. [48]. Design an application named iVoLVER that separates information from images in existing raster graphs and produces a more appealing and user-friendly representation. Saava et al. developed a framework called ReVision to identify diagrams into five classifications and concentrate information from bar graphs and pie outlines [49]. A standard method for making a chart diagram open is to describe it with an alternative text (alt text). A well-known method has been experimentally tested with BLV clients for online pictures [50].

The HighCharts [51] visualization library, for instance, in 2017, began making an open outline course that gives a generous semantic level of depiction, with the strategy moved by MathJax [52]. Visa Chart Components gives users a design-free portrayal plan that permits users to access unrefined data [53]. A model system was presented by Murillo-Spirits and Miesenberger [54] in which the client could get some information about the data's meaning, cutoff points, the range of data. Sharif et al. [55, 20] have utilized a comparable methodology to construct VoxLens. This JavaScript module can respond to inquiries with predefined words like "most extreme," "least," "middle," and "mode" and produce an enlightening summary of a diagram. and. afterward, it permits clients to collaborate with the representation by utilizing voice-initiated commands. Kim et al. [56] examined how BLV individuals might interact with a chart QA system, listed several factors influencing their querying behavior, and confirmed that BLV individuals were sufficiently motivated to utilize the system.

Chundury et al. [57] endeavored to comprehend how people with visual deficiency see environmental elements utilizing non-visual faculties, and given their findings, they illustrated a few design considerations for open representations. Murphy et al. [58] ran an extensive review including 30 visually impaired and some partially sighted system users to figure out the everyday difficulties they face while utilizing the web. The study results revealed that a shortage of alt texts on pictures and inadequately named page things make it difficult and make it attempt to shape a site page format. As indicated by an interview study [11], individuals with visual impairment needed to utilize perceptions and pay attention to listening to texts to draw a picture in their heads. From the examination of interview information, the researchers suggested that the alt text ought to incorporate chart type, axis labels, and information patterns, and users should have access to data of interest.

The iGRAPH-Light framework [59] makes diagrams available to blind clients by producing captions and supporting routes through the graph using the keyboard. They tracked down that the metadata, for instance, frame titles, center point names, and chart types, are critical information to visually disabled readers. Charts that are easy to read for screen readers, including the type of chart, trends in data, and statistical standard deviation, are created using EvoGraph [60–64], a jQuery plugin. Alam MZ et al. [65] created a SeeChart. This browser extension program naturally dismantles SVG representations from a site page, making them open without depending on designers to compose code. Additionally, unlike the above bodywork, SeeChart empowers intelligent choice and filtering using the keyboard. Lundgard and Satyanarayan [66] emphasize the use of natural language generation (NLG) for visual descriptions through a four-level semantic model. Beyond auditory-only approaches, multimodal solutions have also been explored.

3 METHODOLOGY

Access to information is a human right; due to differences or disabilities, no person should encounter barriers to employment, education, or literacy. This is an objective of a nonprofit organization named Benetech. In this research work, the selected dataset is taken from Benetech, having about 65,000 labeled scientific figures of line graphs, dot plots, scatter plots, and bar graphs, both vertical and horizontal, extracted from STEM textbooks across 12 disciplines (e.g., biology, physics, and economics). Various professional production sources extract several thousand figures, most synthetic.

Fig. 1. STEM graphs dataset (line, bar, and scatter)

Figure 1 shows different types of graphs taken from textbooks of STEM subjects. Due to the complex structure of chart images, current OCR-based approaches have shown promising performance, but they suffer from 1) high computational costs for using OCR; 2) inflexibility of OCR models on languages or types of documents; 3) OCR error propagation to the subsequent process. To address these issues, Kim, Geewook, et al [67] introduce a novel OCR-free VDU model named DONUT, which stands for Document Understanding Transformer. As the first step in OCR-free VDU research, authors propose a simple architecture (i.e., Transformer) with a pre-training objective (i.e., cross-entropy loss). In this study, the selected model is DONUT, with different parameters and configurations as presented in Figure 2.

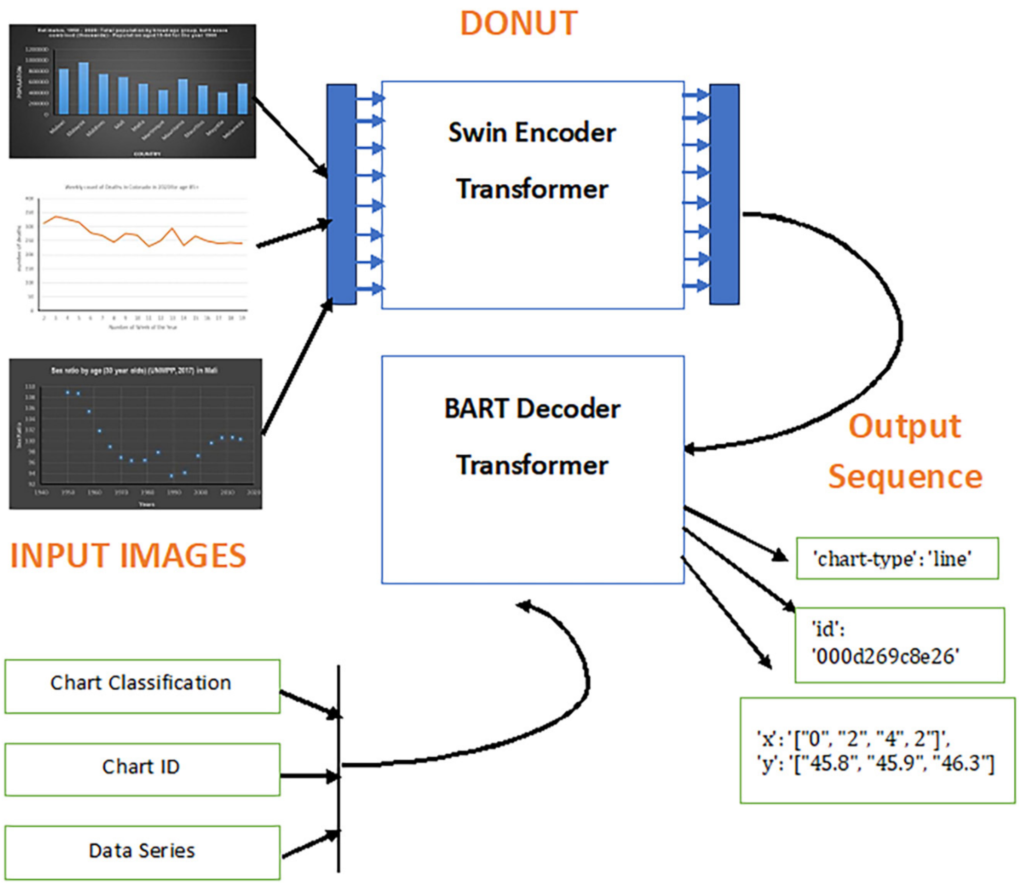

Fig. 2. Proposed methodology of DONUT model

In the proposed model, chart images are taken as an input, which passes to the encoder transformer; here, we use a vision encoder named Swin Transformer [68]; the encoder first encodes the chart images into a tensor of embeddings with the following parameters (of shape batch_size, seq_len, hidden_size) and the output of the final Swin Transformer block is fed into the decoder. We use BART [69], [70] (Bidirectional Encoder Representations from Transformers) decoder architecture that autoregressively generates data points, image ID, and a class of chart images. Finally, the output sequence produces results. There are several parameters used in the DONUT model, which are stated as:

3.1 Configuration

The DONUT model has various configuration (CFG) parameters that can be set during training. Table 1 parameters can be adjusted to customize the model's behavior and optimize its performance for a specific use case. The model was trained using the Adam optimizer with a learning rate of 3e-5, categorical cross-entropy loss, batch size of 2, and 2 epochs. Warmup steps (300), cosine learning rate scheduling, and early stopping were applied. Mixed precision (FP16) was used for computational efficiency. Training was conducted on an NVIDIA GPU, with random seed 42 for reproducibility.

Table 1. Configuration parameters values used in DONUT

Parameter	Value	Description/Unit
Batch Size	2	Number of samples per training batch
Learning Rate	3e-5	Step size for model weight updates (unitless)
Warmup Steps	300	Steps used to increase the learning rate gradually
Number of Epochs	2	Full passes through the training dataset
Seed	42	Random seed for reproducibility
Precision	16	Mixed-precision training (FP16 used)
Check Validation Every N Epochs	1	How frequently validation is run (in epochs)
Log Steps	200	Frequency of logging training progress (in steps)
Validation Check Interval	1.0	Frequency of validation checks (per epoch)
Accelerator	GPU	Hardware used for training
LR Scheduler Type	Cosine	Type of learning rate decay
Optimizer	Adam (3e-5)	Optimization algorithm and its learning rate
Early Stopping	True	Stop training early if performance stops improving
Dataset	Benetech (Charts)	Source of images for training/testing
Preprocessing	Resize 224 × 224, normalize	Standard input preparation steps
Augmentation	None	
Evaluation Outputs	Coordinates + descriptions	Metrics and outputs reported

3.2 Tokenization

According to the dataset, we have defined the special tokens, tokens such as "PROMPT_TOKEN," "X_START," "X_END," "Y_START," and "Y_END." These tokens are used to mark specific positions or elements; then we define the chart type tokens like "<line> <vertical_bar> <horizontal_bar> <dot> <scatter>" would be tokenized into ["<line>," "<vertical_bar>," "<horizontal_bar>," "<dot>," "<scatter>"]. After doing that, we combined the special and chart-type tokens.

3.3 Data generator

During training sessions, data generators produced data on demand using its class, mainly where the entire dataset cannot fit into memory. The training process is considered completed when all batches are executed one by one generated by the generator. In the current scenario, the data generator takes a JSON file as input, along with all variables of the corresponding image, and appends it into a JSON file to make a dataset more efficient to handle.

3.4 Transformers

Vision Encoder Decoder: Configuration is created using the "VisionEncoderDecoderConfig" class from the Hugging Face Transformers library. It loads a pre-trained configuration for a model called "naver-clova-ix/Donut-base" (which appears to be related to the "DONUT" model for vision tasks). Then, specific configurations for the encoder and decoder parts of the model are modified to match the settings specified in the CFG class. By taking the benefits of the load pre-trained configuration for the vision encoder-decoder model named "naver-clova-ix/Donut-base" from the Hugging Face model, from the configuration. The image encoder size function updates the image size for the encoder that we define in the CFG, and then we define the maximum decode length.

3.5 Extracted and generated

Extracted and generated examples and a separate validation set with only extracted examples for evaluation. The dataset contains information about ground truth texts, image paths, and chart types for training and evaluation tasks. The data is being processed for a DONUT-based vision encoder-decoder model, as discussed above.

- Count Chart Types in Extracted Data Sources.
- Perform Stratified K-fold Split.
- Print chart type counts in each fold.

The training and validation prepare the necessary data for training the DONUT-based vision encoder-decoder model; additionally, it prints some information about the data batch to understand its structure.

4 EXPERIMENT AND RESULT ANALYSIS

The training process involves a dataset of charts and graphs in the STEM fields, which are divided into training, validation, and testing splits of 70%, 15%, and 15%, respectively, stratified across chart types to ensure balanced representation. Preprocessing steps included resizing each image to $224 \times 224 \times 3$, normalization, and dataset cleaning. The research implements a DONUT pre-trained model image transformer encoder (SWIN) and an autoregressive text transformer decoder (BART). Vision Encoder Decoder loads a pre-trained configuration for a model called "Naver-Clova-IX/Donut-Base," which appears to be related to the "DONUT" model for vision tasks, and Donut preprocesses the ground truth text and image output Conversion. The output token sequence is converted to a desired structured format. We adopted a JSON format due to its high representation capacity. We add two unique tokens, [START $*$] and [END $*$], which $*$ indicate each field to extract. If the output token sequence is wrongly structured, we treat the field as lost. For example, if only [START name] exists but no [END name], we assume the model fails to extract the "name" field.

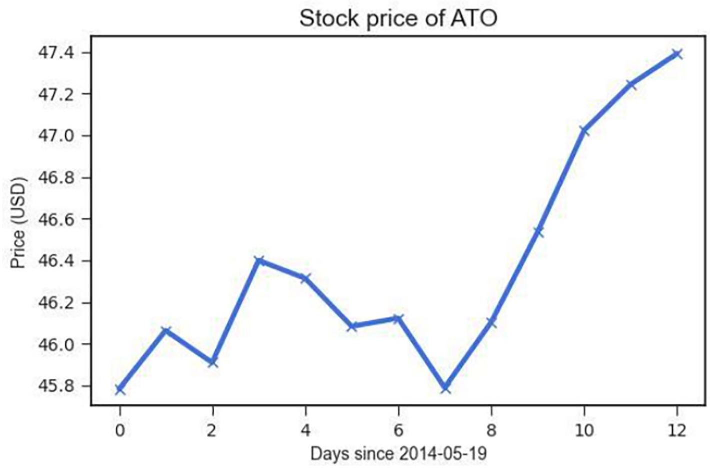

Fig. 3. Line graph

The output shown in Figure 3 demonstrates how the DONUT model successfully extracts essential information from a graph image. Specifically, the model identifies the type of chart (e.g., line chart), the unique chart ID, and the x- and y-axis values. These values represent the actual data points plotted on the graph. For example, in the result shown:

- 'chart-type': 'line' indicates that the chart is a line graph.
- 'id': '000d269c8e26' is a unique identifier assigned to that chart image.
- 'x': ["0", "2", "4", "6", "8", "10", "12"] are the x-axis values (e.g., time or categories).
- 'y': ["45.8", "45.9", "46.3", "46.1", "46.1", "47.0", "47.4"] are the y-axis values (e.g., temperature, score, etc.).

These outputs are formatted in JSON, a widely used structure in computing for organizing data. Here, it helps us show that the model correctly "reads" the graphs such as a human would. This interpretation allows visually impaired users to receive the same information in alternative formats (e.g., text or audio).

Table 2. Test images result generated by DONUT model

Charts Id	Data Trends	Chart Type
000b92c3b098_x	0; 6; 12; 18; 24	Line
007a18eb4e09_x	0.0; 0.4; 0.8; 1.2; 1.6; 2.0; 2.4	Line
00dcf883a459_x	Group 1; Group 2	Vertical Bar
000b92c3b098_y	–0.0; –1.4; –1.4; –2.1; –3.3	Line
007a18eb4e09_y	0.0132; 0.0132; 0.0132; 0.0132; 0.0132; 0.0132; 0.0132	Line
00dcf883a459_y	3.7; 8.4	Vertical Bar

Table 2 presents initial results on selected test cases, demonstrating the feasibility of the proposed approach. For example, when provided with a line chart image, the model correctly identified the chart type as a line graph (ID: 007a18eb4e09) and extracted corresponding data trends along both the x- and y-axes. In the table, the first three rows represent x-axis values from the test image, while the last three rows correspond to y-axis values. The extracted numerical values are reported with four decimal precisions, reflecting the model's ability to capture fine-grained coordinate information. This demonstrates how technical outputs can be mapped into meaningful natural language descriptions suitable for screen readers or audio-based delivery.

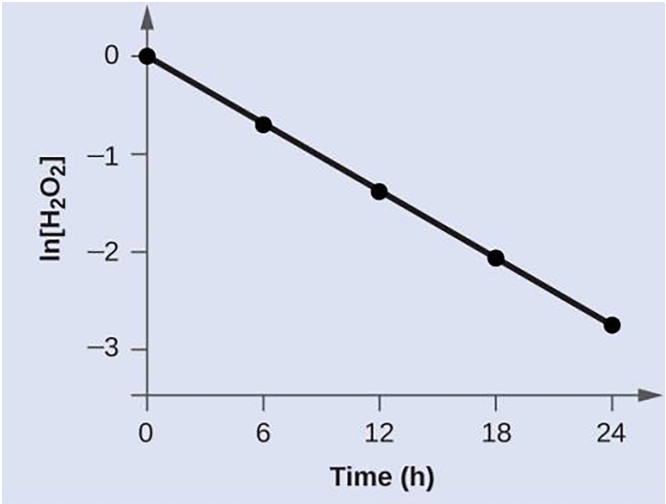

Fig. 4. Test case (ID: 000b92c3b098)

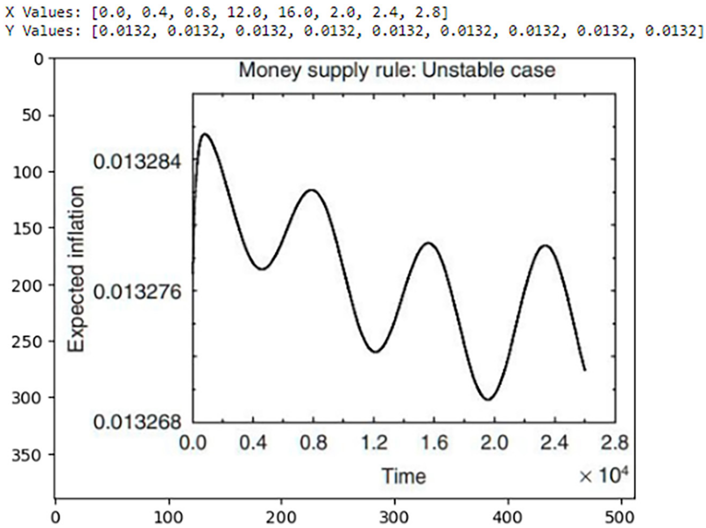

Fig. 5. Test case (ID: 007a18eb4e09)

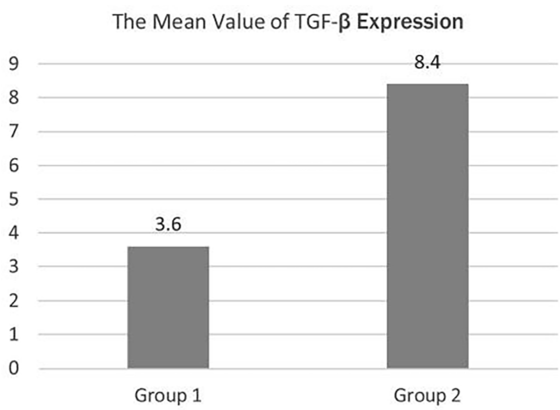

Fig. 6. Test case vertical (00dcf883a459)

Figures 4, 5, and 6 present test case examples are used to validate the performance, evaluate the DONUT model's ability to extract structured data (such as x and y values), and verify the output accuracy, as seen in Table 2. Each figure corresponds to a different chart type extracted from the dataset: Figures 4 and 5 depict line charts with varying data scales and point distributions. In contrast, Figure 6 presents a vertical bar chart. The system occasionally misclassified overlapping chart elements, struggled with low-resolution images, and produced inaccurate extractions when axis labels were partially obscured. These failure cases emphasize the importance of integrating robust OCR for embedded text and refining preprocessing pipelines. A more comprehensive error analysis with quantitative metrics will be conducted in future work.

5 CONCLUSION

This study interprets chart ID, chart type, and data trends to provide alternative text (alt text) by proposing a novel assistive technology system for understanding charts and graphs in STEM education. This information could be helpful for people with vision or cognitive impairments; however, further studies involving actual users are needed to evaluate the effectiveness of the education approach. Our proposed DONUT model extracts core information such as graph type, graph ID, and values from x and y coordinates to analyze line, bar (vertical and horizontal), and point graphs commonly found in STEM textbooks. The proposed technology has potential for future use in school settings, subject to further validation through user studies.

6 FUTURE WORK

The primary objective of this study is to validate the DONUT model's capability to identify chart types and accurately extract data trends from a controlled dataset, serving as a proof-of-concept. While configuration parameters and evaluation design were intentionally lightweight, future work will conduct systematic hyperparameter tuning, quantitative benchmarking, and user-centered studies to strengthen rigor and assess practical accessibility impact. Further this work will be extended by incorporating a larger and more diverse collection of real-world charts to better support BLV users in interpreting graphical data dynamically. The performance of the proposed model will also be evaluated against emerging multimodal large language models such as GPT-4V, LLaVA, and Claude 3 Opus, as well as existing chart understanding frameworks such as ChartVi and iVoLVER, to assess comparative effectiveness in chart interpretation, computational efficiency, and accessibility. Additionally, the structured JSON output will be enhanced through integration with natural language generation techniques to provide accessible visual descriptions for BLV users.

7 REFERENCES

[1] P. Argüeso, "Human ocular mucins: The endowed guardians of sight," *Advanced Drug Delivery Reviews*, vol. 180, p. 114074, 2022. https://doi.org/10.1016/j.addr.2021.114074

[2] P. Boyce, "Light, lighting and human health," *Lighting Research & Technology*, vol. 54, no. 2, pp. 101–144, 2021. https://doi.org/10.1177/14771535211010267

[3] G. Anton, "Graph attention networks for link prediction in semantic word grouping," Uppsala University, Disciplinary Domain of Science and Technology, UPTEC STS, ISSN 1650-8319, Report No. 23001, Uppsala, Sweden, 2023.

[4] Orbis International, "Avoidable blindness," [Online]. Available: https://www.orbis.org/en/avoidable-blindness

[5] F. K. Astuti, E. Ellianawati, M. Masturi, W. Wiyanto, and W. Sumarni, "Central Java teachers' perspective on science, technology, engineering and mathematics (STEM) learning," *Journal of Innovative Science Education*, vol. 12, no. 1, pp. 74–81, 2023. https://doi.org/10.15294/jise.v12i1.53846

[6] Z. Shaikh *et al.*, "Blockchain hyperledger with non-linear machine learning: A novel and secure educational accreditation registration and distributed ledger preservation architecture," *Applied Sciences*, vol. 12, no. 5, p. 2534, 2022. https://doi.org/10.3390/app12052534

[7] K. Marriott *et al.*, "Inclusive data visualization for people with disabilities," *Interactions*, vol. 28, no. 3, pp. 47–51, 2021. https://doi.org/10.1145/3457875

[8] S. Zhu, K. Ota, and M. Dong, "Energy-efficient artificial intelligence of things with intelligent edge," *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7525–7532, 2022. https://doi.org/10.1109/JIOT.2022.3143722

[9] A. Budrionis, D. Plikynas, P. Daniušis, and A. Indrulionis, "Smartphone-based computer vision travelling aids for blind and visually impaired individuals: A systematic review," *Assistive Technology*, vol. 34, no. 2, pp. 178–194, 2020. https://doi.org/10.1080/10400435.2020.1743381

[10] B. Kuriakose, R. Shrestha, and F. E. Sandnes, "Tools and technologies for blind and visually impaired navigation support: A review," *IETE Technical Review*, vol. 39, no. 1, pp. 3–18, 2020. https://doi.org/10.1080/02564602.2020.1819893

[11] G. Alexiou, "How AI is being used to help blind students 'visualize' graphs and charts," Forbes, 2022. https://www.forbes.com/sites/gusalexiou/2022/07/27/how-ai-is-being-used-to-help-blind-students-visualize-graphs-and-charts/?sh=457af8f02c6d

[12] C. Jung, S. Mehta, A. Kulkarni, Y. Zhao, and Y.-S. Kim, "Communicating visualizations without visuals: Investigation of visualization alternative text for people with visual impairments," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 1, pp. 1095–1105, 2022. https://doi.org/10.1109/TVCG.2021.3114846

[13] J. L. Joyner and S. T. Parks, "Scaffolding STEM literacy assignments to build greater competence in microbiology courses," *Journal of Microbiology & Biology Education*, vol. 24, no. 1, 2023. https://doi.org/10.1128/jmbe.00218-22

[14] R. Welsh, "Foundations of orientation and mobility: Technical report," American Printing House for the Blind, Louisville, KY, USA, 1981.

[15] B. W. Stone, D. Kay, and A. Reynolds, "Teaching visually impaired college students in introductory statistics," *Journal of Statistics Education*, vol. 27, no. 3, pp. 225–237, 2019. https://doi.org/10.1080/10691898.2019.1677199

[16] B. Whitburn, " 'A really good teaching strategy': Secondary students with vision impairment voice their experiences of inclusive teacher pedagogy," *British Journal of Visual Impairment*, vol. 32, no. 2, pp. 148–156, 2014. https://doi.org/10.1177/0264619614523279

[17] Y. Yang, K. Marriott, M. Butler, C. Goncu, and L. Holloway, "Tactile presentation of network data: Text, matrix or diagram?" in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020. https://doi.org/10.1145/3313831.3376367

[18] A. Sharif, O. H. Wang, A. T. Muongchan, K. Reinecke, and J. O. Wobbrock, "VoxLens: Making online data visualizations accessible with an interactive JavaScript Plugin," in *CHI Conference on Human Factors in Computing Systems*, 2022. https://doi.org/10.1145/3491102.3517431

[19] C. Engel, E. F. Müller, and G. Weber, "SVGPlott," in *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, 2019. https://doi.org/10.1145/3316782.3316793

[20] S. Bi *et al.*, "A survey on artificial intelligence aided internet-of-things technologies in emerging smart libraries," *Sensors*, vol. 22, no. 8, p. 2991, 2022. https://doi.org/10.3390/s22082991

[21] B. H. Lee and Y. J. Lee, "Evaluation of medication use and pharmacy services for visually impaired persons: Perspectives from both visually impaired and community pharmacists," *Disability and Health Journal*, vol. 12, no. 1, pp. 79–86, 2019. https://doi.org/10.1016/j.dhjo.2018.07.012

[22] C. Clark and S. Divvala, "PDFFigures 2.0: Mining figures from research papers," in *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries*, 2016, pp. 143–152. https://doi.org/10.1145/2910896.2910904

[23] K. Swathi, B. Vamsi, and N. T. Rao, "A deep learning-based object detection system for blind people," in *Smart Technologies in Data Science and Communication*, in Lecture Notes in Networks and Systems, 2021, pp. 223–231. https://doi.org/10.1007/978-981-16-1773-7_18

[24] L. D. Lopez *et al.*, "A framework for biomedical figure segmentation towards image-based document retrieval," *BMC Systems Biology*, vol. 7, no. Suppl 4, p. S8, 2013. https://doi.org/10.1186/1752-0509-7-S4-S8

[25] B. Davis, B. Morse, B. Price, C. Tensmeyer, C. Wigington, and V. Morariu, "End-to-end document recognition and understanding with dessurt," in *Computer Vision – ECCV 2022 Workshops*, 2023, pp. 280–296. https://doi.org/10.1007/978-3-031-25069-9_19

[26] National Federation of The Blind, "Blindness Statistics | National Federation o the Blind," Nfb.org, 2019. https://nfb.org/resources/blindness-statistics

[27] W. A. Erickson, S. VanLooy, S. von Schrader, and S. M. Bruyère, "Disability, income, and rural poverty," in *Disability and Vocational Rehabilitation in Rural Settings*, 2017, pp. 17–41. https://doi.org/10.1007/978-3-319-64786-9_2

[28] A. F. Siu *et al.*, "COVID-19 highlights the issues facing blind and visually impaired people in accessing data on the web," in *Proceedings of the 18th International Web for All Conference*, 2021. https://doi.org/10.1145/3430263.3452432

[29] J. Choi, S. Jung, D. G. Park, J. Choo, and N. Elmqvist, "Visualizing for the non-visual: Enabling the visually impaired to use visualization," *Computer Graphics Forum*, vol. 38, no. 3, pp. 249–260, 2019. https://doi.org/10.1111/cgf.13686

[30] J. Hwang, K. H. Kim, J. G. Hwang, S. Jun, J. Yu, and C. Lee, "Technological opportunity analysis: Assistive technology for blind and visually impaired people," *Sustainability*, vol. 12, no. 20, p. 8689, 2020. https://doi.org/10.3390/su12208689

[31] J. Wang, S. Wang, and Y. Zhang, "Artificial intelligence for visually impaired," *Displays*, vol. 77, p. 102391, 2023. https://doi.org/10.1016/j.displa.2023.102391

[32] A. Shelton and T. Ogunfunmi, "Developing a deep learning-enabled guide for the visually impaired," in *2020 IEEE Global Humanitarian Technology Conference (GHTC)*, 2020. https://doi.org/10.1109/GHTC46280.2020.9342873

[33] J. S. Kallimani, K. G. Srinivasa, and R. B. Eswara, "Extraction and interpretation of charts in technical documents," in *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2013. https://doi.org/10.1109/ICACCI.2013.6637202

[34] A. K. Triantafyllidis and A. Tsanas, "Applications of machine learning in real-life digital health interventions: Review of the literature," *Journal of Medical Internet Research*, vol. 21, no. 4, p. e12286, 2019. https://doi.org/10.2196/12286

[35] K. Manjari, M. Verma, and G. Singal, "A survey on assistive technology for visually impaired," *Internet of Things*, vol. 11, p. 100188, 2020. https://doi.org/10.1016/j.iot.2020.100188

[36] C. Park, C. C. Took, and J.-K. Seong, "Machine learning in biomedical engineering," *Biomedical Engineering Letters*, vol. 8, no. 1, pp. 1–3, 2018. https://doi.org/10.1007/s13534-018-0058-3

[37] B. K. Swenor, P. Y. Ramulu, J. R. Willis, D. Friedman, and F. R. Lin, "The prevalence of concurrent hearing and vision impairment in the United States," *JAMA Internal Medicine*, vol. 173, no. 4, p. 312, 2013. https://doi.org/10.1001/jamainternmed.2013.1880

[38] S. C. Daggubati and J. Sreevalsan-Nair, "ACCirO: A system for analyzing and digitizing images of charts with circular objects," in *International Conference on Computational Science*, 2022, pp. 605–612. https://doi.org/10.1007/978-3-031-08757-8_50

[39] J. Ganesan, A. T. Azar, S. Alsenan, N. A. Kamal, B. Qureshi, and A. E. Hassanien, "Deep Learning Reader for Visually Impaired," *Electronics*, vol. 11, no. 20, p. 3335, 2022. https://doi.org/10.3390/electronics11203335

[40] R. Tasnim, S. T. Pritha, A. Das, and A. Dey, "Bangladeshi Banknote recognition in real-time using convolutional neural network for visually impaired people," in *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, 2021. https://doi.org/10.1109/ICREST51555.2021.9331182

[41] M. Mukhiddinov and J. Cho, "Smart glass system using deep learning for the blind and visually impaired," *Electronics*, vol. 10, no. 22, p. 2756, 2021. https://doi.org/10.3390/electronics10222756

[42] P. Mishra, S. Kumar, M. K. Chaube, and U. Shrawankar, "ChartVi: Charts summarizer for visually impaired," *Journal of Computer Languages*, vol. 69, p. 101107, 2022. https://doi.org/10.1016/j.cola.2022.101107

[43] R. S. Bankar and S. R. Lihitkar, "JAWS (Job Access With Speech)," in *Advances in Educational Technologies and Instructional Design*, S. Dhamdhere and F. Andres, Eds., IGI Global Scientific Publishing, 2022, pp. 19–40. https://doi.org/10.4018/978-1-7998-4736-6.ch002

[44] https://www.nvaccess.org/

[45] D. Jung *et al.*, "ChartSense," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017. https://doi.org/10.1145/3025453.3025957

[46] ROHATGI A. WebPlotDigitizer. https://apps.automeris.io/wpd4/, 201 . 3, 9.

[47] B. Tummers, "Datathief III," DataThief, 2006. http://datathief.org/

[48] G. G. Méndez, M. A. Nacenta, and S. Vandenheste, "iVoLVER," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016. https://doi.org/10.1145/2858036.2858435

[49] M. Savva, N. Kong, A. Chhajta, L. Fei-Fei, M. Agrawala, and J. Heer, "ReVision," in *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, 2011. https://doi.org/10.1145/2047196.2047247

[50] A. Stangl, N. Verma, K. R. Fleischmann, M. R. Morris, and D. Gurari, "Going beyond one-size-fits-all image descriptions to satisfy the information wants of people who are blind or have low vision," in *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, 2021. https://doi.org/10.1145/3441852.3471233

[51] S. C. S. Joyner, A. Riegelhuth, K. Garrity, Y.-S. Kim, and N. W. Kim, "Visualization accessibility in the wild: Challenges faced by visualization designers," in *CHI Conference on Human Factors in Computing Systems*, 2022. https://doi.org/10.1145/3491102.3517630

[52] D. Cervone, P. Krautzberger, and V. Sorge, "Employing semantic analysis for enhanced accessibility features in MathJax," in *2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, 2016. https://doi.org/10.1109/CCNC.2016.7444948

[53] Visa C art (2022). Visa Chart Components. https://github.com/visa/visa-chart-components

[54] T. Murillo-Morales and K. Miesenberger, "AUDiaL: A natural language interface to make statistical charts accessible to blind persons," in *Computers Helping People with Special Needs*, 2020, pp. 373–384. https://doi.org/10.1007/978-3-030-58796-3_44

[55] A. Sharif, A. M. Zhang, A. Shih, J. O. Wobbrock, and K. Reinecke, "Understanding and improving information extraction from online geospatial data visualizations for screen-reader users," in *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, 2022. https://doi.org/10.1145/3517428.3550363

[56] J. Kim, A. Srinivasan, N. W. Kim, and Y.-S. Kim, "Exploring chart question answering for blind and low vision users," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023. https://doi.org/10.1145/3544548.3581532

[57] P. Chundury, B. Patnaik, Y. Reyazuddin, C. Tang, J. Lazar, and N. Elmqvist, "Towards understanding sensory substitution for accessible visualization: An interview study," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 1, pp. 1084–1094, 2022. https://doi.org/10.1109/TVCG.2021.3114829

[58] E. Murphy, R. Kuber, G. McAllister, P. Strain, and W. Yu, "An empirical investigation into the difficulties experienced by visually impaired Internet users," *Universal Access in the Information Society*, vol. 7, nos. 1–2, pp. 79–91, 2007. https://doi.org/10.1007/s10209-007-0098-4

[59] O. G. Ahmed, "Smart building systems: A confluence of architecture and technology," *KHWARIZMIA*, vol. 2025, pp. 11–22, 2025. https://doi.org/10.70470/KHWARIZMIA/2025/002

[60] S. A. Inam *et al.*, "A novel deep learning approach for investigating liquid fuel injection in combustion system," *Discover Artificial Intelligence*, vol. 5, no. 1, p. 32, 2025. https://doi.org/10.1007/s44163-025-00248-2

[61] A. Korneev, Y. Niu, and A. Ibrahim, "Electric vehicles in the 21st century: Historical evolution, environmental impact, and safety challenges for sustainable mobility," *KHWARIZMIA*, vol. 2024, pp. 32–38, 2024. https://doi.org/10.70470/KHWARIZMIA/2024/005

[62] S. A. Inam *et al.*, "PR-FCNN: A data-driven hybrid approach for predicting PM2.5 concentration," *Discover Artificial Intelligence*, vol. 4, no. 1, p. 75, 2024. https://doi.org/10.1007/s44163-024-00184-7

[63] M. Ur Rahim, M. Hussain, S. A. Inam, and H. Hashim, "Ignition behavior of supercritical liquid fuel in combustion system," *Journal of Mechanics of Continua and Mathematical Sciences*, vol. 16, no. 8, 2021. https://doi.org/10.26782/jmcms.2021.08.00003

[64] S. A. Inam, D. Iqbal, H. Hashim, and M. A. Khuhro, "An empirical approach towards detection of tuberculosis using deep convolutional neural network," *International Journal of Data Mining, Modelling and Management*, vol. 16, no. 1, pp. 101–112, 2024. https://doi.org/10.1504/IJDMMM.2024.136232

[65] I. I. Al Barazanchi and D. H. Rasheed, "The role of green technologies in mitigating carbon footprints in industrial sectors," *ESTIDAMAA*, vol. 2024, pp. 30–35, 2024. https://doi.org/10.70470/ESTIDAMAA/2024/005

[66] L. Alan, and A. Satyanarayan, "Accessible visualization via natural language descriptions: A four-level model of semantic content," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 1, pp. 1073–1083, 2021. https://doi.org/10.1109/TVCG.2021.3114770

[67] S. Mohammed, T. kh.Asman, H. M. Salih, and A. M. Mahmood, "A electric vehicle blockchain: Problems and opportunities," *Babylonian Journal of Networking*, vol. 2024, pp. 18–24, 2024. https://doi.org/10.58496/BJN/2024/003

[68] M. Jishkariani, "Assessment of cyber security risks in electrical engineering system of Georgia," *Mesopotamian Journal of Civil Engineering*, vol. 2024, pp. 54–59, 2024. https://doi.org/10.58496/MJCE/2024/008

[69] A. Mahmood, "Experimental study of air jet heat transfer between Orifice and heated plate," *Babylonian Journal of Mechanical Engineering*, vol. 2023, pp. 47–54, 2023. https://doi.org/10.58496/BJME/2023/006

[70] I. Husain *et al.*, "Logic retaining power among genders and effect of time constraints on the performance of undergrads university students," *Journal of Mechanics of Continua and Mathematical Sciences*, vol. 16, no. 4, 2021. https://doi.org/10.26782/jmcms.2021.04.00008

8 AUTHORS

Syed Muhammad Hassan Zaidi is with the Department of Artificial Intelligence & Mathematical Sciences, Sindh Madressatul Islam University, Karachi, Pakistan (E-mail: m.hassan@smiu.edu.pk).

Abdul Hafeez Khan is with the Department of Software Engineering Sindh Madressatul Islam University, Karachi, Pakistan (E-mail: akhan@smiu.edu.pk).

Sarmad Ahmed Shaikh is with the Department of Computer Science Sindh Madressatul Islam University, Karachi, Pakistan (E-mail: sarmad@smiu.edu.pk).

Imtiaz Hussain is with the Department of Artificial Intelligence & Mathematical Sciences, Sindh Madressatul Islam University, Karachi, Pakistan (E-mail: imtiaz@smiu.edu.pk).