

Deep Learning Model for Enhanced Surgical Phase Recognition and Tool Detection in Robot Assisted Surgery

Khairunnesa and Dr. Pushpalatha K R

Research of Scholar, Sri Siddhartha Academy of Higher Education, Tumkur, India

Associate Professor Department of CSE(Data Science)

khairunnesa_12@rediffmail.com and pushpalathakr@ssit.edu.in

Abstract: *This study addresses the advancement of surgical phase recognition and tool detection in robot-assisted minimally invasive surgery (RMIS) through the development of a deep learning-based framework. The transition from conventional open surgery to MIS, and further to RMIS, has resulted in significant improvements in patient outcomes, including reduced trauma, lower infection risk, and faster recovery. Despite these benefits, challenges persist in accurately recognizing surgical phases and detecting tools due to high variability in surgical procedures, complex and dynamic visual scenes, and limitations in existing temporal modeling approaches. Traditional methods often rely on fixed temporal convolutions or constant-length video representations, which inadequately capture the diverse temporal dynamics inherent in surgical workflows. Recent advances leverage convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and transformer-based models to enhance feature discrimination and temporal dependency modeling. However, gaps remain in integrating the semantic structure of surgeries, efficiently exploiting multi-task learning, and handling challenging video frames. This research proposes a cascaded deep learning model that jointly addresses surgical phase recognition and tool detection, incorporating advanced architectures and multi-task strategies. The model's performance is evaluated against established benchmarks using various datasets and metrics, demonstrating improved accuracy and robustness. The findings highlight the potential of deep learning to enhance intraoperative decision support, workflow analysis, and surgical quality assurance in RMIS.*

Keywords: Surgical Phase Recognition; Tool Detection; Robot-Assisted Minimally Invasive Surgery (RMIS); Deep Learning; Multi-Task Learning

I. INTRODUCTION

Surgery, in a general case, has evolved over the decades. From conventional open surgery to minimally invasive surgery (MIS), the medical community has shifted its focus on less invasive procedures. MIS has now become the new trend by providing less risk while having added benefits. In basic concept, MIS involves the inserting of trocars, thin and flexible tube-like retractors, through small incisions made on the patient's body. The trocar serves as a channel for minuscule surgical instruments such as a laparoscope or a laparoscopic grasper. At the proximal end of the trocar, the surgeon can control the laparoscopic tool and perform surgical tasks inside the patient's body. With the use of MIS, pain and trauma for the patient are lessened and the risk of infection is greatly lowered [1].

Robotic-assisted surgery is defined as a type of surgery that performs operations with the use of robotic systems. This branch of surgery was developed to enhance a surgeon's capabilities in operation and to overcome the limitations present in MIS. Although more commonly associated with MIS as RMIS, it is also sometimes used in open surgery. Robotic surgery proves to be more advantageous over the conventional MIS due to its additional benefits. For patients, the benefits they will experience is similar to that of MIS. Patients will have shorter hospital time and faster recovery time. Pain and trauma will also be reduced with minimal scarring. Also, the risk of infection will be greatly lowered

and blood loss is reduced. As for surgeons, RMIS has a greater advantage because it provides: enhanced dexterity, greater precision, and better visualization [2].

MINIMALLY invasive surgery (MIS) has been adopted as the preferred treatment method in various medical departments, such as urology, gynecology, and general surgery, due to its advantages, including less trauma, less pain, and quick recovery [3]. With the advancements in robotics, the accuracy and reliability of robot-assisted MIS have further improved. Surgical interaction recognition (SIR) plays a crucial role in navigation decision support for MIS or robot-assisted MIS. The SIR assists in procedure monitoring, surgical path planning, surgical report generation, surgical reconstruction, surgical quality supervision etc. Therefore, the SIR not only shortens the surgery time but also reduces the risks associated with the surgery [4]. Recently, the task of understanding the surgical scene, including the detection and tracking of surgical instruments, surgical instrument segmentation, and surgical workflow identification, has been a focus of the research community.

The deep learning-based algorithms have achieved excellent performance in surgical instrument detection and segmentation. A lightweight cascaded convolutional neural network (CNN), which is capable of achieving real-time surgical instrument segmentation, was proposed by us. Based on instrument detection, the video-driven surgical workflow or surgical phase recognition have become research hotspots. The time series models, such as conditional random fields and hidden Markov models, are often used in a surgical workflow or phase recognition. In addition, the deep learning models are also adopted to predict the surgical workflow or phase [5].

There exist previous attempts to develop automatic methods for surgical phase recognition and tool presence detection based on hand-crafted features and shallow learning methods [6]. However, accurate recognition of surgical phase and detection of surgical tool presence are challenging due to multiple factors, including variability of surgeon's skills [7], variability of patient's conditions such as clinical profiles and the operative regions, and quality variations of recorded videos caused by camera movement, non-uniformity illumination, blood and smoke in the surgical field. To meet the challenges, lots of studies have been dedicated to extracting discriminative visual features from video frames and to modeling the temporal dependencies among frames in order to improve the performance for surgical phase recognition and tool presence detection.

Despite these efforts, however, it is still challenging for existing methods to fully solve the problem. There are great potentials to improve the performance for the following reasons. First, given that temporal information is crucial for both surgical phase recognition and tool presence detection, most of existing approaches model the surgical video dynamics either using temporal convolutions with fixed kernel size, or inputting constant length of video clip representation into Long Short-Term Memory (LSTM) layers [8]. Both forms of variables are too rigid to model the varieties of temporal space for complex surgical workflow analysis. This has motivated the introduction of transformers-based models for surgical workflow analysis.

where an upper triangular mask is used to prevent information leakage from future frames. Such a causal mask enables not only on-line recognition but also long-range temporal dependency by aggregating information via self-attention from all previous frames for the current frame prediction. It does have, however, one disadvantage that it treats the temporal dependency of different tasks the same way. For example, the temporal dependency of the surgical phase recognition task may be different from that of the tool presence detection task. Second, despite high variability, the same type of surgery is usually executed with a standardized and reproducible workflow [9]. This is due to the fact that surgeons are requested to follow guidelines to perform specific operations with corresponding sets of instruments for different surgery phases.

Nevertheless, pure video-based workflow recognition is highly complex. Complicated surgical scenes lead to limited inter-phase variance but high intra-phase variance. Scenes are often blurred due to the frequent motion of the surgical instruments. The changes in lighting conditions, inevitable visual occlusions from smoke and blood, and artifacts introduced by the lens cleaning process also increase the difficulty of surgery perception. As the surgical procedure can be described as a sequence of tasks at different granularity levels, including phases, steps, and activities, with gradually fine-grained definition [10], each phase can be naturally presented as the compositions of actions, in which we observe that these actions are characterized by high variability in their temporal duration.

However, it would be insufficient for these low-level features to represent the complicated surgical visual appearance. With the revolution of deep learning and its successful applications on medical imaging, recent methods proposed to enhance the feature discrimination capability by employing convolutional neural networks (CNNs) [11]. Meanwhile, given that the surgical video is actually a form of sequential data, leveraging the temporal information and effectively capturing the sequential dynamics are crucial for accurate workflow recognition. A number of approaches have also been proposed in this direction by utilizing dynamic time warping, conditional random field, and derivations of hidden markov model (HMM).

Objectives

- To Design and develop a deep learning model for automatic surgical phase recognition in robotic surgery.
- To Design and develop a deep learning model for automatic tool detection in robotic surgery
- To Design and develop a deep learning model for automatic surgical phase classification.
- To evaluate the model using various datasets and deep learning metrics, comparing it with existing techniques to demonstrate its efficiency.

II. RELATED WORK

Surgical phase recognition is another most studied application in surgical video analysis. [12] proposed PhaseNet which was based on single-task learning and comprised a nine-layer CNN. In order to enforce temporal constraints, they used two different approaches: HMM-based and LSTM-based. They further compared the performance of PhaseNet with that of EndoNet, which was designed to perform jointly surgical phase recognition and tool presence detection. Their results indicated that 1) multi-task learning achieved better results than single-task learning; and 2) LSTM-based approaches achieved better results than HMM-based approaches. In their methods, CNNs and LSTM networks were trained separately.

In contrast, [13] designed SV-RCNet which leveraged the complementary information of visual and temporal features learned from surgical videos. This was achieved by integrating a deep residual network and an LSTM. They additionally proposed the prior knowledge inference (PKI) as a post-processing step for improved performance. Based on a CNN and a causal, dilated multi-stage temporal convolutional network (MS-TCN) [14], proposed TeCNO for online surgical phase recognition, exploring long-range temporal relations in precomputed spatial features. Recently, [15] proposed TMRNet for relating long-range and multi-scale temporal patterns to augment the present features. Despite the fact that significant progress has been achieved, none of the previously mentioned work for phase recognition considers the semantic structure of a surgery during network training. The methods in the latter category are trained for multiple tasks. EndoNet is the first method based on multi-task learning for joint surgical tool presence detection and surgical phase recognition without explicitly exploring the interaction between these two tasks.

In contrast, [16] developed a multi-task recurrent convolutional network with correlation loss (MTRCNet-CL) to exploit the relatedness between the surgical phase recognition task and the surgical tool presence detection task. In their method, extracted frame features were used to predict tool information while also serving as input to an LSTM model for the surgical phase prediction. A correlation loss was designed to model the relatedness between tool presence and phase identification of each video frame. introduced MTMS-TCNs, which was an extension to the MS-TCN with multi-task learning, to jointly learn the tasks of phase and step recognition. In several studies, multi-task learning has been shown to achieve better results than single-task learning. However, how to efficiently exploit the relatedness between different tasks is still an open problem.

[17] presented radio frequency identification (RFID) tags for real-time tracking of laparoscopic instruments. Researchers have tried marker-less tool detection using image processing methods based on color or texture. Advancement in deep learning architectures has prompted researchers to achieve good performance on surgical tool presence and phase recognition. used a hybrid embedding aggregation transformer to enhance the importance of spatial features for phase recognition. Some work employed multi-stage architecture, e.g., a predictor stage and a refinement stage, in surgical phase recognition, where the misclassification in the predictor stage can be well rectified during the refinement stage. For example, TeCNO adapted the multi-stage temporal convolution network (MS-TCN) to online

surgery scenario via causal and dilated convolutions, found that directly using MS-TCN brings little improvement and then proposed a new non-end-to-end training strategy.

used a ResNet [18] to predict the binary predictions for the tool presence and then combine the predictions and features for phase recognition. MTRCNet-CL introduced a novel correlation loss to explicitly model the relations between tool presence and phase classification. In addition to the tool information, leveraged more cues, such as management tools, ontology and production rules to improve the performance.

Recently, [19] realize the negative effect of hard frames and propose data cleansing and online hard frames mapper to detect and handle them respectively. [20] find that simply applying multi-stage architecture e.g. multi-stage TCN makes the refinement fall short and thus design not end-to-end training manner to alleviate this problem. OperA leverages attention weight to yield further insights into the decision-making process. Trans-SVNet [21] proposes a hybrid embedding aggregation Transformer to fuse spatial and temporal embedding. Ding and Li emphasize the importance of segment-level semantics and extract semantic-consistent segments to refine the erroneous predictions. Notably, some related methods utilize additional tool presence labels to perform a multi-task learning to facilitate surgical phase recognition.

EndoNet and DeepPhase [22] on cholecystectomy and cataract surgeries, respectively. EndoNet proposed a Convolutional Neural Network (CNN) followed by a hierarchical Hidden Markov Model (HMM) to perform both phase and tool detection. Similarly, DeepPhase introduced an architecture with ResNet [23] and Recurrent Neural Network (RNN), instead of HMMs, for temporal modeling, for both phase recognition and tool detection. EndoLSTM, extended EndoNet by utilizing a Long Short-Term Memory (LSTM) for temporal refinement of spatial features. Similarly, SV-RCNet [24] trained a ResNet and LSTM model end-to-end and proposed a prior knowledge inference scheme for surgical phase recognition. MTRCNet-CL [25] presented a multi-task model to detect tool presence and perform phase recognition along with a novel correlation loss to capture the relationship between tool presence and phase identification.

III. RESEARCH GAP

The research gap identified includes:

Temporal Dynamics Modelling: Existing methods predominantly use fixed kernel sizes in temporal convolutions or constant-length video clip representations, which may be too rigid to capture the diverse temporal dynamics of complex surgical workflows.

Semantic Structure Integration: Most current phase recognition systems do not consider the semantic structure of surgeries during network training, which can be crucial for understanding the context and sequence of surgical actions.

Variability in Surgical Procedures: High variability in surgeon skills, patient conditions, and operative regions poses a challenge, and current systems may not adequately adapt to these variations for accurate phase recognition and tool detection.

Complexity of Surgical Scenes: The complexity of surgical scenes, including blurred imagery due to instrument motion, variable lighting conditions, and visual occlusions, has not been fully addressed, leading to challenges in accurately perceiving surgeries.

Unified Multi-Task Learning Approaches: While multi-task learning shows promise, efficiently exploiting the relatedness between different tasks such as phase recognition and tool detection is still an open problem that requires further exploration.

Handling of Hard Frames in Videos: The negative effect of hard frames in surgical videos is a relatively unexplored area, with potential for development of specific strategies to detect and handle these challenging segments.

Leveraging Additional Data Cues: The use of supplementary data cues such as management tools, ontology, and production rules for improving model performance has not been fully explored in existing research.

Integration of Advanced Architectures: The application of newer deep learning architectures like transformers, which can handle long-range dependencies and complex temporal relationships, is still in its nascent stage in the context of surgical phase recognition and tool detection.

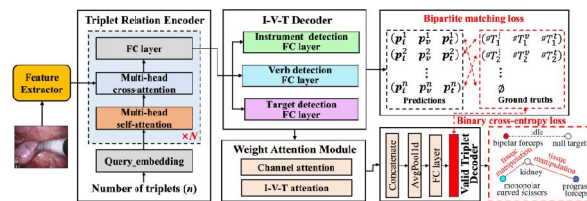
Data Cleansing and Online Mapping Strategies: Strategies for data cleansing and online mapping of hard frames for better model training and performance are areas that need more in-depth research and development.

Segment-Level Semantics in Videos: The importance of segment-level semantics in surgical videos for refining predictions and enhancing model accuracy is another area that warrants further investigation.

IV. PROPOSED METHODOLOGY

Network

In a MIS scenario, relations among surgical interaction triplets are complex, including coexistence, conversion, rejection, etc. Given an endoscopic image, there is often more than one surgical interaction triplet. Although instruments or targets with obvious features on the image are well recognized, it is difficult to recognize a surgical interaction triplet directly through image texture. For example, there is usually more than one verb corresponding to an instrument, and some instruments would perform the same verb. The verb is a comprehensive indirect judgment from whether the instrument is idle, instrument type, the type of objects grasped by the instrument tip, and the tissue state. When an instrument is located at the edge of the image, the performance of recognizing it can be improved indirectly through verb and target recognition. Shows that Figure 1



SIRNet architecture

The proposed SIRNet (Surgical Instrument Relation Network) architecture, designed for processing endoscopic images in Minimally Invasive Surgery (MIS), consists of five key components:

Feature Extractor: Utilizes the Mlp_mixer model to extract features from endoscopic images. It generates learnable embeddings for each surgical interaction triplet, allowing for the capture of distinctive image features.

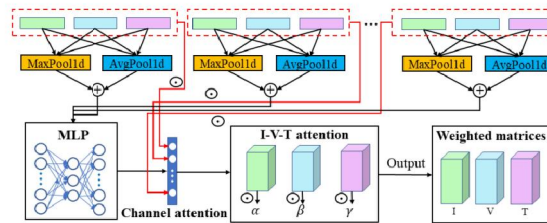
Triplet Relation Encoder: This crucial part of the network uses a stack of transformer layers for self-attention and cross-attention computations. It focuses on learning the complex relationships among surgical interaction triplets without relying on pre-defined dependencies. The encoder also establishes connections between the entire endoscopic image features and each surgical interaction triplet, moving beyond the conventional approach of using region-specific instrument features.

I-V-T Decoder: The network's decoder is specifically designed for the fine-grained Surgical Instrument Relation (SIR) task. It treats each surgical interaction triplet as an object comprising three separate components: the instrument, the verb (action), and the target. This design ensures that each component of the triplet is accurately classified. The decoder separately predicts preliminary instruments, verbs, and targets, mitigating the risk of misclassification.

Weight Attention Module: A specialized module designed to focus on important features and relationships, improving the network's ability to accurately interpret surgical scenes.

Valid Triplet Decoder: This component further processes the output from the I-V-T decoder, ensuring the generation of valid and coherent surgical interaction triplets.

SIRNet addresses the issue of surgical instrument overlapping in endoscopic images, a common challenge in MIS. It does not consider the overlapping of instruments as a significant problem, assuming that blocked instruments are typically not in active use during overlapping occurrences. Surgeons usually resolve overlapping by adjusting the instruments or the endoscope's angle, a factor considered in SIRNet's design. The architecture takes an endoscopic image and the number of surgical interaction triplets as input, efficiently extracting and processing features to understand and predict the dynamics of surgical interactions.



The weight attention module

Weight Attention Module:

The weight attention module is used to measure the importance of each query and the influence of instrument, verb, and target prediction results on the final surgical interaction triplet prediction. As presented in Fig. 2, the weight attention module consists of channel attention and I-V-T attention. Since different surgical interaction triplets appear differently in surgical scenes, different surgical interaction triplet queries have different weights. The channel attention is used to learn the weights during the training process. We aggregate the spatial information of the feature maps generated by the instrument detection FC layer, verb detection FC layer, and target detection FC layer by max-pooling and average-pooling. The max-pooling and average-pooling features are added in each query, and a spatial context descriptor is generated. Multilayer perceptron (MLP) uses this descriptor, comprising one hidden layer with three nodes. Finally, the weighted instrument, verb, and target with channel attention are computed by performing element-wise multiplication of the output of MLP with the original feature maps.

Valid Triplet Decoder

The last module of the proposed SIRNet is the valid triplet decoder. This module is used to generate the final valid surgical interaction triplets. The weighted feature matrices generated by the weight attention module are concatenated and then redirected to an average-pooling layer. Finally, a FC layer with a sigmoid function is used to generate the final probability of each surgical interaction triplet.

Loss

Two relevant loss functions are adopted when training the SIRNet: a bipartite matching loss and a weighted binary cross-entropy loss. The bipartite matching loss measures the difference between the predicted preliminary instrument set, verb set, target set in the I-V-T decoder and the ground truth triplets. The weighted binary cross-entropy loss computes the difference between the final predicted valid triplets and the ground truths.

Expected outcome

The expected outcome for the project is the development of a fine-tuned deep learning model that will significantly revolutionize robot-assisted Minimally Invasive Surgery (MIS). This model is designed to excel in two critical areas: surgical phase recognition and tool detection. By adeptly analyzing surgical video data, the model will accurately identify and classify various phases of robotic surgeries, enhancing the precision and efficiency of these complex procedures. In addition, the model is expected to demonstrate exceptional capability in detecting and categorizing a wide range of surgical tools utilized during operations. This dual functionality will not only streamline the surgical process but also bolster safety and accuracy, ultimately contributing to improved surgical outcomes and patient care. The successful implementation of this model will mark a substantial advancement in the integration of deep learning technologies in the field of robotic surgery, setting new standards for operational excellence and innovation.

REFERENCES

- [1]. K. M. Tahlil et al., "Crowdsourcing to support training for public health: A scoping review," PLOS Glob. Public Health, vol. 3, no. 7, p. e0002202, Jul. 2023.PLOS
- [2]. Z. Xue et al., "Toward Equitable Access: Leveraging Crowdsourced Reviews to Investigate Public Perceptions of Health Resource Accessibility," arXiv preprint arXiv:2502.10641, Feb. 2025.



- [3]. K. Kazari, Y. Chen, and Z. Shakeri, "Scaling Public Health Text Annotation: Zero-Shot Learning vs. Crowdsourcing for Improved Efficiency and Labeling Accuracy," arXiv preprint arXiv:2502.06150, Feb. 2025.
- [4]. J. Tucker's team, "Crowdsourcing transforms public health messaging," UNC Research, Mar. 2025. [Online]. Available: [https://www.unc.edu/posts/2025/03/12/crowdsourcing-transforms-public-health-messaging/UNC Chapel Hill](https://www.unc.edu/posts/2025/03/12/crowdsourcing-transforms-public-health-messaging/UNC%20Chapel%20Hill)
- [5]. Z. S. Hossein Abad et al., "Crowdsourcing for Machine Learning in Public Health Surveillance: Lessons Learned From Amazon Mechanical Turk," J. Med. Internet Res., vol. 24, no. 1, p. e28749, Jan. 2022. Journal of Medical Internet Research
- [6]. N. Ghildayal et al., "Public Health Surveillance in Electronic Health Records: Lessons From PCORnet," Prev. Chronic Dis., vol. 21, p. 230417, Jul. 2024. CDC
- [7]. H. Rilkoff et al., "Innovations in public health surveillance: An overview of novel use of data," Can. Commun. Dis. Rep., vol. 50, no. 3/4, Mar./Apr. 2024. Government of Canada
- [8]. A. E. Duhaime et al., "Gamified Crowdsourcing as a Novel Approach to Lung Ultrasound Data Set Labeling: Prospective Analysis," J. Med. Internet Res., vol. 26, no. 1, p. e51397, Jan. 2024. Journal of Medical Internet Research
- [9]. S. J. Schueller et al., "Crowdsourcing integrated into a digital mental health platform for depression: A pilot study," Internet Interv., vol. 29, p. 100567, Jan. 2024. ScienceDirect
- [10]. N. Divi et al., "Using EpiCore to Enable Rapid Verification of Potential Health Threats: Illustrated Use Cases and Summary Statistics," JMIR Public Health Surveill., vol. 10, p. e52093, Mar. 2024. IJMR+IJMIR Public Health+1
- [11]. M. M. Duggan et al., "Gamified Crowdsourcing as a Novel Approach to Lung Ultrasound Data Set Labeling: Prospective Analysis," J. Med. Internet Res., vol. 26, no. 1, p. e51397, Jan. 2024. Journal of Medical Internet Research
- [12]. C. C. Freifeld et al., "Participatory Epidemiology: Use of Mobile Phones for Community-Based Health Reporting," PLoS Med., vol. 7, no. 12, p. e1000376, Dec. 2010.
- [13]. E. H. Chan et al., "Global capacity for emerging infectious disease detection," Proc. Natl. Acad. Sci. U.S.A., vol. 107, no. 50, pp. 21701–21706, Dec. 2010.
- [14]. M. J. Paul and M. Dredze, "You Are What You Tweet: Analyzing Twitter for Public Health," in Proc. Fifth Int. AAAI Conf. Weblogs Soc. Media, 2011, pp. 265–272.
- [15]. A. Sadilek, H. Kautz, and V. Silenzio, "Modeling Spread of Disease from Social Interactions," in Proc. Sixth Int. AAAI Conf. Weblogs Soc. Media, 2012, pp. 322–329.
- [16]. D. R. Karger, S. Oh, and D. Shah, "Iterative Learning for Reliable Crowdsourcing Systems," in Adv. Neural Inf. Process. Syst., 2011, pp. 1953–1961.
- [17]. V. C. Raykar et al., "Learning From Crowds," J. Mach. Learn. Res., vol. 11, pp. 1297–1322, Apr. 2010.
- [18]. J. Whitehill et al., "Whose Vote Should Count More: Optimal Integration of Labels from Labelers of Unknown Expertise," in Adv. Neural Inf. Process. Syst., 2009, pp. 2035–2043.
- [19]. A. P. Dawid and A. M. Skene, "Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm," J. R. Stat. Soc. Ser. C Appl. Stat., vol. 28, no. 1, pp. 20–28, 1979.
- [20]. J. Snow et al., "Cheap and Fast—but is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks," in Proc. 2008 Conf. Empir. Methods Nat. Lang. Process., 2008, pp. 254–263.
- [21]. M. Lease, "Quality Control in Crowdsourcing," in Crowdsourcing for Search and Data Mining, ACM, 2011, pp. 97–102.
- [22]. A. Sorokin and D. Forsyth, "Utility Data Annotation with Amazon Mechanical Turk," in Proc. 2008 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, 2008, pp. 1–8.
- [23]. S. Nickoloff, "Capsule Commentary on Ranard et al., Crowdsourcing—Harnessing the Masses to Advance Health and Medicine, a Systematic Review," J. Gen. Intern. Med., vol. 29, no. 1, pp. 186–186, Jan. 2014.



- [24]. B. L. Ranard and R. M. Merchant, "Harnessing Digital Media to Promote Health and Research," *Circulation*, vol. 137, no. 21, pp. 2197–2199, May 2018.
- [25]. M. Swan, "Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem," *J. Med. Internet Res.*, vol. 14, no. 2, p. e46, Apr. 2012.
- [26]. J. A. Brownstein, C. C. Freifeld, and L. C. Madoff, "Digital Disease Detection — Harnessing the Web for Public Health Surveillance," *N. Engl. J. Med.*, vol. 360, no. 21, pp. 2153–2157, May 2009.
- [27]. C. C. Freifeld et al., "Participatory Epidemiology: Use of Mobile Phones for Community-Based Health Reporting," *PLoS Med.*, vol. 7, no. 12, p. e1000376, Dec. 2010.
- [28]. R. Chunara, J. R. Andrews, and J. S. Brownstein, "Social and News Media Enable Estimation of Epidemiological Patterns Early in the 2010 Haitian Cholera Outbreak," *Am. J. Trop. Med. Hyg.*, vol. 86, no. 1, pp. 39–45, Jan. 2012.
- [29]. M. J. Paul and M. Dredze, "You Are What You Tweet: Analyzing Twitter for Public Health," in *Proc. Fifth Int. AAAI Conf. Weblogs Soc. Media*, 2011, pp. 265–272.
- [30]. A. Sadilek, H. Kautz, and V. Silenzio, "Modeling Spread of Disease from Social Interactions," in *Proc. Sixth Int. AAAI Conf. Weblogs Soc. Media*, 2012, pp. 322–329.
- [31]. E. H. Chan et al., "Global capacity for emerging infectious disease detection," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, no. 50, pp. 21701–21706, Dec. 2010.
- [32]. A. Kittur, E. H. Chi, and B. Suh, "Crowdsourcing User Studies With Mechanical Turk," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2008, pp. 453–456.
- [33]. P. G. Ipeirotis, "Analyzing the Amazon Mechanical Turk Marketplace," *XRDS*, vol. 17, no. 2, pp. 16–21, Dec. 2010, doi: 10.1145/1869086.1869094.
- [34]. D. E. Chandler and A. Kapelner, "Breaking Monotony with Meaning: Motivation in Crowdsourcing Markets," *J. Econ. Behav. Organ.*, vol. 90, pp. 123–133, Jun. 2013, doi: 10.1016/j.jebo.2013.03.003.
- [35]. H. Becker, M. Naaman, and L. Gravano, "Beyond Trending Topics: Real-World Event Identification on Twitter," in *Proc. Fifth Int. AAAI Conf. Weblogs Soc. Media*, 2011, pp. 438–441.
- [36]. D. H. Nguyen, A. Patrick, and J. D. Rojas, "A Crowdsourcing Framework for Medical Image Annotation Using Amazon Mechanical Turk," *J. Digit. Imaging*, vol. 33, no. 4, pp. 1040–1052, Aug. 2020, doi: 10.1007/s10278-020-00348-7.
- [37]. A. Irshad et al., "Crowdsourcing Image Annotation for Nucleus Detection and Segmentation in Computational Pathology: Evaluating Experts, Automated Methods, and the Crowd," *Pacific Symp. Biocomput.*, pp. 294–305, 2014, doi: 10.1142/9789814583001_0029.
- [38]. M. Dumontier et al., "Crowdsourcing Analysis of Physical Activity Patterns to Identify Population Health Trends," *BMC Public Health*, vol. 21, p. 555, Mar. 2021, doi: 10.1186/s12889-021-10526-4.
- [39]. C. MacPherson, J. Ghosh, and N. Mohanty, "Ethical Challenges of Crowdsourced Health Research," *Hastings Cent. Rep.*, vol. 52, no. 3, pp. 21–29, May 2022, doi: 10.1002/hast.1375.
- [40]. Lee YJ, Arida JA, Donovan HS. The application of crowdsourcing approaches to cancer research: a systematic review. *Cancer Med.* 2017;6: 2595–605