Enhanced Landmines Recognition with Deep Learning Models Using Generative AI Image Augmentation

Rudwan A. Husain
Faculty of Information Technology, University of Tripoli,
r.husain@uot.edu.ly

Abstract

Maintaining the high accuracy of landmines recognition is an essential objective for deminres safety. Those who are involved in mine actions need to be informed about the true type, and the current condition, of the unexploded ordinance they will be dealing with. Therefore, timely and accurate definition of such lethal items is a crucial task in explosives hazard areas. Deep Learning models have been experimented in landmines recognition, and have shown impressive result. However, the lack of field realistic data sets disrupts adequate model training efficiency. The deterioration and damage of landmine structures have caused huge safety concerns. Corrosion has been identified as a big reason for the deterioration and damage, which causes demining members confused with other materials. Hence, this paper proposes a landmine data sets enrichment method based on Generative Artificial Intelligence tools that are used to produce more realistic images for an augmented landmines data set. The proposed process is based on OpenAI Dall-E image generator as the landmine data augmentation tool, and MobileNetV2 deep learning model as the landmine recognition tool. The goal is to effectively predict the corrosion behavior of steel structured landmines. The system can simulate corrosion based on the dataset collected from different internet resources. Based on comparative experiments, this system demonstrates outstanding performance and outperforms the baseline model. Thus, the method proposed in this study can be a promising tool to assist in building DL agents deployed in mobile applications useful for landmines and UXOs identification.

Keywords: Landmines, Generative AI, Deep Learning, Dataset Augmentation.

التعرف المحسن على الألغام الأرضية باستخدام نماذج التعلم العميق باستخدام زيادة الصور التوليدية بالذكاء الاصطناعي

رضوان علي بلقاسم حسين
كلية تقنية المعلومات، جامعة طرابلس،
r.husain@uot.edu.ly

الملخص

الحفاظ على الدقة العالية في التعرف على الألغام الأرضية هو هدف أساسي لضمان سلامة العاملين على نزع الألغام. يحتاج أولئك الذين يشاركون في عمليات نزع الألغام إلى أن يكونوا على علم بالنوع الفعلي والحالة الحقيقية للمخلفات غير المنفجرة التي سيتعاملون معها. لذلك، فإن تعريف هذه العناصر القاتلة بشكل سريع ودقيق هو مهمة حاسمة في مناطق الخطر الناجمة عن المتفجرات. تم تجربة نماذج التعلم العميق في تعريف الألغام الأرضية، وقد أظهرت نماذج التعلم العميق نتائج مثيرة للإعجاب. ومع ذلك، فإن نقص مجموعات البيانات الحقلية الواقعية يعوق كفاءة تدريب النموذج بشمل كاف. أدى تدهور وتلف هياكل الألغام الأرضية إلى وجود مخاوف كبيرة لسلامة العاملين على نزعها. تم تحديد التآكل كسبب كبير للتدهور والتلف، مما يجعل أفراد نزع الألغام يشعرون بالارتباك مع مواد أخرى مشابهة للمخلفات المتفجرة. ولذا، يقترح هذا البحث طريقة تحسين مجموعات بيانات الألغام الأرضية بناءً على أدوات الذكاء الاصطناعي التوليدية التي تستخدم لإنتاج صور أكثر واقعية لمجموعة معززة لبيانات الألغام. تستند العملية المقترحة على مولد الصور OpenAI Dall-E كأداة لزيادة بيانات الألغام الأرضية، ونموذج التعلم العميق MobileNetV2 يستخدم كأداة للتعرف على الألغام. الهدف هو التنبؤ بفعالية بسلوك التآكل للألغام ذات الهياكل المعدنية. يمكن للنظام محاكاة التآكل استنادًا إلى مجموعة البيانات التي تم جمعها من مصادر إنترنت مختلفة. بناءً على التجارب المقارنة، يظهر هذا النظام أداءً متميزاً ويفوق النموذج الأساسي. وبالتالي، يمكن

أن تكون الطريقة المقترحة في هذه الدراسة أداة واعدة للمساعدة في بناء وكلاء التعلم العميق التي توظف في تطبيقات الهواتف المحمولة المفيدة للتعرف على الألغام الأرضية والمخلفات غير المنفجرة.

الكلمات الدالة: الألغام الأرضية، الذكاء الاصطناعي التوليدي، التعلم العميق، زيادة مجموعة البيانات

1. Introduction

Landmines are still commonly used in armed conflicts, posing a serious threat to civilian communities even long after the conflict has ended. Proper detection of buried objects is essential for subsequent clearance operations. Among the widely used equipment in humanitarian efforts, electromagnetic induction metal detectors stand out due to their affordability and user-friendliness. The metal detector's response signal is affected by several factors, including detector technology, the type of target, its orientation, and the properties of the surrounding soil. However, in countries like Libya, they still depend on the experience of the deminers to identify the unexploded ordinance (UXO). Common devices such as smart mobile phones can be utilized to help in definition of UXOs, if supported by Artificial Intelligence (AI) models.

The existence of explosive remnants of war (ERW), and specifically antipersonnel landmines, are considered as a serious threat for civilians as well as militaries around the World (Shire *et. al*., 2015). Many countries have suffered from mining for the last seven decades (John, 2019). According to Libyan Mine Action Center regarding mines and other UXO victims in Libya, it has been reported that 19 people were killed, including 14 children, killed by explosives remnants of war only in 2022. Since May 2020 the were 180 accidents reported causing 369 victims (LibMAC, 2023). The UN Department of Humanitarian Affairs (UNDHA) set a group of strict legislations that determine the strategy of civil area demining. Never-the-less, there is no accurate estimation of these areas which were trapped by landmines. In fact, in order to consider an area free-

mining, 99.6% of remnants of war must be safely eliminated. Generally, landmines detection, localization, and recognition are the steps followed by most landmine clearness systems. In reality, most of these systems depend on metal detectors and on deminer experience. Thus, with the purpose of protecting the lives of civilians and militaries, and facilitating demining operations, this paper focuses on the recognition problem which is essential to determine whether an object is an APM or not, and specify its model type (Bestagini *et. al*., 2018).
Corrosion in Metal devices, such as landmines, is caused by chemical and electrochemical interactions. This phenomenon is usually seen in environmental conditions featuring a high level of moisture. There are different kinds of corrosion such as general corrosion which occurs as uniformly distributed nonprotective flakes of rust and pitting which is a localized point of corrosive attack (Hoang and Tran, 2019). Manual methods performed by human to generate augmented datasets for damaged or corroded landmines are labor intensive and time consuming. Moreover, the processes of data processing and reporting are also very tedious for human technicians. Therefore, there is a practical need to come up with a more productive and accurate method of image dataset augmentation.

Deep learning (DL) is an important area of machine learning that accomplishes classification tasks straight from images, video, texts or sounds, DL is described as a major enhancement of the Artificial Neural Networks (ANN). DL model architectures are consisting of a considerable number of neuron layers that give a superior level of abstraction and progression to the target data. Therefore, deep learning is count as the key automatic learning technique in computer vision and image processing fields (LeCun *et. al*., 2015). Some of the advanced models of deep learning include AlexNet, VGG net, GoogleNet, ResNet from, and You Only Look Once (YOLO) (Shu *et. al*., 2019). CNN plays an important role mostly in analyzing visual data items. A CNN architecture consists of a number of layers, which include a convolutional layer, a pooling layer, a ReLU layer, a fully connected layer, and a loss layer. Where the main layer is considered to be the Convolutional layer, that is due to the fact that it composed of kernel filters. The filters are to identify target image features, and the pooling layer is used to decrease the number of parameters in the NN. The ReLu layer represents the activation function that is used to set any negative weight values to the zero. The fully connected layer smoothens the output features. Finally, Loss layer illustrates the differences between the predicted labels and the actual values.

2. Related Work

Elwaer *et. al.* in 2021 experimented Multiple Deep Learning Models for Antipersonal Landmines Recognition. The findings of the experiments reveal that MiniGooglenet, ResNet, and MobileNet models achieve a very good level of accuracy between 86% and 97%, on 1750 test landmine images. In fact, MiniGooglenet seemed to be the valuable choice for experiment with the best results on landmine dataset. Authors recommended to improve their landmine dataset by increasing the number and the diversify of dataset images. In 2022, Jiang and Hirohata build a simulation system based on Generative Adversarial Network (GAN) for data augmentation in order to enrich the dataset, and used MobileNetV2 as the discriminator. Their goal was to effectively predict the corrosion deterioration and damage behavior of uncoated steel structures over time and under different circumstances. Another work used GAN in dataset augmentation was done by Haile Woldesellasse and Tesfamariam in 2023. They generated new data samples to utilize a conditional generative adversarial network (cGAN) for dealing with imbalance problem in a corrosion dataset. Utility of the cGAN data augmentation was evaluated by training an ANN model.
For demining purposes, a number of methods have been designed and developed. In general, most of these detection methods are made of three main components; for capturing signs, a sensor is found; an image processor for handling the obtained signals; and for landmine recognition, a component of decision making is used (Kasban *et. al*., 2010). Several studies focused on landmine detection rather than recognition. The authors of (Ward *et. al*., 2019) proved that the standard deep learning techniques can be applied for

automating image recognition of landmines. Furthermore, the study (Jang *et. al*., 2012) proposed and assessed algorithms with the purpose of tackling the problem of detection and identification of landmines using GPRs. Genc and Akar (Genc and Akar, 2019) classified the techniques for buried target detection into four categories: shape-based, physics-based, and image-based techniques, and convolutional neural networks (CNNs).

3. Methodology

This study aims to solve the class imbalance problem in the landmines and other UXOs databases. Most of data samples are collected from the internet UXO websites. The photos shown on the internet were taken in labs for items in perfect conditions. Due to fact that explosive devices are always treated securely, and not easy to take pictures of, as a result, class imbalance handling techniques are used to generate new samples related to the dangerously damaged UXO classes. The techniques developed include standard image augmentation, and Generative Artificial Intelligence (GAI) technique. The methodology includes dataset collection, and evaluating the utility of these techniques using artificial neural network deep learning algorithm. Prior to any model development, the corrosion dataset was pre-processed and divided into training and testing dataset, each containing 80% and 20% of the corrosion augmented data, respectively. The pre-processing steps includes removing outliers from the corrosion database and data normalization.

3.1. Landmines Dataset Collection

A set of 855 samples of Landmines were obtained. The original dataset was collected for training and testing purposes, and it was gathered using one of the following three ways: Firstly, downloading images manually from landmines data websites such as CAT UXO (Brownlee, 2019). This way consumed a lot of time simply due to the amount of human work involved. Secondly, using an automated python script works to download landmine images from the internet, which helped gathering images faster and with less effort. Finally, a few images were captured from real mine fields in Libya using a mobile phone. The dataset is divided into two classes of Landmines images, metal and nonmetal mines. Sample images of metal mines are given in Figure 1, with names written underneath each type. Additionally, the size of a training dataset was expanded using data augmentation. Therefore, a new dataset of 7000 samples were generated of augmented images after using data augmentation technique.

Figure 1. Samples of metal landmines collected dataset.

3.2. Data Augmentation

In order to build effective Deep Learning models, it is important to keep the validation error as low as possible while keeping the training error to a minimal. The effectiveness of this strategy may be proven by using data augmentation. Image data augmentation includes several techniques used for increasing the size of training datasets and add variety to datasets in order to build more accurate deep learning models. However, modified versions of original images were added into the dataset (Wang and Perez, 2017).

3.2.1. Standard Augmentation

In many machine learning tasks, Data augmentation is a commonly used technique. A variety of augmentation strategies have been proposed and shown to capture important characteristics of natural images. The augmentation strategies can generate variants of the images, which can increase the fit models' capacity to generalize their knowledge to new images. One of the more effective data augmentations strategies is the traditional transformations, include flipping, Rotations, Zooming, and shifting. To perform the basic transformation, only affine transformations are used (Simonyan and Zisserman, 2015). Each input image was output in two distinct forms; an exact copy shifted (width/height), zoomed out, rotated, flipped(horizontal), sheared, or rescaled. Eight augmentation strategies were utilized to generate new training sets. The following Table 1, illustrates the data augmentation techniques used in this experiment for randomly generated images.

Table 1. Standard Applied Augmentation Techniques

Action	**Description**	**Value**
Rotation	Rotational augmentation can be performed by turning the image clockwise or counterclockwise	40
Rescale	In scaling or resizing, the image is resized to the given size.	64x64
Shear	Shear the image with increasing rate magnitude along the horizontal (vertical) axis.	0.2
Zoom	Zooming is for randomly zooming inside pictures.	0.2
Flipping	Randomly flipping half of the images.	True
Width shift	Shift the image along the X-axis.	0.2
Height shift	Shift the image by along the Y-axis.	0.2
Fill mode	Filling the area that was left over of shifting.	Nearest

To illustrate the effect of augmentation strategies that were used for the purpose of this study, Figure 2, presents a set of the training images before and after augmentation.

3.2.2. Generative AI Augmentation

Within the field of computer vision, it is highly complex to generate an image from various sources of data, such as text, scene graphs, and object layouts. Additionally, manually capturing images from multiple perspectives to create objects or products can be laborious and costly.

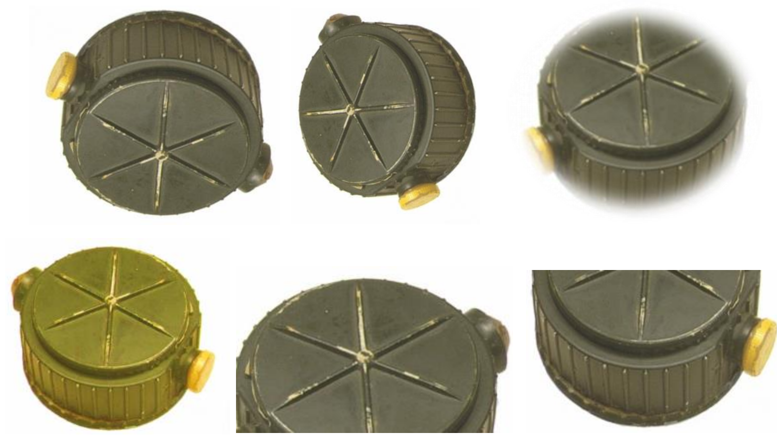

Figure 2. Sample augmentation effects applied on FMK – 1 landmine.

However, thanks to advancements in deep learning and artificial intelligence, it is now feasible to generate new images from diverse types of data. (Elasri *et. al.*, 2022) Consequently, significant efforts have been dedicated to the development of image generation techniques, leading to remarkable accomplishments in this area. OpenAI company introduced the image generation tool DALL-E, which is a highly advanced model derived from GPT-3. DALL-E has been meticulously trained to generate visually coherent images by leveraging text descriptions. This training routine involved a comprehensive dataset containing annotated pairs of text and corresponding images. Impressively, DALL-E exhibits an expansive array of capabilities, encompassing the capacity to produce anthropomorphized representations of animals and objects, seamlessly integrate disparate concepts into plausible visual compositions, faithfully render textual input as images, and apply complicated transformations to pre-existing visual content. In the work, DALL-E is utilized for dataset augmentation in order to produce new corroded landmine images. Figure 3 illustrates the structural procedure of the generation process, which is based on input landmine image, augmentation prompt, and image augmentation mask. Initially, the variables of the operating system under use have to identified, followed by establishing the connection with OpenAI DALL-E library. Default output image size was set to 256x256. The augmentation prompt text describes the visual effects requested from GAI to be applied on the target input image.

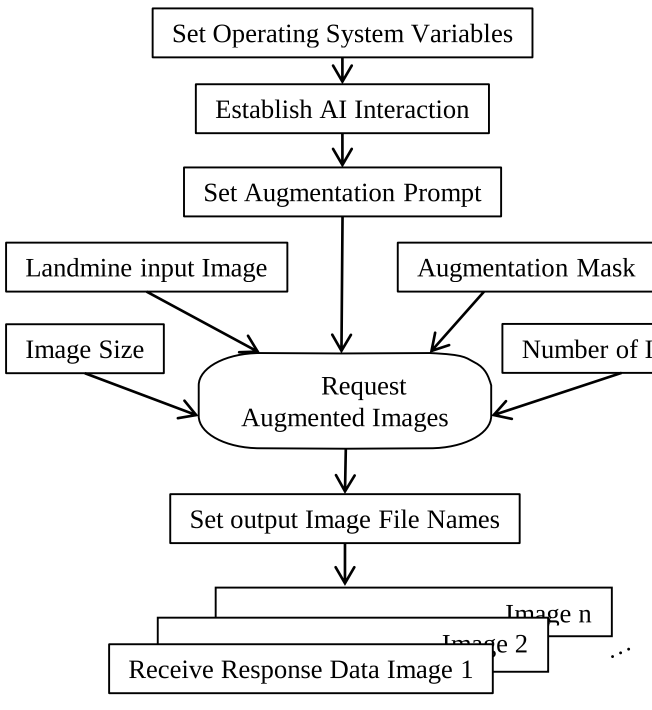

Figure 3. Structure of augmented image generation with GAI DALL - E.

Generated AI images of 20, for each landmine sample, were stored in in separate files. These new images are to be used in training, testing, and validation of the MobileNetV2 deep learning model. Sample generated image of DALL-E dataset augmentation is given in Figure 4 showing corroded metal landmines. The Augmentation mask specifies the landmine shape as a transparent area. This will set the landmine boundaries out of the image background and other surround environmental objects.

International Science and Technology Journal
المجلة الدولية للعلوم والتقنية
Volume 32 العدد
April 2023 ابريل

Figure 4. Corroded metal landmine images generated by DALL-E.

4. Experiments and Results

The experiments doubted the deep learning model MobileNet recommended by Elwaer *et. al.* in 2021for accurate landmines identification mobile Apps. The DL model was trained with a high-performance computing (HPC) unit, which has the following specifications: AMD Ryzen 2600 CPU, 16 GB RAM, and Nvidia GTX 980TI. Data processing and training was carried out using Keras library (Kaiming *et. al*., 2015). This API was built using Python, and runs on TensorFlow (Rosebrock, 2017).

4.1. MobileNetV2 DL model

The MobileNet DL model is built on a simplified design that utilizes depthwise separable convolutions to construct lightweight deep neural networks. The model introduces 2 global hyper-parameters which make an effective trade-off between latency and accuracy. Depthwise Separable Convolution is a core layer of MobileNet (Janocha and Czarnecki, 2017), which is a type of factorized convolution in which a conventional convolution is factorized into a depthwise and pointwise convolution. In the first convolution, each input channel receives a single filter. The second convolution then uses one-by-one convolution to combine the depthwise convolution outputs. the first layer, which constitutes a complete convolution, the MobileNet has 28 layers. In ImageNet, the model obtains an accuracy of 0.895 percent for top-5 tests. This ANN DL model is to be examined for the original dataset, as well as the newly augmented dataset that includes the GAI data.

4.2. Results Discussion

The result illustrates in Figure 5 shows that after only 40 epochs the MobileNet model achieved an overall accuracy of ≈ 98% training accuracy. Loss on both the training and validation data continues to fall with limited number of minor "spikes" due to the learning rate staying constant and not decaying. At the end of the 140th epoch, it is reaching 99% training accuracy on original training dataset.

Experimental results give in Figure 6 below demonstrate how after only 25 epochs the MobileNet model achieved an overall accuracy of ≈ 99% training accuracy. At the end of the 80th epoch, it is reaching 99% training accuracy on the GAI augmented training dataset. The above plot shows a typical graph as the loss decreases each time the accuracy increases, moreover the training and validation loss and accuracy mimic each other indicating that the model network is learning the underlying patterns without overfitting. In addition, the result illustrates in Figure 6 shows that after very much fewer epochs the MobileNet model achieved an overall accuracy of ≈ 99% training accuracy.

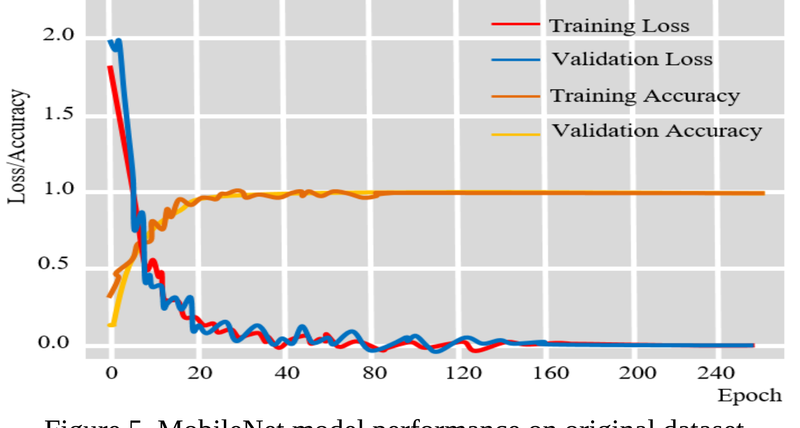

Figure 5. MobileNet model performance on original dataset.

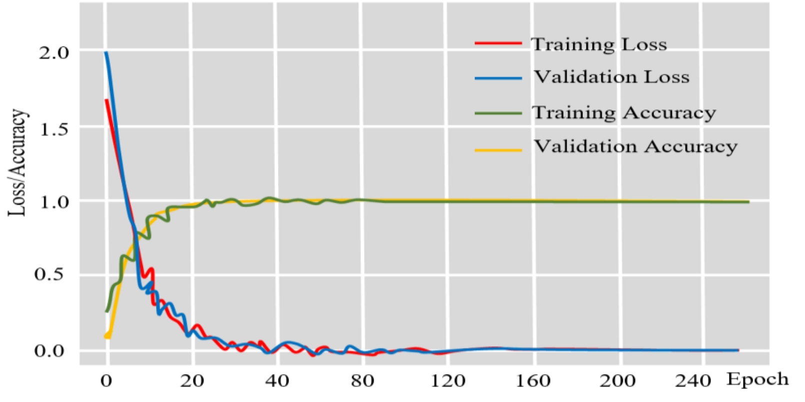

Figure 6. MobileNet model performance on GAI augmented dataset.

5. Conclusions

A dataset augmentation framework based on limited realistic samples for landmines recognition based on image generation AI is proposed in this paper. The effort done on this research was motivated by the rising number of UXO victims in Libya, especially of those who are working on humanitarian mine actions. The method illustrated here is aiming at solving the problem of the unsatisfactory generalization ability of deep learning models when trained on limited number of realistic data samples. The augmentation model used two transformations such as standard geometric image transformations, and Generative AI image tools to deal with complex visual transformations. To facilitate the data collection for the MobileNetV2 deep learning landmine recognition model, DALL-E was utilized to create a synthetic dataset of natural effects on landmines. With DALL-E's capabilities, it was possible to generate 20 different images for each class metal mines. The effect was metal corrosion, and other effects can be used too. In addition, diverse of backgrounds, lighting conditions, and natural environmental objects can be added for more realistic field activities. The resulting dataset was both realistic and varied, providing the proposed model with an impressive learning resources. To further enhance the dataset, a technique called data augmentation using the Keras ImageDataGenerator was used. This technique allowed the model to generate additional images with variations in lighting, orientation, and other image transformations. The augmented dataset generated here was more robust and better able to generalize to naturally unseen UXO datasets. The fine-tuned pre-trained MobileNet model for landmines recognition task using deep learning, it was possible to achieve an accuracy of almost 99% on the validation data with less than 30 training and validation epochs. These results are demonstrating the effectiveness of proposed method. The integration of AI-generated data augmentation, standard data augmentation, and deep learning enables the development of precise and efficient models, presenting opportunities for various humanitarian mine action endeavors.

6. References

[1] Abdelhamid Elwaer, Marwa Solla, Fatima Ben Lashihar, Hassan Ebrahem, Hala Shaari, and Rudwan Husain, "*Applying Multiple Deep Learning Models for Anti-Personal Landmines Recognition*", Proceedings of The Libyan Conference on Automation and Robotics (LCAR-2021) No. 1, July 7-8, 2021, Tripoli – Libya.

[2] Nhat-Duc Hoang, and Van-Duc Tran, *Image Processing-Based Detection of Pipe Corrosion Using Texture Analysis and Metaheuristic-Optimized Machine Learning Approach*, Hindawi, Computational Intelligence and Neuroscience, 2019.

[3] F. Jiang, M. Hirohata, *A GAN-Augmented Corrosion Prediction Model for Uncoated Steel Plates*, Applied Science. 2022.

[4] Haile Woldesellasse and Solomon Tesfamariam, *Data augmentation using conditional generative adversarial network (cGAN): Application for prediction of corrosion pit depth and testing using neural network*, Journal of Pipeline Science and Engineering, Elsevier, 2023.

[5] Mohamed Elasri, Omar Elharrouss, Somaya Al-Maadeed, and Hamid Tairi, Image Generation: A Review, Neural Process Letter, 4609–4646, 2022.

[6] A. Shire, U. Jawarkar and M.S. Chavhan, "Overview of Buried Mines Detection Using Radar Bullet", *International Journal of Engineering Sciences & Research Technoloey*, vol. 4, no. 2, pp. 250-254, 2015.

[7] John, P. J, Review Paper on Object Detection using Deep Learning- Understanding different Algorithms and Models to Design Effective Object Detection Network, International Journal for Research in Applied Science and Engineering Technology, 7(3), 1684–1689, 2019.

[8] LibMAC, Statement on international day for mine awareness and assistance in mine action, https://lmac.gov.ly/?p=814, 2023.

[9] P. Bestagini, F. Lombardi, M. Lualdi, F. Picetti and S. Tubaro, "Landmine detection using autoencoders on multi-polarization GPR volumetric data", arXiv:1810.01316, 2018.

[10] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” nature, vol. 521, no. 7553, p. 436, 2015.

[11] M. Shu, ‘‘Deep learning for image classification on very small datasets using transfer learning,’’ M.S. thesis, Iowa State Univ., Ames, IA, USA, 2019.

[12] H. Kasban, O. Zahran, Sayed M. Elaraby, M. El-Kordy and F. E. Abd El-Samie, "A Comparative Study of Landmine Detection Techniques", Sens Imaging 11:89–112, Springer, 2010.

[13] Chris M. Ward, Josh Harguess, Cameron Hilton, Chelsea Mediavilla, Keith Sullivan, and Rick Watkins, “Deep learning for automatic ordnance recognition,” Geospatial Informatics IX, 10992, 109920H, International Society for Optics and Photonics, 2019.

[14] Gyubin Jang, Kyungmi Park, and Kangwook Kim, “GPR-based landmine detection and identification using multiple features,” Int. Jurnal of Antennas Propag., Vol. 2012, 1–7, 2012.

[15] A. Genc and G. B. Akar, "Combination of physics-based and image-based features for landmine identification in ground penetrating radar data", *J. Appl. Remote Sens.*, vol. 13, no. 2, pp. 1, Apr. 2019.

[16] Jason Brownlee, "deep learning for computer vision: image classification, object detection and face recognition in python", http://machinelearningmastery.com, 2019.

[17] Jason Wang, and Luis Perez, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning", In: Stanford University research report, https://arxiv.org/abs/1712.04621, 2017.

[18] Karen Simonyan, and Andrew Zisserman, "Very Deep Convolutional Networks For Large-Scale Image Recognition", https://arxiv.org/abs/1409.1556v6, 2015.

[19] He Kaiming, Zhang Xiangyu, Ren Shaoqing, and Sun Jian, "Deep Residual Learning for Image Recognition", https://arxiv.org/abs/1512.03385, 2015.

[20] Adrian Rosebrock, "Deep Learning for Computer Vision with Python: Starter Bundle", PyImageSearch, 2017.

[21] Katarzyna Janocha, and Wojciech Marian Czarnecki, "On Loss Functions for Deep Neural Networks in Classification," Theoretical Foundations of Machine Learning, https://arxiv.org/abs/1702, 2017.