Journal of Theoretical and Applied Information Technology
31st December 2024. Vol.102. No. 24

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

MAKING THE VISUALLY IMPAIRED FEEL IMAGES USING DLEVIS (DEEP LEARNING EYE FOR VISUALLY IMPAIRMENTS)

[1]S. MURUGESAN, [2]DR. N. BALAJIRAJA

[1]Research Scholar, J.J. College of Arts and Science (Autonomous), Affiliated to Bharathidasan University,Tiruchirappali Pudukkottai, Tamil Nadu, India.
[2]Assistant Professor, PG and Research, Department of Computer Science, J.J. College of Arts and Science (Autonomous), Affiliated to Bharathidasan University,Tiruchirappali Pudukkottai, Tamil Nadu, India
E-mail: kpsmurugesan@gmail.com, nbalajiraja@gmail.com

ABSTRACT

In the world, 25% of the 285 million persons who suffer from visual impairments (VIs) are totally blind. The thrill of sight is also taken away from people with VIs. Even through visuals, they are unable to appreciate nature, green spaces, or surroundings. Furthermore, social media and recent technological advancements have served as the primary means of bringing people together on a worldwide scale. Common people's lives are impacted by mobile social media, and their usage patterns are the subject of in-depth research. Users with VIs as special user groups, however, are frequently disregarded. Few studies have been done to find out how they connect with popular mobile social media today. By proposing a method that interprets images and provides information to individuals with VIs, this study aims to close these gaps and enable their introduction as a unique user group on social media. The suggested method, which is based on DLTs, is dubbed DLEVIs (Deep Learning Eye for Visually Impairments). The accuracy of the scheme is also assessed in the study, where the findings show relatively low categorization mistakes.
Keywords: *Deep Learning, Visually Impaired, Mobile Social media, Captioning Images, Social Interactions, Social Networking Sites*

1. INTRODUCTION:

The Los Angeles Times (November 11, 2019) estimates that 2,450,000,000 people have a home on social networking sites (SNSs), where they can interact, communicate, socialize, and exchange information. Even if a sizable percentage of SNS users are blind or have VIs, it is challenging to pinpoint their precise numbers, therefore 114 million is a fair estimate. Research on this group of people is necessary to understand how they interact with the outside world and each other through SNSs. It should be noted that over the past 10 years, the average amount of time these users spend on SNSs has almost tripled. Over 100,000 blind and visually impaired people have bought an Apple iPhone since the launch of a screen reader (VoiceOver) in 2007, according to a recent estimate [1]. The main question is how these SNSs help visually impaired persons overcome social isolation by interacting and interacting with other users worldwide. Global data on individuals with VIs is shown in Figure 1

Ages (in years)	Population (millions)	Blind (millions)	Low Vision (millions)	Visually Impaired (millions)
0-14	1,848.50	1.421	17.518	18.939
15-49	3548.2	5.784	74.463	80.248
50 and older	1,340.80	32.16	154.043	186.203
all ages	6,737.50	39.365 (0.58)	246.024 (3.65)	285.389 (4.24)

Figure 1 – Details On Global Visual Impairments

Social media's advancements, which now include features that allow people with specific needs to participate with them, are extremely beneficial to persons with VIs who rely on different technical ways to communicate themselves. In the past, the main resource for the blind and those with VIs was the Braille system [2]. In actuality, social networking sites are crucial for interacting with blind and VI individuals, especially for their

Journal of Theoretical and Applied Information Technology
31st December 2024. Vol.102. No. 24

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

education and social cue awareness. Additional technologies,

such as screen readers that alter written messages on computer screens, have been developed by SNSs in recent years to increase accessibility for individuals with VIs. As a result, there were more active users on SNSs who were impacted by VIs. General explanations for VIs are shown in Figure 2.

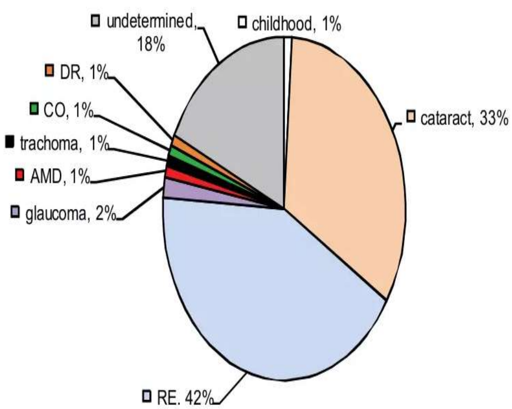

Figure 2 - Global Causes For People With People With Vis

SNSs have made it simpler for people to access a range of duties, especially for the blind who are unable to do their daily tasks in person. Although social media has opened up new opportunities and made many everyday tasks easier for blind people, nothing can truly make up for their loss of sight. This enables persons with VIs to engage with others, work independently, access information, perform errands like shopping, participate in education and training, and navigate the real world. Furthermore, artificial intelligence (AI)-based computer systems can currently identify persons, objects, and landscapes in photos [3]. When people with VIs are aware of the content of visuals in social situations, they can participate in social interactions more fully, learn more from news, and enjoy entertainment or humor [6]. Instead of depending on manually written alternate text, which is often missing [7], or approaching people who can see, which can be time-consuming or burdensome, people with VIs may be able to learn more about these photos with the aid of automatic captioning systems. The effectiveness of AI systems has been evaluated by comparing machine outputs with sighted human outputs, despite the fact that persons with VIs were not taken into consideration when creating them [8]. These are occasionally achieved by user tests in which blind people choose the best caption from a collection of captions or score the quality of the captioning of a particular image [9]. This study examined the usefulness of generated captions on social media for individuals with VIs who use SNSs to access information [10]. The study focused on strangers and accessed a range of information, including news and humor from various relationships. People with VIs rely heavily on captioning, often contributing information to clarify differences between tweet texts and strangely phrased captions where image captions don't seem to match tweet contents or contexts, with the use of contextual questions. They tested the effects of captions on building trust or skepticism online and looked at the phenomenon more broadly. After this introduction, the next section reviews relevant literature, and the third piece provides specifics on the suggested DLEVIs schema. The results of the suggested schema are shown in part four, and this paper is concluded in section five.

2. LITERATURE REVIEW:

People with VIs also utilize SNSs extensively. Prior studies have examined the experiences of individuals with VIs on Twitter and social networking sites. Wu and Adamic [11] found many similarities between sighted and visually impaired SNS users, but they also noted that although sighted users post more status updates, visually impaired users publish and engage with less images. This phenomenon was examined in further detail in the study in [12], which focused on how visually challenged people interact with visual information on social networking sites. In addition to their concern about making a mistake and their discontent with the lack of useful contextual information, they talk about blind users' enthusiasm in participating in simple SNS activities like sharing photos or leaving comments on friends' photos. The study outlined techniques that people may use to assess visual content and found that author-generated descriptions were the most beneficial when available. More recently, SNSs have enabled captioning for photos, which enumerates scenes and items that have been recognized in the photos. Figure 3 depicts social interactions between humans and those with VIs.

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

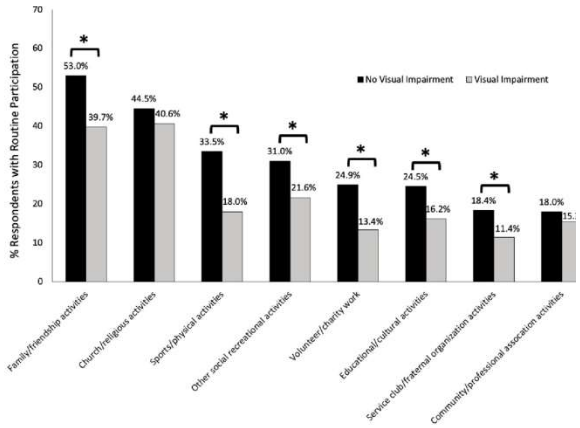

Figure 3 - Social Interactions Of People With People With Vis ([*Https://Journals.Plos.Org/Plosone/Article?Id=10.1371/Journal.Pone.0218540*](Https://Journals.Plos.Org/Plosone/Article?Id=10.1371/Journal.Pone.0218540))

An interestingly new approach to addressing the lack of explanations for social media photos is to automatically generate these alt text captions, even though research has examined how to manually construct better captions [16]. The advances in artificial intelligence needed to do such a task with a respectable degree of accuracy (particularly in computer vision and natural language generation) have generally been the focus of research in this area [17]. These tools are being used in a wide range of applications, including social media platforms, as a result of subsequent achievements. Uniform measurements that compare a machine's output with that of a sighted human are frequently used to evaluate these solutions [18]. These metrics facilitate comparisons between algorithms when they are evaluated on comparable datasets. To compare the quality of human and machine-generated captions, researchers have employed visually impaired individuals to evaluate the quality of a caption for a particular image [19] or asked them to choose the best description from a collection of captions [20]. If accessibility were the primary priority, the evaluation criteria for caption quality and trade-offs between the costs and benefits of specific mistake kinds would be different. A reconsideration of these technologies' fundamental principles is required because they are now being employed for accessibility purposes. Previous studies have shown that people with VIs prefer to hear subtitles that are complete sentences rather than a list of terms [33]. This necessitates advanced ways for converting the data into natural language in addition to advanced computer vision and deep learning techniques. To better emphasize correctness, some SNS implementations give a list of keywords [21]. However, little is known regarding the implications of the wording used in these captions. Some implementations, like Microsoft's Cognitive Services Computer Vision API [36], include the captions into whole phrases. The Framing Effects study [22] looked into several ways to word captions to give viewers a more accurate sense of dependability.

The study of framing effects looks at how people react to information presented differently and how this influences their decision-making. This is frequently done as a comparison of positive and negative framing. For instance, in [23] it was found that people tended to be more risk-averse in the negative framing scenario and more risk-tolerant in the positive framing scenario depending on whether the decision to undergo surgery was presented with the phrasing "X% chance of dying" (negative framing) or "Y% chance of surviving" (positive framing). Comparing natural language with numerical risk descriptors is another technique used to study framing effects. When given numerical values and semantic descriptions (plain English phrases), participants' perceptions of risk were compared in the study in [24]. It was found that when provided descriptions in normal language, people tend to overestimate the likelihood of low probability events. Framing effects have been examined in a range of contexts and have been used to guide the development of persuasive technologies in the field of HCIs (Human Computer Interactions), which has investigated the use of frame picture captioning as a method of accurately delivering captions [25]. While attempting to assist individuals with VIs by suggesting the technique DLEVIs as an alternate eye for individuals with VIs, this research work considers earlier studies on picture captioning.

3. DLEVIS SCHEMA:

The study in [13] looked at how persons with VIs used text-based Twitter interfaces. For people with VIs, the impact was changed by the increased image-sharing choices [14]. Their research highlighted problems with the growing popularity of photographs. The ability to attach photo alt text descriptions to a tweet has since been provided by Twitter [15], but it must be specifically enabled. Figure 4 displays the proposed DLEVIs schema.

Journal of Theoretical and Applied Information Technology
31st December 2024. Vol.102. No. 24

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

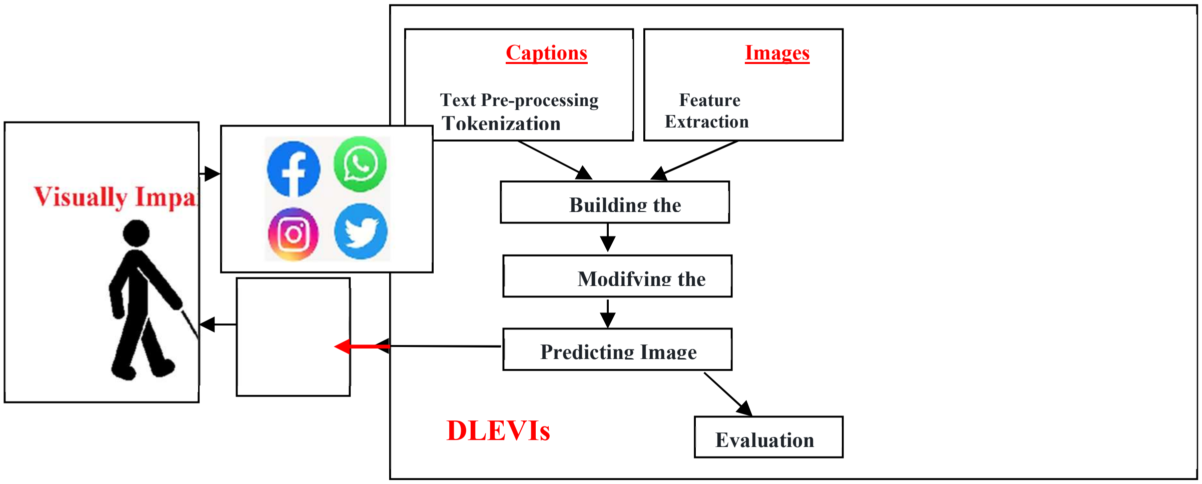

Figure 4 – Schema Of Dlevis

DLTs are used in DLEVIs to pre-process caption texts. Cleaning texts and then tokenizing and encoding them are examples of text pre-processes. Since digital circuits may manipulate groups of bits, encoding information is carried out using only bits (0, 1). The difficulty lies in determining how many bits are needed to encode a given piece of information and how this encoding should be carried out. In this work, visual features are simultaneously retrieved from images. Machine image loading, reading, and processing are intricate procedures. All images are interpreted by machines as a matrix of integers. The size of this matrix is actually determined by the number of pixels in the entering image. Each pixel's pixel values describe or show the color and brightness of that specific pixel. In the most basic example, a pixel value of a 1-bit number indicates the foreground or background of a binary image. Thus, they are numbers that represent the pixel's brightness or intensity. White is represented by numbers nearer 255, whereas black is represented by values nearer zero. As a result, machines can identify images without eyes. Using Google Voice, the schema for this study creates a model based on DLTs to predict image contents with captions for users with VIs. The classification accuracy of the captions of the suggested DLEVIs in this work is assessed. Algorithm1 is a list of the procedures used in this paper.

4. ALGORITHMIC STEPS OF THE PROPOSED DLEVIS

Inputs : Social Media Images (Image Datasets)
Outputs: Audio Descriptions of Captioned Texts for images

Read Image IM = $\{Im_1.......Im_N\}$ // where 1..N number of pixels in Image
Read Caption Texts
Captions preprocessing:
  convert to lowercase
  remove special characters and punctuations
  split text into words
  Add startseq to end
  Add enddeq to end
 For i = 1 to words count
  Tokenize Words
  Encode Words
 End for
// Feature Extractions from images using DLT
  Define DLT model's input/output layers
  img_size = 224
  For i – 1 to N
   //Convert pixels into image array img = img_to_array(img)
    For j = 1 to pixels count
      Image array [j] = Image pixel [j]
    End for
  End for
 features = extracted features from Image array
// DLT Model Building
  Define Inputs
  Define dense nets
  Define relu activations
  Join Image features and Captions
  Define caption model with categorical_loss and crossentropy
  Define adam optimizer
// Model Modification for performance
 Pass DLT Model Outputs to fully connected layers
 DLT Caption Predictions

Journal of Theoretical and Applied Information Technology
31st December 2024. Vol.102. No. 24
© Little Lion Scientific

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

```
    feature = features[]
    in_text = "startseq"
    for i in range(max_length):
        sequence = tokenized texts (pad_sequences max_length)
        ypred = model.predict([feature,sequence])
        word = idx_to_word(y_pred, tokenizer)
        if word is None:
            break
        in_text+= " " + word
        if word == 'endseq':
            break
    return Captions
//Caption Predictions and  Image Descriptions
Pred captions (Captions) :
    // Example #r '<start> black dog is digging in the snow <end>'
    # Using Google Text to Speech API (online), which will convert the caption to audio
    speech = gTTS('Predicted Caption : ' + pred_caption, lang = 'en', slow = False)
    play (Captions)
```

5. RESULTS AND DISCUSSIONS:

Python 3 running on an AMD processor running Windows 10 was used to implement the suggested DLEVIs. The results by stage are displayed as figures. Natural Language Processes (NLPs) are data science subfields that deal with textual data in pre-processing. To answer business problems, text data is also used in addition to numerical data. Data must be pre-processed before being used for analysis or forecasting. As the first steps in getting textual data ready for model builds, DLEVIs carry out text preprocessing, which includes removing special characters. The pre-processing output of the DLEVIs schema is shown in Figure 5.

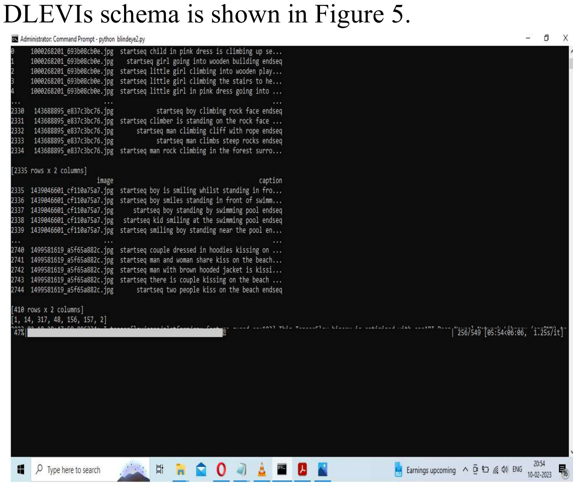

Figure 5 - Pre-Processing Output Of Dlevis Schema

Tokenization, which is the process of replacing a sensitive piece of information, such as a credit card number, with a value known as a token, is the pre-processing of caption texts that DLEVIs uses. Sensitive information still frequently needs to be securely stored for future use. The foundation of tokenization techniques is the protection of sensitive values through the generation of surrogate values and their mapping back to the original values. The tokenization and encoding outputs of the DLEVIs schema are shown in Figure 6.

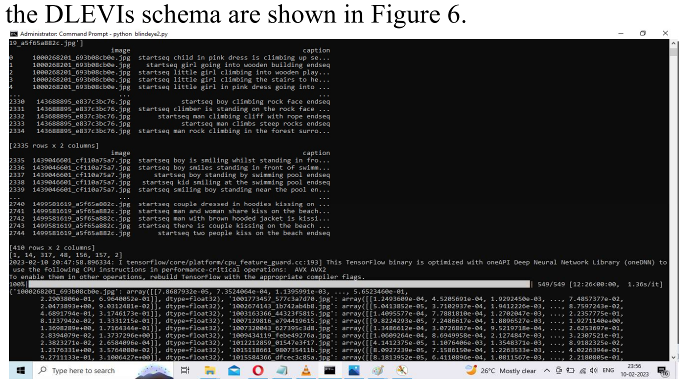

Figure 6 - Dlevis Schema's Tokenization

DLEVIs Image Feature Extractions: Object detection is aided by feature extraction from photos. In colored graphics, the three matrices, or channels, Red, Green, and Blue, are used to represent the pixel colors of the image, with values ranging from 0 to 255. By superimposing these three channels, a colored image is produced. Both features and pixels have the same quantity. The most important aspect of these massive data sets is the large number of variables they contain. It requires a lot of processing power to process these variables. Feature extraction helps to extract the best feature from those enormous data sets by selecting and merging variables into features, hence reducing the amount of data. Therefore, feature extraction is a component of the dimensionality reduction method, which makes processing easier by breaking down a starting set of raw data into digestible chunks. DenseNet architecture is used in the DLEVIs schema to extract features from images. Furthermore, the last DenseNet model layers that DLEVIs use are the Global Average Pooling layer. The study's sample loaded photos are shown in Figure 7.

Journal of Theoretical and Applied Information Technology
31st December 2024. Vol.102. No. 24

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

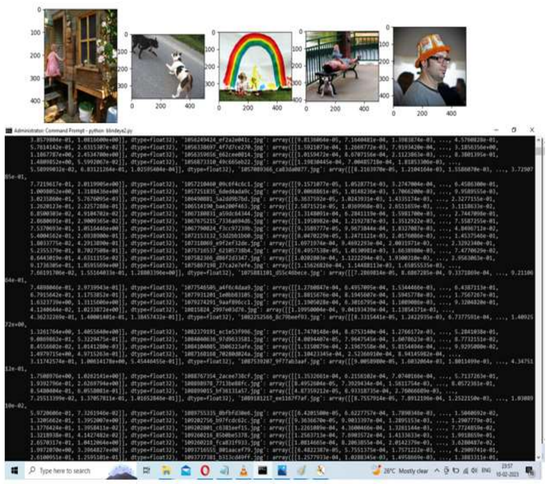

Figure 7 – Sample Images Used In The Study

Model Building in DLEVIs: Since training an image caption model, like training any other neural network, requires a lot of resources, the first stage is data generation. Since we cannot put the data into the main memory at once, we must generate the data in the required format in batches. The image embeddings and matching caption text embeddings will be the inputs for the training process. In order to create captions, the text embeddings are passed word by word at inference time. The network of DLTs starts producing words after each input, finally finishing a sentence, after receiving the image embedding representations and the phrase's first word, starseq. The top thirty terms utilized and the words found in the schema are shown in Figure 8.

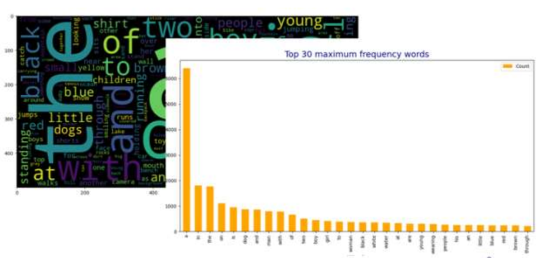

Figure 8 – Caption Word Assessments In Dlevis

Model Modifications in DLEVIs: To improve speed, a little modification has been made to the original model architecture. Before being delivered to the fully linked layers, the LSTMs' output is improved using the picture feature embeddings. The performance of the first suggested model is slightly improved by this. Figure 9 illustrates the DLT model used by DLEVIs.

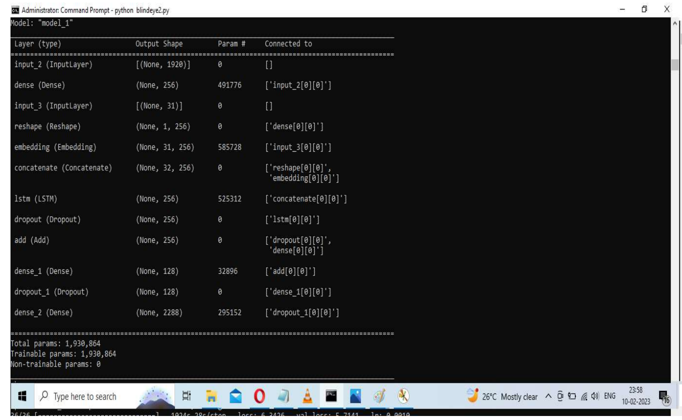

Figure 9 – Model Of Neural Networks Used By Dlevis

Predicting Image Captions: In order to predict and explain image contents to the blind, this technique blends picture features with tagged texts. Neural networks are used to combine many modalities into a single representation space by combining unimodal features or joint representations, which can be done at different points in the model's architecture. The model in this work initially extracts text features (by combining word position data with the WordPiece tokenized phrases) and image area characteristics for each pair of picture texts. This guarantees the genericity of the learned embeddings. Before learning the joint embedding, the extracted features undergo a series of transformations. The results of embedded photos and captions are shown in Figure 10.

Figure 10 - Output Of Embedded Captions And Images

Evaluations of DLEVIs: Starting with sample data (often called "training data"), the majority of machine learning algorithms extract features. These traits are used to construct a mathematical model, which is then used to produce judgments or predictions without being explicitly programmed to do so. Since improving the model is the aim, monitoring model performance on a validation set is an excellent way to get input on how well it performs. The losses incurred by the suggested system during training or learning were evaluated. Training a model is essentially learning (figuring

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

out) acceptable values for each weight and bias from labeled examples. The process by which a machine learning algorithm builds a model in supervised learning by examining multiple examples and searching for a model that minimizes loss is known as empirical risk minimization. Loss is the result of a bad prediction. To put it another way, loss is a gauge of how badly the model forecasted a single instance. If the model's prediction is correct, there is no loss; if not, there is a greater loss. The goal of training a model is to identify a collection of weights and biases that, on average, have low loss across all cases. Loss-based model training is depicted in Figure 11, with the high loss model on the left and the low loss model on the right.

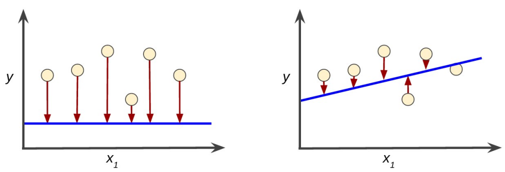

Figure 11 – Model's Losses While Training Samples

The proposed model was trained using only 20 EPOCHS. At 4-5, the ultimate loss rate did not change. Additional evaluations with different epochs were not employed because the objective was to understand caption integration with images rather than to construct the state-of-the-art model. By calculating the likelihood of words based on their frequency in the given vocabulary list, effective searches made it possible to forecast captions by identifying the keywords with the highest probabilities. Figure 12 displays the loss rate graph for the proposed DLEVIs architecture. The output and chart demonstrate how effectively the suggested approach predicts photographs with captions.

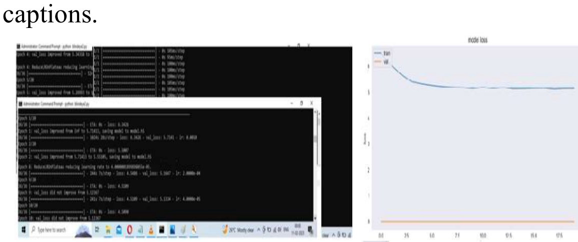

Figure 12 - Loss Rate Graph Of The Proposed Dlevis Schema

6. CONCLUSIONS AND FUTURE WORK

Around the world, millions of people suffer with VIs; 25% of them are completely blind and have never experienced the joy of sight again. Even through photographs, they are unable to appreciate the surrounding environment or events. Furthermore, social media and contemporary technical developments have become the main tools for connecting people worldwide. Regular people's lives are being impacted by mobile social media, and a lot of research is being done to find out how they use it. Conversely, users with VIs are often disregarded as distinct user groups. There haven't been many research done to learn more about how they interact with the popular mobile social media sites of today. A tool that can assist individuals with VIs in understanding social media photos has been proposed by this research project. In a broader sense, our work has uncovered important problems that persons with VIs have when trying to lead normal lives. In order to enable the blind to view photographs on social media, this effort has combined image processing with natural language processing and speech technologies. The proposed schema DLEVIs have been applied with good results and few mistakes. The creation of descriptive sentences that may be used at a regional level, that is, across several languages, is the future scope of this effort.

REFERENCES

[1] Qiu, S., Hu, S., & Rauterberg, G. (2015). Mobile social media for the blind: preliminary. Observations. Proceedings of the International Conference on Enabling Access for Persons with Visual Impairment (ICEAPVI),12-14)

[2] 1. Jiménez, J., Olea, J., Torres, J., Alonso, I., Harder, D., Fischer, K. (2009). Biography of Louis Braille and Invention of the Braille Alphabet. Survey of Ophthalmology, 54(1).

[3] Hao Fang, Saurabh Gupta, Forrest Iandola, Rupesh K. Srivastava, Li Deng, Piotr Dollar, Jianfeng Gao, Xiaodong He, Margaret Mitchell, John C. Platt, C. Lawrence Zitnick, and Geoffrey Zweig. 2015. From captions to visual concepts and back. In Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on, 1473–1482.

[4] Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Jill Schiller. 2017. Automatic Alt-text: Computer-generated Image Descriptions for Blind Users on a Social Network Service. In Proceedings of the 20th ACM Conference on Computer Supported Cooperative Work and Social Computing. https://doi.org/10.1145/2998181.2998364

[5] Meredith Ringel Morris, Annuska Zolyomi, Catherine Yao, Sina Bahram, Jeffrey P. Bigham, and Shaun K. Kane. 2016. "With

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

Most of It Being Pictures Now, I Rarely Use It": Understanding Twitter's Evolving Accessibility to Blind Users. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16), 5506–5516.

[6] Meredith Ringel Morris, Annuska Zolyomi, Catherine Yao, Sina Bahram, Jeffrey P. Bigham, and Shaun K. Kane. 2016. "With Most of It Being Pictures Now, I Rarely Use It": Understanding Twitter's Evolving Accessibility to Blind Users. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16), 5506–5516.

[7] Morten Goodwin, Deniz Susar, Annika Nietzio, Mikael Snaprud, and Christian S. Jensen. 2011. Global Web Accessibility Analysis of National Government Portals and Ministry Web Sites. Journal of Information Technology & Politics 8, 1: 41–67.

[8] Nasrin Mostafazadeh, Ishan Misra, Jacob Devlin, Larry Zitnick, Margaret Mitchell, Xiaodong He, and Lucy Vanderwende. 2016. Generating Natural Questions About an Image. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.

[9] Krishnan Ramnath, Simon Baker, Lucy Vanderwende, Motaz El-Saban, Sudipta N. Sinha, Anitha Kannan, Noran Hassan, Michel Galley, Yi Yang, Deva Ramanan, Alessandro Bergamo, and Lorenzo Torresani. 2014. AutoCaption: Automatic caption generation for personal photos. In Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on, 1050–1057.

[10] Erin L. Brady, Yu Zhong, Meredith Ringel Morris, and Jeffrey P. Bigham. 2013. Investigating the Appropriateness of Social Network Question Asking As a Resource for Blind Users. In Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13), 1225–1236.

[11] Shaomei Wu and Lada A. Adamic. 2014. Visually Impaired Users on an Online Social Network. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14), 3133–3142.

[12] Violeta Voykinska, Shiri Azenkot, Shaomei Wu, and Gilly Leshed. 2016. How Blind People Interact with Visual Content on Social Networking Services. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16), 1584–1595.

[13] Michele A. Burton, Erin Brady, Robin Brewer, Callie Neylan, Jeffrey P. Bigham, and Amy Hurst. 2012. Crowdsourcing Subjective Fashion Advice Using VizWiz: Challenges and Opportunities. In Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '12), 135–142.

[14] Jack Dorsey. 2011. search+photos. Twitter Blogs. Retrieved August 12, 2016 from https://blog.twitter.com/2011/searchphotos

[15] Todd Kloots. 2016. Accessible images for everyone. Twitter Blogs. Retrieved August 12, 2016 from https://blog.twitter.com/2016/accessible-images-for-everyone

[16] Valerie S. Morash, Yue-Ting Siu, Joshua A. Miele, Lucia Hasty, and Steven Landau. 2015. Guiding Novice Web Workers in Making Image Descriptions Using Templates. ACM Transactions on Accessible Computing 7, 4: 1–21.

[17] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In Proceedings of The 32nd International Conference on Machine Learning, 2048–2057.

[18] Ting-Hao Kenneth Huang, Francis Ferraro, Nasrin Mostafazadeh, Ishan Misra, Aishwarya Agrawal, Jacob Devlin, Ross Girshick, Xiaodong He, Pushmeet Kohli, Dhruv Batra, and others. 2016. Visual Storytelling. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

[19] Kenneth Tran, Xiaodong He, Lei Zhang, Jian Sun, Cornelia Carapcea, Chris Thrasher, Chris Buehler, and Chris Sienkiewicz. 2016. Rich Image Captioning in the Wild. Proceedings of CVPR 2016.

[20] Jacob Devlin, Hao Cheng, Hao Fang, Saurabh Gupta, Li Deng, Xiaodong He, Geoffrey Zweig, and Margaret Mitchell. 2015. Language Models for Image Captioning: The Quirks and What Works. In ACL – Association for Computational Linguistics, 100.

[21] Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Jill Schiller. 2017. Automatic Alt-text: Computer-generated Image Descriptions for Blind Users on a Social Network Service. In Proceedings of the 20th ACM Conference on

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

Computer Supported Cooperative Work and Social Computing.
[22] Amos Tversky and Daniel Kahneman. 1985. The Framing of Decisions and the Psychology of Choice. In Environmental Impact Assessment, Technology Assessment, and Risk Analysis: Contributions from the Psychological and Decision Sciences, Vincent T. Covello, Jeryl L. Mumpower, Pieter J. M. Stallen and V. R. R. Uppuluri (eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 107–129.
[23] Theresa M. Marteau. 1989. Framing of information: Its influence upon decisions of doctors and patients. British Journal of Social Psychology 28, 1: 89–94.
[24] Sean Young and Daniel M. Oppenheimer. 2009. Effect of communication strategy on personal risk perception and treatment adherence intentions. Psychology, Health & Medicine 14, 4: 430–442. https://doi.org/10.1080/13548500902890103
[25] Young-Ho Kim, Jae Ho Jeon, Eun Kyoung Choe, Bongshin Lee, KwonHyun Kim, and Jinwook Seo. 2016. TimeAware: Leveraging Framing Effects to Enhance Personal Productivity. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16), 272–283.