Smart Shield: Innovative Content Filtering for Safe

(Browsing on INTERNET)

|  |  |  |
| --- | --- | --- |
| Andrew Adel Andrews *Computer Science*  Department.  Faculty of Computer & Information Sciences, Ain Shams University  Cairo, Egypt  [andrew.adel.a@outlook.com](mailto:andrew.adel.a@outlook.com) | Veronica Emad William *Computer Science*  Department.  Faculty of Computer & Information Sciences, Ain Shams University Cairo, Egypt | Mehrael Ashraf Rahif *Computer Science*  Department.  Faculty of Computer & Information Sciences, Ain Shams University Cairo, Egypt |
| Monica Sameh Azer *Computer Science* Department*.* Faculty of Computer & Information Sciences, Ain Shams University  Cairo, Egypt | Mariz Erian Ibrahim  *Computer Science*  Department.  Faculty of Computer & Information Sciences,  Ain Shams University Cairo, Egypt | Filobateer Essam Motamed  *Computer Science Department.*  Faculty of Computer & Information Sciences, Ain Shams University  Cairo, Egypt |

***Abstract*— The internet, despite its vast array of resources and benefits, also serves as a conduit for harmful content such as cyberbullying, hate speech, violence, and inappropriate material. This exposure is particularly alarming for vulnerable individuals, including children and adolescents, who may inadvertently come across such detrimental content. Therefore, it is crucial to have effective strategies to ensure a safer online environment. To address this concern, our project focuses on developing a sophisticated real-time content filtering extension that uses advanced machine learning techniques to classify both text and images.**

**The primary features of our project include:**

**1. Text Classification:** The system identifies and filters harmful text content such as cyberbullying, hate speech, and offensive language.

**2. Binary Image Classification:** It distinguishes between violent and nonviolent images.

**3. Multiclass Image Classification:** The system categorizes images into specific classes, including fire, accidents, damaged buildings, and normal content.

**In developing this system, we curated and prepared diverse datasets for both text and image classification, addressing challenges such as imbalanced data and irrelevant content. Through extensive experimentation with various models, we identified the most effective approaches for each classification task. Specifically, we selected DistilBERT [1] for text classification, MobileNetV3 [2] for binary image classification, and EfficientNetV2B2 [2] for multiclass image classification**

**The results of our developed system have demonstrated its ability to effectively filter harmful content in real time, thus significantly enhancing the safety and inclusivity of the online environment. By proactively identifying and mitigating the risks associated with harmful content, our project contributes to fostering a positive and secure digital space for all users. This, in turn, promotes a healthier and more supportive online community, free from the detrimental impacts of cyberbullying, hate speech, and exposure to violent or inappropriate material.**

**Contribute to the well-being of online communities by minimizing the spread of harmful content and fostering a healthier digital environment.​**

**Overall, Respect users' freedom of expression while addressing the challenges associated with content filtering, harmonizing freedom of expression and mutual respect​**

***Keywords—content Filter, Cyberbullying Detection, Extension, Python, Natural Language Processing.***

1. INTRODUCTION

Our motivation stems from the increasing prevalence of harmful online content, particularly hateful speech and violence, on social media and educational platforms. This content negatively impacts user's well-being, highlighting the urgent need for a safer online environment. We are dedicated to creating a secure space for vulnerable users, ensuring they are not exposed to inappropriate and violent material. Despite the internet's essential role in daily life, it also harbors harmful content, especially on social media platforms where inappropriate and violent material is prevalent. Although some graphic content is filtered, many posts containing violent imagery merely receive warnings and remain accessible. Additionally, curse words often evade removal by being obscured with various symbols. The challenge, therefore, is to develop an effective filtering system to ensure a safe and positive online experience for all users, Any content that can cause emotional, psychological, or physical harm to individuals. This includes text and images containing threats, insults, offensive language, or any form of aggressive behavior online.

A white paper with black text

Description automatically generated

Fig. 1. Harmful Content Categories Covered

Traditional machine-learning approaches offer simplicity, efficiency, and interpretability, making them suitable for certain applications, especially with limited data and computational resources. They often fall short in handling complex, large-scale, and context-rich data compared to deep learning models, which can automatically learn intricate patterns and features from the data.

Deep learning approaches offer significant advantages in terms of automatic feature extraction, contextual understanding, and performance on large-scale and complex datasets for both text and image classification. These benefits come with challenges, including high computational requirements, large data needs, and difficulties in interpretability.

One of the key user groups targeted by this project is unsupervised children who browse the web. These young users are particularly vulnerable to encountering harmful content due to their limited understanding of online risks and the potential for content filtering systems to fail to protect them adequately.

1. Related Work
   * *Social Media Platforms*

The related work primarily discusses the content filtering and moderation efforts of major social media platforms like Facebook and Twitter. These platforms employ a combination of automated systems, utilizing machine learning algorithms and natural language processing (NLP) techniques, along with human moderators to review flagged content. The related work also mentions Instagram and X (formerly Twitter) and their content guidelines, highlighting the challenges these platforms face in enforcing their policies due to the vast amount of user-generated content. [3] [4]. In contrast, this project focuses on developing a browser extension for real-time content filtering. While the project draws inspiration from the techniques employed by social media platforms, it aims to provide a more personalized and user-controlled filtering experience. The browser extension will leverage machine learning models and rule-based filtering to empower users to tailor their content preferences and block harmful content according to their individual needs. Additionally, the project emphasizes addressing biases in content filtering by utilizing diverse datasets and incorporating user feedback, which is a distinct aspect not explicitly mentioned in the related work concerning social media platforms.

TABLE 1. Social media platforms guidelines

|  |  |  |
| --- | --- | --- |
| **Platform Name** | **Guidelines** | **Observations From previous experience** |
| Instagram  [21] | •Intellectual Property  • Appropriate Imagery • Spam  • Illegal Content  • Hate Speech, Bullying and Abuse  • Self-Injury  • Graphic Violence | • Even though  Graphic content is filtered out there are posts that contain graphic and violence imagery that just get a warning and can be viewed.  • There are curse words written in a lot of posts and separated by a variety of symbols that don’t get removed |
| Facebook [22] | • Violence and Criminal Behavior  • Objectionable content • Integrity and authenticity |
| X (Twitter) [23] | CNN |

Instagram: Instagram's guidelines focus on intellectual property, appropriate imagery, spam, illegal content, hate speech, bullying, abuse, self-injury, and graphic violence. The platform aims to maintain a positive and safe environment for its users by prohibiting harmful and offensive content [21].

Facebook: Facebook's guidelines address a broader range of issues, including violence, criminal behavior, objectionable content, and maintaining integrity and authenticity. The platform strives to create a space where users can connect and share experiences while upholding community standards [22].

X (Twitter): X's guidelines specifically target hateful references, incitement, slurs, tropes, dehumanization, and hateful imagery, including profile 8 information. The platform is committed to combating hate speech and promoting healthy online discourse [23].

TABLE 2. Related work table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Authors** | **Dataset** | **Focus of paper** | **Models used** | **Results** |
| Rubio, J. L. S., Almeida, A. V., & Segura-Bedmar, I. (2023) [9]. | dataset comprising 2,996 Spanish tweets related to violent acts | Detection of Violent Incidents on Social Networks (Text) | BERT | 0.918  F1- score |
| Escalante-Hernandez, A., Joaquín-Arellano, L., Lavalle-Martínez, J. D. J., Villaseñor-Pineda, L., & Jair Escalante, H. (2023) [14] | DA-VINCIS dataset of 5000 tweets in Mexican and Spanish | Identification of violent incidents using social networks (Text) | MLP | 0.65  F1- score |
| Pericherla, S., & Ilavarasan, E. (2023) [7]. | MS-COCO,  Flickr 30K | Cyberbullying detection in image (images) | RoBERa | Accuracy  88.82% |

The first paper utilizes a Decision Tree model for text classification, achieving a remarkable accuracy of 99.90%. Decision Trees are known for their interpretability and ability to handle both categorical and numerical data, making them a suitable choice for this task [9]. • The

second paper employs a Multinomial Naive Bayes (MNB) model, also for text classification, but with a lower accuracy of 64.00%. MNB is a probabilistic model that assumes feature independence, which may not 9 always hold true in real-world text data, potentially explaining the lower accuracy [14]. • The third paper focuses on image classification and utilizes a Medium Gaussian SVM model, achieving an accuracy of 98.73%. SVMs are effective in high-dimensional spaces and can handle non-linear relationships between features, making them well-suited for image analysis [7].

• HaramBlur Extension:

HaramBlur is a browser extension designed to filter out content that may be considered inappropriate or offensive according to Islamic values. It also aims to protect user privacy and reduce browsing distractions. The extension automatically detects, and blurs content deemed "Haram", allowing users to navigate the web while adhering to their religious beliefs [24]. While both HaramBlur and our project share the goal of content filtering, there are key differences in their scope and target audience. HaramBlur is specifically 11 tailored for users who wish to filter content based on Islamic values, whereas our project aims to create a more general-purpose content filter that can be customized to individual preferences. Additionally, HaramBlur focuses on blurring inappropriate content, while our project aims to block or remove it entirely. Another distinction lies in the type of content being filtered. HaramBlur primarily targets images and videos, while our project encompasses both text and image filtering. This broader scope allows our project to address a wider range of harmful content, including cyberbullying, hate speech, and violence material.

TABLE 3. Models used in previous work

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Name** | **Accuracy** | **Image / text** | **Reference** |
| Decision Tree | 99.90% | Text | [4] |
| Multinomial Naïve Bayes (MNB) | 64.00% | Text | [4] |
| AdaBoost Classifier | 45.50% | Text | [4] |
| Logistic regression | 77.70% | Text | [4] |
| Linear SVC | 97.00% | Text | [4] |
| Stochastic gradient descent optimizer (SGD) | 92.80% | Text | [4] |
| Random Forest | 99.00% | Text | [4] |
| Bagging classifier | 93.80% | Text | [4] |
| Medium Gaussian SVM | 98.73% | Image | [11] |
| KNN | 82.26% | Image | [11] |
| Logistic regression & SVM | 75.00% | Text | [16] |
| decision tree, SVM & Naive | 91.00% | Text | [16] |
| Combinational Network for Bullying Detection (CNBD) | 91.40% | Image | [7] |

Table 3 presents a list of machine learning models used in previous research for text and image classification tasks, along with their reported accuracy.

• Models 1-8 are designed for text classification, while models 9-10 are for image classification.

• Model 11 combines logistic regression and SVM for text classification, while model 12 combines decision trees, SVM, and Naive Bayes for the same purpose.

• Model 13, the Combinational Network for Bullying Detection (CNBD), is specifically designed for image-based bullying detection.

Table 3 presents a list of machine learning models used in previous research for text and image classification tasks, along with their reported accuracy.

• Models 1-8 are designed for text classification, while models 9-10 are for image classification.

• Model 11 combines logistic regression and SVM for text classification, while model 12 combines decision trees, SVM, and Naive Bayes for the same purpose. • Model 13, the Combinational Network for Bullying Detection (CNBD), is specifically designed for image-based bullying detection.

1. PROPOSED METHODOLOGY

The system architecture we have designed is a comprehensive framework that encompasses multiple layers (as shown in Figure 2) to facilitate a rich and interactive user experience. Each layer of the architecture serves a specific purpose and contributes to the overall functionality of the system.

A diagram of a diagram

Description automatically generated

Fig. 2. System Architecture

Let's delve into the key layers that comprise this system architecture:

***1. Application Layer***

− User Interface and communication layer where the end-user interacts with the application.

• Functionalities:

1) Choose Content-Type: The user selects the type of content to be filtered.

2) Select Specific Categories to display: The user can choose specific image categories to whitelist and bypass the filtering.

3) Activate the Extension: The user initiates the extension to start filtering the content on the current active web page.

4) Blurring Decision: The application injects a blurring script based on the prediction of the models, to blur text or images that match the user's filtering criteria

**2. Logic Layer:**

• Processes the data collected from the Application Layer and applies the trained models to get the filtering decision.

• Functionalities:

1) Preprocessing:

− Images: Converts to image arrays, resizes, and expands dimensions as needed.

− Text: Removes irrelevant data, eliminates empty strings, performs tokenization, and adds padding.

2) Classifiers:

− Image classification:

+ EfficientNet V2 B2 model to classify images into 4 classes: Fire, Damaged Building, Accidents, and Normal.

A diagram of a diagram of a flowchart

Description automatically generated

Fig.3 EfficientNetV2[3]

+ MobileNet V3 model to classify images into 2 classes: Violent and Non-violent.

A diagram of a block diagram

Description automatically generated

Fig. 4. Mobilenet V3[2]

− Text classification:

+DistilBERT model to analyze text and identify inappropriate language and topics.

A diagram of a process

Description automatically generated

Fig. 5. DistilBERT Transformer [1]

3) Mapping Fetched Data:

− Maps each data item to its filtering decision

**3. Data Layer:**

− Stores the datasets used to train the deep learning

models.

− Provides an interface to access and load the trained

models during operation.

**A. Intended Users:**

Our intended user is anyone that browses the internet daily that faces a lot of undesired content and needs help to filter out unwanted content.

Such as:

1. Parents and Guardians:

• Parents who want to protect their children from

inappropriate or harmful content online.

• Guardians oversee the online activities of minors to

ensure a safe browsing environment.

2. Educational Institutions:

• Schools and universities to ensure students are not exposed to harmful content while using school-provided internet and devices.

• Libraries and other educational facilities to provide a safe browsing environment for all users.

3. Content Creators, Moderators and Social Media Users:

• Safeguard users from offensive and harmful content.

• Bloggers, vloggers, and online communities want to ensure their content remains clean and appropriate for all audiences.

• Moderators of forums and online communities who need tools to help maintain a safe and respectful environment.

4. Mental Health, Support Groups and Healthcare Providers:

• Prevent exposure to triggering or harmful content in

online support communities.

• Clinics and hospitals that want to protect patients,

especially those in vulnerable mental states, from

exposure to triggering content.

**B. User Characteristics**

One of our distinct features is that the user does not need to have any technical skill or technical background to be able to operate the extension since it has a user friendly interface that makes it accessible for all ages and tech proficiency levels.

*IV. Datasets*

Our dataset is merged from 5 different datasets.

1. Text Datasets

TABLE 4.Textual dataset

A table with numbers and words

Description automatically generated

− Challenges:

• Some datasets contained abbreviated slang words (lol, asap).

• It also contained irrelevant content such as days, months, mentions, links, etc.

• Most of the datasets were imbalanced.

− How we solved them:

• Searched for a dataset that contains the original words of the slang word abbreviation.

• Created a dictionary to replace them with their original words. • Ex: lol = laughing out loud, asap = as soon as possible.

•Removed irrelevant content by

• Creating a list of short and long versions of each word.

• Ex: Mon = Monday, Apr = April.

• Used regular expressions to remove these special characters.

• Applied down sampling as a balancing technique.

B) Image Datasets

− Binary Image Dataset

TABLE 5. Binary Image Classification Dataset

A screenshot of a computer

Description automatically generated

Both datasets were clean and devoid of significant issues.

However, to enhance the deep learning model's performance, the decision was

made to merge them.

By combining the datasets, we capitalized on their strengths, resulting in a more

diverse and effective training set.

− Multiclass Image Dataset

Our dataset is merged from 12 different datasets as each category is a combination of selected folders from 3-5 different datasets.

TABLE 6. Multiclass Image Classification Dataset

A screenshot of a data sheet

Description automatically generated

• For Normal and Fire:

− Both categories were clean and free of significant issues.

• For Accidents and Damaged Buildings:

− Challenges:

• Not enough data available.

• Available datasets were very small.

• Other datasets were video frames and didn’t have scene

varieties. 26

− How we solved them:

• Extracted the good images from the available datasets.

• Resized the images to dimensions of 224x224 pixels as

a preparatory step for utilization with pretrained models.

• Applied data augmentation on the extracted images to

increase the size of the dataset.

• Combined all the found images and the augmented ones

into one dataset.

*V. RESULTS*

In this section, we explored various model approaches by trying traditional, modern, and deep learning techniques. We enhanced each model with suitable preprocessing methods and improved their performance through hyperparameter tuning. Finally, we documented each experiment thoroughly.

− Text Results

TABLE 7. Text Classification Results

A graph with numbers and text

Description automatically generated

− Binary Image Results

TABLE 8. Binary Image Classification Results

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Train**  **Accuracy** | **Validation**  **Accuracy** | **Train Loss** |
| ResNet50 | 99.48% | 98.8% | 0.0284 |
| MobileNetV3 | 98.78% | 98.87% | 0.2812 |
| EfficientNetB0 | 97.6% | 97.34% | 0.0671 |
| VGG19 | 94.7% | 94% | 0.1451 |
| DensNet121 | 91.6% | 90% | 0.2140 |
| MobileNetV2 | 89.48% | 89.75% | 0.2528 |
| Xception | 85.15% | 78.8% | 0.3617 |
| AlexNet | 80% | 79.7% | 0.5149 |
| NASNetLarge | 87.7% | 78.9% | 0.2935 |
| InceptionV3 | 75.45% | 65.13% | 0.5812 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Models** | **epochs** | **Validation**  **Loss** | **Test**  **Accuracy** |
| ResNet50 | 40 | 0.0504 | 98.8% |
| MobileNetV3 | 10 | 0.9846 | 98.55% |
| EfficientNetB0 | 40 | 0.0702 | 97.3% |
| VGG19 | 27 | 0.1491 | 94.6% |
| DensNet121 | 40 | 0.2664 | 91.9% |
| MobileNetV2 | 40 | 0.2570 | 89.74% |
| Xception | 27 | 0.4223 | 84.4% |
| AlexNet | 20 | 0.6003 | 79.7% |
| NASNetLarge | 40 | 0.4127 | 78.9% |
| InceptionV3 | 9 | 0.8313 | 76% |

sA blue squares with white text

Description automatically generated

Fig. 6. DistilBERT Confusion Matrix

A graph with a line

Description automatically generated

Fig. 7. DistilBERT Accuracy Function

A graph showing a loss function

Description automatically generated

Fig. 8. DistilBERT Loss Function

A diagram of a graph

Description automatically generated with medium confidence

Fig. 9. MobileNetV3 Confusion Matrix

A graph of a line

Description automatically generated

Fig. 10. MobileNetV3 Accuracy Function

A graph of loss and validation

Description automatically generated

Fig. 11. MobileNetV3 Loss Function

We selected these models based on their efficiency and compact size, considering the trade-off among accuracy, performance, and efficiency within the constraints of our available resources.

− Multiclass Image Results

TABLE 9 Multiclass Image Classification Results

A table of numbers and symbols

Description automatically generated

*VI. CONCLUSION*

Create a safer environment for our vulnerable users that do not

wish to be exposed to this type of inappropriate and violent content.

Performed sequence classification instead of token classification to

be able to understand hidden meanings form inappropriate text.

We've implemented two methods: binary and multiclass image

classification. Our filtration combines the parallel predictions of both models to determine the outcome. Our extension operates universally across all web pages, enhancing the browsing experience by filtering out inappropriate content from online sources, not just limited to social media platforms.

*VII.* *Future Work*

* + Enhance the accuracy of models while simultaneously reducing the latency of their predictions.
  + Image Deploy the models in the cloud to leverage the benefits of cloud computing infrastructure and services.
  + Enable the system to support multiple languages, expanding accessibility and enhancing user experience for a global audience
  + Integrate the ability to scan and read text from images, enhancing functionality by detecting whether the text should be filtered.

*VIII. REFERENCES*

[1] Adel, Hadeer, Abdelghani Dahou, Alhassan Mabrouk, Mohamed Abd Elaziz, Mohammed Kayed, Ibrahim Mahmoud El-Henawy, Samah Alshathri, and Abdelmgeid Amin Ali. "Improving crisis events detection using distilbert with hunger games search algorithm." Mathematics 10, no. 3 (2022): 447.

[2] Koonce, Brett, and Brett Koonce. "MobileNetV3." Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization (2021): 125-144.

[3] Tan, Mingxing, and Q. V. Le. "Efficientnetv2: Smaller models and faster training. arXiv 2021." arXiv preprint arXiv:2104.00298.

[4] Sahana, V., Anil Kumar KM, and Abdulbasit A. Darem. "A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Form Spring in Textual Modality. “

[7] Pericherla, Subbaraju, and E. Ilavarasan. "Overcoming the Challenge of Cyberbullying Detection in Images: A Deep Learning Approach with ImageCaptioning and OCR Integration." International Journal of Computing and Digital Systems 14, no. 1 (2023): 1-xx.

[9] Rubio, Jorge Luis Saavedra, Alejandro Valbuena Almeida, and Isabel Segura-Bedmar. "UC3M at Da-Vincis-2023: using BETO for Detection of Aggressive and Violent Incidents on Social Networks." In Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), CEUR Workshop Proceedings. CEUR-WS. org. 2023.

[11] Khan, Muhammad Attique, Usman Tariq, Taerang Kim, and Jae-Hyuk Cha. "Anomalous Situations Recognition in Surveillance Images Using Deep Learning.“

[14] Escalante-Hernandez, Alejandro, Luis Joaquín-Arellano, Jose de Jesús Lavalle-Martínez, Luis Villaseñor-Pineda, and Hugo Jair Escalante. "Towards the Monitoring of Violent Events in Social Media through Visual Information." Computación y Sistemas 27, no. 1 (2023): 153-162.

[16] Bharadwaj, Vedadri Yoganand, Vasamsetti Likhitha, Vootnoori Vardhini, Adari Uma Sree Asritha, Saurabh Dhyani, and M. Lakshmi Kanth. "Automated Cyberbullying Activity Detection using Machine Learning Algorithm." In E3S Web of Conferences, vol. 430, p. 01039. EDP Sciences, 2023.

[21] Instagram Community Guidelines FAQ’s.

[22] Facebook Community Standards | Transparency Center.

[23] X's policy on hateful conduct | X Help

[24] Alganzory. (2024). HaramBlur: Browser extension that allows you to navigate the web with respect for your Islamic values [Software].

GitHub. <https://github.com/alganzory/HaramBlur>

[25] KARANDEEP, S. (2021, November). Real Life

December 20, 2023 from November). Violence vs. Non-Violence: 11K Images Dataset, Version 1. Retrieved December 20, 2023

[27] Monica, S. (2024, February). Text Datasets, Version 1. Retrieved December

20, 2023 from <https://www.kaggle.com/datasets/monicasamehsalib/text-datasets>.

[28] SHIKAMARU. (2023, September). Toxic Comments Detection, Version 1.

Retrieved December 24, 2023 from

<https://www.kaggle.com/datasets/shikamaru073/toxic-comments-detection>.

[29] CHANGLU, G. (2023, December). AIDERdata. Retrieved January 25, 2024, from <https://www.kaggle.com/datasets/clguo1/aiderdata>.

[30] KUTAY, K. (2021, July). Forest Fire. Retrieved January 25, 2024, from

<https://www.kaggle.com/datasets/kutaykutlu/forest-fire>.

[31] Almond, N. (2023, December). Ukraine Images 2023. Retrieved January 26,

2024, from <https://www.kaggle.com/datasets/almondnguyen/ukraine-images-2023>.

[32] Arash, F. (2022, June). Wildfire surveillance. Retrieved February 10, 2024,

from <https://www.kaggle.com/datasets/arashfarahdel/wildfire-surveillance>.

[33] Rupak, R. (2022, March). Cyclone Wildfire Flood Earthquake Database.

Retrieved February 10, 2024, from

<https://www.kaggle.com/datasets/rupakroy/cyclone-wildfire-flood-earthquakedatabase>.

[34] Cristian, C. (2021, July). FOREST FIRE IMAGE DATASET. Retrieved

February 10, 2024, from

<https://www.kaggle.com/datasets/cristiancristancho/forest-fire-image-dataset>.

[35] Amerzish, M. (2023, December). Forest Fire Smoke and Non-Fire Image

Dataset. Retrieved February 20, 2024, from

<https://www.kaggle.com/datasets/amerzishminha/forest-fire-smoke-and-non-fireimage-dataset>.

[36] ISMAIL, M (2021, March). The wildfire dataset. Retrieved February 20,

2024, from <https://www.kaggle.com/datasets/elmadafri/the-wildfire-dataset>.

[37] JOÃO, G (2021, March). Fire Images Database. Retrieved February 20,

2024, from <https://www.kaggle.com/datasets/gondimjoaom/fire-images-database>.

[38] Yaroslav, C (2023, Feb). RescueNet. Retrieved February 20, 2024, from

<https://www.kaggle.com/datasets/yaroslavchyrko/rescuenet>