Investigating Data Collection and Reporting Practices of Human Annotations in Societally Impactful Machine Learning Applications

A Systematic Review of Top-Cited IEEE Access Papers

Ahmed Ibrahim

Supervisor(s): Cynthia Liem, Andrew Demetriou, Frank Broz

EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 25, 2023

Name of the student: Ahmed Ibrahim
Final project course: CSE3000 Research Project
Thesis committee: Cynthia Liem, Andrew Demetriou, Frank Broz

Abstract

This systematic review investigates the practices and implications of human annotations in machine learning (ML) research. Analyzing a selection of 100 papers from the IEEE Access Journal, the study explores the data collection and reporting methods employed. The findings reveal a prevalent lack of standardization and formalization in the annotation process. Key details such as annotation sources, number of annotators, and formal instructions are frequently neglected, possibly compromising the quality and effectiveness of ML algorithms. Domain-specific implications are discussed, highlighting the need for comprehensive annotation practices in areas like medical diagnostics, language processing, and intelligent vehicle systems. The study contributes to the field by emphasizing the importance of standardized procedures and transparency in ML research. Future research is recommended to develop systematic annotation methodologies and examine the impact of subpar annotation on data quality.

1 Introduction

Machine Learning (ML) algorithms have recently seen an increase in development and deployment across a wide range of domains with significant societal impact, such as healthcare, criminology, surveillance, and fraud detection [1] [2]. To perform different classification tasks, these ML algorithms employ data from diverse sources such as images, videos, sound recordings, and text documents. The success or failure of these algorithms is evaluated by comparing their output to a reference, also referred to as the 'ground truth.' Ground truth is usually obtained through human annotators who label and classify the data. Such annotations are employed during the training and evaluation phases of machine learning algorithms. The efficacy of these algorithms is thus contingent on the quality of these annotations. Inaccurate annotations can potentially affect system performance and hinder accurate assessment of their true performance.

Despite the critical role of ground truth in these ML algorithms, there is a noticeable absence of systematic reviews and best practices within the literature regarding the process of collecting annotations from humans [3]. This may lead to a compromise in the quality and usefulness of ground truth data in societally impactful applications. An illustrative example of the catastrophic societal impact that can result from ML algorithm failures is the 'Childcare Benefits Scandal.' The Dutch tax office implemented an algorithmic system to identify suspicious tax-related activities. The system automatically flagged individuals with tax debts over ten thousand euros as high-risk fraud cases [4]. This approach led to false fraud accusations affecting around 26,000 families and consequently impacted approximately 71,000 children [5], out of whom 1,115 were separated from their parents as a result of high debts [6]. Although it is not explicitly clear whether the 'Childcare Benefits Scandal' was a direct result of poor data annotation, it underlines the potential for catastrophic societal effects when errors occur in the operation of ML algorithms. Inaccurate annotation could result in comparable scandals when used for ML algorithms in societally impactful domains.

Previous studies like that by [3] have provided some insights into this area, but there remains uncertainty surrounding human annotation practices. The focus of that study was directed towards a specific area, namely, ML application papers in the field of social computing that use Twitter data. The primary objective was to closely examine whether these papers followed and documented well-established standards for data labeling. This study revealed considerable variability in following these practices and highlighting potential risks associated with unreliable data. Whilst providing crucial insight, their work invites further exploration as the link between poorly reported annotation practices and subpar ML system performance is not explicitly addressed.

Building upon the insights provided by [3], it is crucial to further explore this link between poor-quality data and ML system performance. Understanding a data set before employing it for ML applications is essential, and any negligence in this regard could result in inaccurate classifications [7]. Gupta et al. [8] have also noted that understanding the data is not the only vital aspect, but also the quality of the data and its annotation, which can directly influence the efficiency and precision of ML algorithms. An inadequately documented annotation process can result in variability and inconsistencies when assigning labels to data, which in turn compromises the quality of the data. This could potentially result in suboptimal training of ML algorithms, negatively affecting their performance and reliability. It is therefore crucial to investigate the human annotation practices across diverse ML applications beyond the field of social computing and Twitter data, and discuss the impact incorrect ground truth could have on different domains.

To answer this identified gap, this research project aimed to provide a systematic review, examining the practices employed in the collection and reporting of human annotations across different ML applications. The study has specifically analyzed papers published in the IEEE Access Journal[1], which is known for its broad coverage and highly cited articles across multiple domains. The primary objective of this study revolved around answering the following research question: **"What are the data collection and reporting practices of human annotations/labels in societally impactful applications of Machine Learning Research as reflected in top-cited papers from the IEEE Access Journal?"**. This study seeks to uncover human annotation practices across different domains, trace data set origins, and examine methods employed by top-cited papers, as these papers are the most likely to have had a societal impact in recent years. It thereby aims to provide a better understanding of

[1] https://ieeeaccess.ieee.org/

the reliability and implications of human annotations in ML applications.

2 Methodology

This study's objective was to systematically review the prevailing practices of human annotations in ML algorithms. The adopted methodology consisted of three steps: selection of papers, data extraction, and data analysis. This methodology was inspired by Birhane et al.'s [9] similar study, which focused on values within ML research. Their approach of analyzing highly cited papers served as a valuable reference for this research. PRISMA guidelines [10] were also applied to ensure a systematic and transparent review. The following subsections describe the steps in detail.

2.1 Selection of papers

The first step of the research involved the selection of top-cited papers within the IEEE Access journal. The selected papers had to meet some requirements as inclusion criteria. The papers should have been published within the past three years to ensure that the review covers current discussions and practices in the field of ML. ML algorithms have seen rapid advancements, and this could also hold for the annotation practices within the field. Another requirement was that the papers should be written in English to exclude any possible errors resulting from incorrect translations.

The selection process relied on the use of Scopus[2] as a database. This choice is explained by the reproducibility of the search results, as Scopus displays consistent records, unlike search engines like Google Scholar, which may yield variable results based on the user [11]. Additionally, Scopus's strong search capabilities allowed for the incorporation of a targeted search string which helped apply the inclusion and exclusion criteria with little effort.

The specific search string utilized for this study is as follows: *"TITLE-ABS-KEY("machine learning" OR "deep learning" OR "neural network" OR "supervised learning") AND TITLE-ABS-KEY("annotation" OR "label" OR "ground truth" OR "class" OR "categorization") AND (LIMIT-TO (EXACTSRCTITLE "IEEE Access")) AND (LIMIT-TO (PUBYEAR, 2023) OR LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020))"*. This search string was designed to exclude papers published prior to the year 2020 and those not published in the IEEE Access journal. Furthermore, it ensured that the selected papers contained at least one of the following terms in their title, abstract, or keywords: "Machine Learning," "Deep Learning," "Neural Network," or "Supervised Learning," along with at least one of the following terms: "Annotation," "Label," "Ground Truth," "Class," or "Categorization." This increased the likelihood of including papers that involved ML techniques with human annotations. It is important to note that the search terms were not case-sensitive.

[2]https://www.scopus.com

The employed search criteria yielded a set of 2002 papers, which were then sorted according to the number of citations. From this set, the top 100 most-cited papers were selected for the systematic review. The citation count of these papers ranged between 32-234. It is important to highlight that the selection of this set is subject to change and reflects the data as captured on May 2, 2023.

2.2 Data Extraction

After concluding the paper selection process, a comprehensive review was conducted on the selected papers. This review involved addressing fifteen questions related to the human annotation practices drawn from and inspired by [3]. A more detailed description of these questions will be given in the next section. Each paper was examined individually and the answers to the questions were carefully recorded in an Excel spreadsheet, with each row representing one paper. The use of Excel not only provided an organized data set but also facilitated the way for subsequent data analysis.

Apart from the review of the papers, an extensive examination was also carried out on each data source mentioned in the papers. This was done to gain a better understanding of the nature and extent of human annotation used and to gather insights into how top-cited papers employ and incorporate external human annotations in their research. The data sources were tracked down following the references in each paper and were located using Google Scholar[3] and Google Search[4]. In many cases, the corresponding data source was found. For each of these data sources, the same questions were answered and reported in a second worksheet within the same Excel file.

2.3 Data Analysis

The final phase of this research was the visualization and interpretation of the collected data. This was done to identify the dominant patterns and trends within the selected papers regarding the data collection and reporting of the ground truth. Possible implications that these patterns and trends can have on the different domains (such as Medical Diagnostics, Security, etc.) were then examined. It is crucial to understand the potential effects and how these practices can shape the future use of ML algorithms in societally impactful domains. Questions such as - Are ML algorithms described in the selected papers reliable? Is it safe to employ this data in algorithms that might have a considerable societal impact?- were then answered.

3 Geiger et al.-Inspired Questions for the Systematic Review

This section provides an overview of the key questions that were addressed for each paper in the course of the systematic review. Drawing largely from the study conducted by [3], these questions have been slightly modified in formulation to fit this research. Each question presented in this

[3]https://scholar.google.nl/

[4]https://google.nl/

overview came with precise instructions explaining how it was answered. The aim was to ensure consistency throughout the review process.

3.1 Original Classification Task

The first question addressed whether the selected paper represented an 'original classification task.' This term was defined similarly to [3], where any paper that involved training a new classifier using data was classified as an 'original classification task.' A crucial aspect to consider was that the classifier had to be trained specifically within the scope of the paper and not pre-trained in other research. The relevance of this question stems from the expectation that when training a new classifier, the data (including the ground truth) used should be explicitly mentioned and described in detail, as it is essential for assessing the effectiveness of the trained classifier.

3.2 Utilization of Human Annotations

The second question aimed to answer whether a paper utilized data that was labeled by humans. A paper was considered to employ human annotations if it involved either original human annotations (new annotations gathered specifically for the current research) or external annotations (annotations previously collected in other papers or data sets). Although one might assume that this question is straightforward to answer, it often required extensive reading and referencing to trace the origins of the annotations used. Uncertainty arose when there was a lack of explanation about data set creation, as seen in [12]. Additional complexities appeared in situations like medical cases, where data labels were represented by official diagnostics by doctors, as referenced in [13], or when participants were instructed to deliberately exhibit a specific emotional state, as in [14]. To resolve this, a clearer definition was adopted: any instance where a human was asked to label existing data was regarded as human annotation. Therefore, medical diagnostic scenarios were included, while cases like the deliberate emotional state recordings were not.

3.3 Utilization of Original/External Human Annotations

As mentioned in the previous section, a paper is considered to utilize original human annotation, if within the scope of the research, new annotations are gathered. External human annotations are defined as labels gathered in previous research or external data sets. Some papers like [15] use both original as external human annotations.

3.4 Human Annotation Source

Human annotation can be gathered in different ways. During the systematic review, the following options were encountered: Experts, Paper's Authors, Crowdwork Platforms, Public and Data Uploaders. In certain cases, it was not specified who performed the data annotation. Experts were defined as individuals with substantial knowledge in the domain of the data. This category encompassed professionals such as radiologists, professional annotators, and professors. Crowdwork platforms such as Amazon MTurk[5] and MeMoSa Annotate[6] were also identified during the review. These platforms enable researchers to gather human annotations from users within the respective platforms. Other forms of human annotations involved public events or online engagement with (a group from) the general public to annotate data. Additionally, some papers in the review utilize data that is uploaded on online platforms where annotations were requested alongside the uploaded data. These annotators are defined as 'Data Uploaders.' An example of such a paper can be found in [16].

3.5 Number of Human Annotators

The subsequent question aimed to determine whether the number of annotators was specified in papers that utilized human annotations (original or external). In certain cases, papers did not explicitly mention the number of annotators but instead provided information such as the number of annotators per data point or the number of annotators for a specific set of data points. If, with this information, the number of annotators could be computed, the question was also affirmatively addressed.

3.6 Prediction of the Number of Human Annotators

The sixth question in the systematic review examined whether the authors of each paper had anticipated the number of human annotators required prior to the annotation process. The question was answered affirmatively only if the paper explicitly mentioned that the number of annotators was predicted prior to commencing the annotation process.

3.7 Formal Definitions and Instructions Provided

The following question sought to determine if the authors had explicitly defined annotation labels and if they had offered nuanced instructions to the annotators during the labeling process. Explicit definitions and guidelines are crucial to ensure consistency and reliability of the human annotations, thereby enhancing the quality of the ground truth.

3.8 Training for Human Annotators Provided

Training is an important step as it equips annotators with the necessary skills and knowledge to label the data accurately and consistently. For this question, a paper was marked a 'Yes' only if it specifically mentioned that the annotators were provided with test examples for annotation based on the given instructions and definitions before they annotated actual data. These training sessions often allowed annotators to highlight any potential areas of confusion or ambiguity, which could be clarified before the main annotation process began.

3.9 Pre-screening for Crowdwork Platforms

The ninth question of the systematic review considered the use of pre-screening on crowdwork platforms such as Amazon Mechanical Turk. Pre-screening is a process that allows the researchers to filter out potential annotators based on their expertise, past performance, or other relevant criteria. Such practice ensures that only qualified annotators contribute to the annotation task, thus increasing the quality of the data.

[5] https://www.mturk.com/

[6] https://memosa.my/

This question was answered affirmatively if a crowdwork platform was employed for annotation and a pre-screening phase was undertaken that resulted in the exclusion of certain annotators.

3.10 Multiple Annotator Overlap

The next question determined the existence of multiple annotator overlaps within the data. Employing multiple annotators for the same task could significantly reduce the likelihood of incorrect labeling, thereby improving the overall quality of annotated data. This approach also entails an increase in resources required for the annotation process. The answer to this question was answered affirmatively if multiple annotators were deployed to label the same items. Domains such as medical image analysis often necessitate the use of multiple annotators due to the high stakes involved - for instance, detecting cancerous or unhealthy tissue.

3.11 Reporting of Inter-Annotator Agreement or Another Metric

With the deployment of multiple annotators, there may be scenarios where the same data item receives varying labels. In such instances, there is a need for a mechanism to resolve this ambiguity, such as an expert intervention or an agreed consensus. This question addressed whether any type of inter-annotator agreement or another metric was reported. If no mention of an inter-annotator agreement was made, despite the involvement of multiple annotators for each data item, the response was marked as 'No.'

3.12 Link to the Data Set Provided

The next question aimed to assess whether a link to the utilized data sets (original and external) was provided in the paper. Affirmation for this question was given only if all data sets were appropriately referenced, while a negative response was marked if none of the data sets were mentioned. In instances where some but not all data sets were referenced, the answer was answered with 'Not All.' It is important to note that a link to the data set does not necessarily mean a direct link to the data - a reference to the original paper from which the data was sourced is considered valid as well.

3.13 Paper/Data Set Accessible

The final question sought to determine the accessibility of the paper or data set. This is rather straightforward and was answered affirmatively if the paper or data set could be accessed through the internet using the link provided or referred to in the paper. In some cases, the papers were behind a paywall, yet, as long as they were accessible through a valid link, the response was recorded as 'Yes.'

4 Results

This section presents the results of the systematic review of the 100 papers shown in Table 1. The table categorizes the papers and their corresponding data sources into different domains. To allow specific conclusions only about papers published in the IEEE Access journal, the results obtained from the selected papers are distinguished from those obtained from the utilized data sources. The data is publicly available and can be accessed through [17].

The data shown in Table 2 reveals that a majority of the papers (92 out of 100) introduced an original ML classification task, and over half of the associated data sources (58/150) were originally constructed/employed for an original classification task. This highlights the effectiveness of the search parameters that were employed to collect the papers from Scopus. As for the data sources that were linked in the papers, a significant portion comprised of papers introducing new public data sets, papers offering links to existing data sets, or papers that developed an original data set for their specific study.

Table 2: Original Classification Task

	Papers	Data sources
Yes	92	58
No	8	92

Table 3 presents the utilization of human annotations in the studied papers and their linked data sources. It reveals that 70 out of the 100 papers used human annotations, as did 99 of the associated data sources. On the other hand, human annotations were not used in 23 papers and 39 of the data sources. For 7 papers and 6 data sources, it was unclear whether human annotations were involved.

Table 3: Utilization of Human Annotations

	Papers	Data sources
Yes	70	99
No	23	39
Unclear	7	6

Table 3 explores the origin of human annotations. Out of the 70 papers that used human annotations, a total of 63 papers used external human annotations, while 15 relied on original human annotations. A total of 8 papers used both external and original human annotations. In contrast, when it comes to the data sources, 85 used original human annotations, while 21 used external annotations. Out of those, 7 used a combination of both original and external annotations.

Table 4: Utilization of Original/External Human Annotations

	Papers	Data sources
Original	15	85
External	63	21

Table 5 investigates the sources of human annotations reported in the papers and the data sets. Among the 70 papers that utilized annotations, 22 sourced their annotations from experts, 6 from the authors themselves, and 4 from public or online events. Only a single paper mentioned the use of a crowdwork platform. Notably, a significant portion of the

Table 1: Papers examined during the systematic review, categorized by domain and their respective utilized data sources

Domain	**Papers**	**Data Sources**
Agriculture	[18] [19] [20]	[21]
Computer Science	[22] [23] [24] [25] [26] [27]	[28] [29]
Computer Vision	[30] [31] [32] [33] [33] [34] [35] [36] [37] [38] [39] [40] [41] [42]	[43] [44] [45] [46] [47] [48] [49] [50] [51] [52] [53] [54] [55] [56] [57] [58] [59] [60]
Cybersecurity	[61] [62] [63] [64] [65] [66] [67] [68] [69] [70] [71] [72] [73]	[74] [75] [76] [77] [78] [79] [80] [81] [82] [83] [84] [85] [86]
Economics	[87] [88]	[89] [90] [91]
Education	[92] [93]	[94] [95] [96] [97] [98] [99] [100] [101] [102]
Electrical Engineering	[103] [104] [105] [106] [107]	None
Emotion Recognition	[108] [109] [110]	[111] [112] [113] [114] [115] [116]
Intel. Vehicle Systems	[117] [118] [119] [120]	[121] [46] [47] [122] [123] [124] [125]
Language Processing	[126] [127] [128] [129] [130] [131] [132] [120] [133] [134]	[135] [136] [137] [138] [139] [140] [141] [142] [143] [144] [145] [146] [46] [147] [148]
Mechanical Engineering	[149] [150] [151]	[152] [153] [154]
Medical Diagnostics	[155] [156] [157] [158] [159] [160] [161] [162] [163] [164] [165] [166] [167] [168] [169] [170] [171] [172] [173] [174] [175] [176] [177] [178] [179] [180] [181] [182] [183] [184] [185] [186] [187] [188] [189] [190]	[191] [192] [193] [194] [195] [196] [197] [198] [199] [200] [201] [16] [202] [203] [204] [205] [206] [207] [208] [209] [210] [211] [212] [213] [214] [215] [216] [217] [218] [219] [220] [221] [222] [223] [224] [225] [226] [227] [228] [229] [230] [231] [232] [233] [234] [235] [236] [237] [238] [239] [240] [241] [242] [243] [244] [245] [246] [247] [248] [249] [250] [251] [252] [253] [254] [255]

papers, 40 in total, did not disclose who provided the annotations. There were two papers that employed multiple types of annotators.

In comparison, the annotator types for the data sources were more often disclosed. Out of the 99 sources, 52 cited the use of expert annotators, while the authors and public or online events contributed to 12 and 9 data sets respectively. Crowdwork platforms were used for 6 data sources. Some data sources referenced the use of 'Data Uploaders', as defined in the previous section, for gathering annotations. A total of 20 data sources did not disclose the type of annotators.

Table 5: Human Annotation Source

	Papers	Data sources
Experts	22	52
Crowdwork	1	6
Paper's Authors	6	12
Data Uploaders	0	4
Public/Online Events	4	9
Not Mentioned	40	20

Table 6 illustrates whether the number of human annotators was mentioned. Only 12 of the papers mentioned the number of human annotators that were used. The majority, comprising 58 papers, did not disclose this information. When looking at the data sources, a similar pattern was found. Only 29 out of the 95 data sets provided the number of human annotators, while 66 data sources did not include this information.

Table 6: Number of Human Annotators Mentioned

	Papers	Data sources
Yes	12	29
No	58	66

Table 7 presents information about whether any details about the anticipated number of annotators prior to gathering annotations were provided. Remarkably, neither of the 70 reviewed papers nor the 95 data sources attempted to predict or estimate the number of human annotators required.

Table 7: Prediction of the Number of Human Annotators Needed

	Papers	Data sources
Yes	0	0
No	70	95

Table 8 provides data on whether the papers or data set sources reported formal definitions and instructions for their human annotators. Only a total of 7 papers and 26 data set sources provided explicit instructions and definitions for their annotation tasks.

Table 8: Formal Definitions and Instructions Provided

	Papers	Data sources
Yes	7	26
No	63	69

Table 9 highlights whether any training for human annotators was reported. A total of 2 papers and 10 data sources reported providing any form of training to annotators. The rest of the papers (68) and data sources (85) did not mention any type of training for the annotators.

Table 9: Training for Human Annotators Provided

	Papers	Data sources
Yes	2	10
No	68	85

Table 10 discusses pre-screening practices on crowdwork platforms. While the one research paper that employed a crowdwork platform did not specifically mention pre-screening, its employment of MeMoSa, which is only used by experts, served as an indirect form of pre-screening. Among the data sources, pre-screening was used in 3 instances, while in 3 others, no form of pre-screening was employed.

Table 10: Pre-screening for Crowdwork Platforms

	Papers	Data sources
Yes	1	3
No	0	3

Table 11 details the existence of multiple annotator overlap. Only 4 out of 70 papers had multiple annotators annotating the same data. Similarly, among the data sources, 21 out of 95 had annotator overlap. The majority of the papers did not (or did not report) the use of multiple annotators.

Table 11: Multiple Annotator Overlap

	Papers	Data sources
Yes	4	21
No	66	74

Table 12 examines whether the papers/data sources that mentioned the use of multiple annotators for the same item reported any metric of inter-annotator agreement. Of the 4 papers that mentioned the use of multiple annotators, 2 also reported the inter-annotator agreement. On the other hand, among the data sources, 9 out of the 21 mentioned a metric for the inter-annotator agreement.

Table 12: Reporting of Inter-Annotator Agreement or Another Metric

	Papers	Data sources
Yes	2	9
No	2	21

Table 13 demonstrates whether links to employed data were provided. Out of the papers using any type of data (92), the majority (68) provided a link to utilized data sources, 18 did not reference the data sources they employed, and 6 only referred to some of the data they used. In terms of data sources, we see that a total of 133 referenced all employed data, 4 did not reference the data sets they used, and one only referred to part of the data sets in use. The remaining data set sources (17) were either inaccessible or were public online databases.

Table 13: Links to data sets Provided

	Papers	Data sources
Yes	68	133
No	18	4
Not All	6	1

Table 14 addresses the accessibility of the reviewed papers and data set sources. All 100 papers that were reviewed were accessible. For the data sources, 141 out of 150 were accessible, 8 were not, and 1 was only partially accessible.

Table 14: Paper/Data set Accessible

	Papers	Data sources
Yes	100	141
No	0	8
Partly	0	1

5 Discussion

5.1 General Overview of the Results

The research examined a selection of 100 ML papers using a tailored search string, and it was found that the vast majority deployed ML algorithms for original classification tasks. From the pool of papers, 70% used human annotations, both original and external. This demonstrated the efficacy of the search criteria that were mentioned in Section 2. The papers were spread across 12 domains, showcasing the expansive diversity that IEEE Access encompasses. These papers were assigned to these domains manually, based on the keywords mentioned in the articles. As a result, the domains ranged from agriculture, computer science, and computer vision to cybersecurity, economics, and education, also spanning electrical engineering, emotion recognition, intelligent vehicle systems, language processing, mechanical engineering, and medical diagnostics. However, this classification was self-imposed and may contain minor errors due to the absence of a systematic methodology for classifying the papers into various domains, as this was out of the scope of this research.

A detailed exploration into the human annotations practices within the IEEE Access papers revealed that most papers (74%) linked to the original data source. A notable proportion of these papers (90%) also leveraged external annotations. Although it was feasible to track back to the original data sets, the review revealed that the papers did not delve deep into the quality of the annotations within these data sources. For instance, the human annotation source was not mentioned in 57% of the papers, while the number of annotators and formal instructions for them were left out in 83% and 90% of the papers respectively. This suggests that papers generally rely on the quality and the explanation provided by the original data source.

In addition, papers often failed to report important information concerning the annotation process. No papers reported performing any research prior to the annotation process to estimate the number of annotators required, indicating that researchers often do not focus on establishing data quality needs and formalization prior to the data collection and annotation process. The use of multiple annotators was not mentioned in 94% of the papers. Of those that did 5%, only half reported an inter-annotator agreement. But this count is too small to conclusively determine the practice of reporting inter-annotator agreement metrics in papers that employ multiple annotators.

This research also revealed a wide variety of data types across the cited data sources, reflecting the various domains covered by the IEEE Access papers. Data types included medical videos and images, audio and video tapes, images sourced from the internet, and numerical data sets. This variety of data types emphasized the need for specific and unique data in many papers. As a result, most of the data sources that employed human annotations opted for original annotations, as seen in 86% of the sources.

In comparison to the papers analyzed, data set sources demonstrated better human annotation practices. Only 20% of the data set sources did not mention the annotation sources, while 31% did mention the number of annotators, and formal instructions were provided in 27%. The use of multiple annotators was also slightly more prevalent in data set sources, and around half of these reported an inter-annotator metric. However, none of the data set sources predicted the number of annotations required before the annotation process, demonstrating a similar trend in the absence of predictive measures, as encountered in the analyzed papers.

Upon further analysis, it was observed that commonly used data sets like ImageNet, COCO, and Pascal VOC had better annotation practices, often featuring formal instructions and definitions, which were mentioned in only 25% out of all papers that used human annotations. Interestingly, COCO and Pascal VOC also provided training for their annotators, and all three data sets made use of Amazon MTurk to source their annotators.

This research highlights that, although human annotation practices tend to be superior in data sources, a general lack of standardization and best practices still prevails across all domains in reporting the collection of human annotations. While correct referencing and accessibility ensure transparency and reproducibility, more efforts need to be put towards formalizing annotation practices and planning the necessary resources to gather high-quality annotations prior to an annotation process starting.

5.2 Implications for Different Domains

The study highlights a prevalent lack of standard annotation procedures across various domains. This absence undermines the transparency and reproducibility of the reviewed papers. Ensuring the high-quality of data is important. However, most papers lack discussions concerning the impacts of poor-quality data or substandard annotation on research outcomes. In many cases, it was observed that the construction of ML algorithms was emphasized more than the quality of the data used for their training. This critical research aspect is often overlooked, although data quality is central to the performance of ML algorithms. This aspect should be given more attention, given the importance of data to the performance of ML algorithms.

The collection of papers was classified into 12 different domains, as noted before, and each domain could be variably affected by poor annotation practices. Some studies did not focus on a specific societal domain but conducted research within scientific fields, like Computer Science, Mechanical Engineering, Electrical Engineering, Cybersecurity, and Economics. These studies often made minimal use of human-annotated data sets or involved only simple human annotations. For instance, a cybersecurity data set might be classified into 'attack' and 'normal' modes based on an IP address. For these types of studies, extensive human annotation practices might not be a primary concern. Their main focus should be on comprehensive data collection practices, which were usually described in detail. Other domains might necessitate more robust annotation practices to ensure the production of high-quality ML algorithms.

In the domain of Intelligent Vehicle Systems, four papers were reviewed, three of which employed human annotations. These studies typically focused on ML algorithms for object detection or driver emotion recognition. Considering the increasing impact of intelligent vehicles on society [256], subpar annotation quality could translate into poor data, potentially leading to serious consequences. For example, studies on object detection often dismiss ambiguous instances, leaving ML classifiers untrained for such situations. If an intelligent vehicle's object-detection classifier fails to identify an object due to these ambiguities or poorly annotated data, accidents could occur. This makes high-quality data and comprehensive annotation crucial in this domain.

In the Language Processing domain, most papers analyzed textual content to predict sentiment based on language use. Formalization of the annotation practices was generally absent. Only one paper provided a detailed outline of the annotation process, including the number of annotators, their training, and inter-annotator agreement metrics. Despite the seeming insignificance of these papers to societal issues, their findings could impact applications like AI chatbots or online content moderation. Poor annotation practices might compromise sensitive content detection or sentiment analysis.

The Emotion Detection domain was limited to just three papers and focused on classifying video-recorded emotions. Two of these papers relied on human annotation and surpassed other papers by specifying the number of annotators and providing formal instructions. This consistency was mirrored in their data sources. Emotion detection, with potential applications in psychology [257], holds significant importance. However, flawed machine learning algorithms in this area could affect patients. For instance, improper emotion recognition could misdirect therapeutic strategies or misinterpret patients' needs.

The systematic review included several papers within computer vision, with a subset specifically focusing on agriculture. Many of these papers relied on large, frequently used data sets such as COCO [47], ImageNet [46], PascalVOC [48], and Places365 [147], either as standalone models or for pre-training. These data sets are annotated through Amazon MTurk, providing a relatively detailed overview of the annotation process. Smaller, paper-specific data sets often neglected to discuss the annotation process. While using a limited number of data sets within a vast domain might aid result comparability, it also risks introducing biases towards data specific to these selected data sets.

The systematic review frequently encountered papers within the medical diagnostics domain. These papers carry significant societal implications, as incorrect diagnoses can lead to severe patient outcomes. As the global population grows, the medical sector will increasingly rely on computational aids for diagnostics [1]. These papers, however, often present limitations. For example, expert annotators, although essential, are often few compared to the volume and significance of patient data under analysis. Also, papers often offer limited details about the annotation process besides indicating expert involvement. There is also a lack of reliable, extensive medical data, which leads to improvisation by researchers. Less reliable sources such as Radiopaedia[7], which is a public database with uncertified medical data from all over the world, is used in many papers. The lack of data verification and consistency within these sources could impact the reliability of machine learning applications in healthcare.

[7] https://radiopaedia.org/

6 Conclusion and Future Work

This systematic review investigated the data collection and reporting practices of human annotations in ML algorithms and its possible implications on societally impactful applications. To accomplish this, the research question that guided the study was formulated as follows:

"What are the data collection and reporting practices of human annotations/labels in societally impactful applications of Machine Learning Research as reflected in top-cited papers from the IEEE Access Journal?"

The analysis of the results obtained from the systematic review of papers from various domains revealed a prevalent lack of formalization in the annotation process. While the majority of papers and data sources used human annotations, the reporting of important annotation details was often insufficient. The lack of information on annotation sources, the number of annotators, and formal instructions undermine the transparency and reproducibility of the research findings. This could also compromise the quality of the data and may negatively affect the efficacy of the ML algorithms.

The discussion highlighted the implications of these findings across various domains. Within these domains, it was observed that different practices and challenges exist. For example, in the domain of Intelligent Vehicle Systems, comprehensive annotation practices are crucial to ensure the reliability of object detection algorithms and driver emotion recognition systems. In Language Processing, poor annotation practices can compromise the accuracy of sentiment analysis and content moderation applications. Emotion Detection research requires robust annotation practices to avoid misinterpretation of patients' emotions and misguided therapeutic strategies. In the Medical Diagnostics domain, the use of reliable and extensively verified medical data is essential to avoid erroneous diagnoses and ensure patient safety.

To address these concerns and improve the reporting and standardization of human annotation practices in machine learning research, the following recommendations for future work are proposed:

- Construction of a systematic and standardized methodology for the collection of human annotations.
- Conducting further research to investigate the actual impact of poor annotation on data quality and the subsequent performance of ML algorithms.
- Investigating the influence of data in pre-trained models on classifier performance and generalization to new data.

7 Responsible Research

Responsible research involves addressing potential ethical issues while ensuring that research methods are transparent and reproducible. In reflecting upon this research, several aspects related to ethical conduct are worth considering.

The first aspect is related to the sample of papers reviewed in this study. The collection of papers reviewed in this study does not represent the actual population published within the IEEE Access Journal, rather it provides a cross-sectional snapshot of papers that met specific search criteria. It's therefore essential to interpret the results of this study bearing in mind that they may not necessarily apply to all papers published in the journal or to all fields of machine learning research.

Another key consideration relates to the classification of the research papers into different domains. This task was performed manually without using any formal or systematic classification approach. This could potentially limit the reproducibility of this research as other researchers could classify papers differently, leading to different outcomes. This highlights the need for using a more systematic classification method to enhance reproducibility.

The study might also be impacted by the author's expertise and possible biases. As this study was conducted by a Computer Science student, the researcher might not have extensive expertise in all domains represented in the papers reviewed, including the psychology behind the annotation practices. This limitation could result in unintentional bias or errors in the interpretation of results.

Another issue related to this study was its execution by a single researcher, which can present both advantages and drawbacks. Lack of validation/verification increases the potential for errors during the review process and reporting of the results. However, the consistency in the approach towards the review process is likely to be enhanced since only one reviewer conducted the research. One solution to overcome this limitation could include adding secondary researcher to independently validate outcomes and report any inconsistencies or errors back to primary researcher.

References

[1] K. A. Bhavsar, "Medical diagnosis using machine learning: A statistical review," *Computers, Materials & Continua*, 2021.

[2] H. Anahideh, "Fair active learning," *Expert Syst. Appl.*, 2020.

[3] R. S. Geiger, K. Yu, Y. Yang, M. Dai, J. Qiu, R. Tang, and J. Huang, "Garbage in, garbage out?" *Association for Computing Machinery*, pp. 1–12, 2020.

[4] E. Pasman, *Kafka in de rechtsstaat. De toeslagenaffaire ontleed*. Prometheus, 2021.

[5] E. Van Dam, "Martine raakte samen met 26.000 andere ouders in de knel door toeslagenaffaire," *Omroep Gelderland*, December 2020. [Online]. Available: https://www.gld.nl/nieuws/6639460/martine-raakte-samen-met-26000-andere-ouders-in-de-knel-doortoeslagenaffaire

[6] K. Van der Schoot, "Toeslagenaffaire in nederland: 1.115 kinderen weggehaald bij ouders die onterecht als fraudeur werden bestempeld," *vrtnws.be*, October 2021. [Online]. Available: https://www.vrt.be/vrtnws/nl/2021/10/21/toeslagenaffaire-nederland-1-115-kinderen-weggehaald-bij-gedupe/

[7] A. Jain, H. Patel, L. Nagalapatti, N. Gupta, S. Mehta, S. Guttula, S. Mujumdar, S. Afzal, R. S. Mittal, and V. Munigala, "Overview and importance of data quality for machine learning tasks," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery &amp Data Mining*. ACM, Aug. 2020. [Online]. Available: https://doi.org/10.1145/3394486.3406477

[8] N. Gupta, S. Mujumdar, H. Patel, S. Masuda, N. Panwar, S. Bandyopadhyay, S. Mehta, S. Guttula, S. Afzal, R. S. Mittal, and V. Munigala, "Data quality for machine learning tasks," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery &amp Data Mining*. ACM, Aug. 2021. [Online]. Available: https://doi.org/10.1145/3447548.3470817

[9] A. Birhane, P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao, "The values encoded in machine learning research," *Association for Computing Machinery*, p. 173–184, 2022. [Online]. Available: https://doi.org/10.1145/3531146.3533083

[10] M. J. Page, D. Moher, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and J. E. McKenzie, "PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews," *BMJ*, p. n160, Mar. 2021. [Online]. Available: https://doi.org/10.1136/bmj.n160

[11] G. Pozsgai, G. L. Lövei, L. Vasseur, G. M. Gurr, P. Batáry, J. Korponai, N. A. Littlewood, J. Liu, A. Móra, J. J. Obrycki, O. L. Reynolds, J. A. Stockan, H. VanVolkenburg, J. Zhang, Y. Liu, and M. You, "Irreproducibility in searches of scientific literature: A comparative analysis," *Ecology and Evolution*, vol. 11, no. 21, pp. 14 658–14 668, 2021.

[12] H. Kang, D. H. Ahn, G. M. Lee, J. D. Yoo, K. H. Park, and H. K. Kim, "Iot network intrusion dataset," 2019. [Online]. Available: https://ieee-dataport.org/open-access/iot-network-intrusion-dataset

[13] K. Pogorelov, K. R. Randel, C. Griwodz, S. L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D.-T. Dang-Nguyen, M. Lux, P. T. Schmidt, M. Riegler, and P. Halvorsen, "KVASIR," in *Proceedings of the 8th ACM on Multimedia Systems Conference*. ACM, Jun. 2017. [Online]. Available: https://doi.org/10.1145/3083187.3083212

[14] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, "YawDD," in *Proceedings of the 5th*

ACM Multimedia Systems Conference. ACM, Mar. 2014. [Online]. Available: https://doi.org/10.1145/2557642.2563678

[15] Y. Yu, K. Zhang, H. Liu, L. Yang, and D. Zhang, "Real-time visual localization of the picking points for a ridge-planting strawberry harvesting robot," *IEEE Access*, vol. 8, pp. 116 556–116 568, 2020. [Online]. Available: https://doi.org/10.1109/access.2020.3003034

[16] X. Yang, X. He, J. Zhao, Y. Zhang, S. Zhang, and P. Xie, "Covid-ct-dataset: A ct scan dataset about covid-19," 2020.

[17] A. Ibrahim. (2023) Annotation practices. https://github.com/ahsmi47/AnnotationPractices_IEEEACCESS/. Data set.

[18] Q. Wu, Y. Chen, and J. Meng, "DCGAN-based data augmentation for tomato leaf disease identification," *IEEE Access*, vol. 8, pp. 98 716–98 728, 2020.

[19] Y. Yu, K. Zhang, H. Liu, L. Yang, and D. Zhang, "Real-time visual localization of the picking points for a ridge-planting strawberry harvesting robot," *IEEE Access*, vol. 8, pp. 116 556–116 568, 2020.

[20] J. Huixian, "The analysis of plants image recognition based on deep learning and artificial neural network," *IEEE Access*, vol. 8, pp. 68 828–68 841, 2020.

[21] A. Zbiciak and T. Markiewicz, "A new extraordinary means of appeal in the polish criminal procedure: the basic principles of a fair trial and a complaint against a cassatory judgment," *Access to Justice in Eastern Europe*, vol. 6, no. 2, pp. 1–18, Mar. 2023.

[22] N. M. Rezk, M. Purnaprajna, T. Nordstrom, and Z. Ul-Abdin, "Recurrent neural networks: An embedded computing perspective," *IEEE Access*, vol. 8, pp. 57 967–57 996, 2020.

[23] N. A. Asif, Y. Sarker, R. K. Chakrabortty, M. J. Ryan, M. H. Ahamed, D. K. Saha, F. R. Badal, S. K. Das, M. F. Ali, S. I. Moyeen, M. R. Islam, and Z. Tasneem, "Graph neural network: A comprehensive review on non-euclidean space," *IEEE Access*, vol. 9, pp. 60 588–60 606, 2021.

[24] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020.

[25] D. Chicco, M. J. Warrens, and G. Jurman, "The matthews correlation coefficient (MCC) is more informative than cohen's kappa and brier score in binary classification assessment," *IEEE Access*, vol. 9, pp. 78 368–78 381, 2021.

[26] S. S. Sarwar, A. Ankit, and K. Roy, "Incremental learning in deep convolutional neural networks using partial network sharing," *IEEE Access*, vol. 8, pp. 4615–4628, 2020.

[27] E.-S. M. El-Kenawy, S. Mirjalili, F. Alassery, Y.-D. Zhang, M. M. Eid, S. Y. El-Mashad, B. A. Aloyaydi, A. Ibrahim, and A. A. Abdelhamid, "Novel metaheuristic algorithm for feature selection, unconstrained functions and engineering problems," *IEEE Access*, vol. 10, pp. 40 536–40 555, 2022.

[28] Y. LeCun. (1998) Mnist handwritten digit database. [Online]. Available: http://yann.lecun.com/exdb/mnist

[29] A. Krizhevsky, "Learning multiple layers of features from tiny images," 2009.

[30] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120 234–120 254, 2020.

[31] L. Schmarje, M. Santarossa, S.-M. Schroder, and R. Koch, "A survey on semi-, self- and unsupervised learning for image classification," *IEEE Access*, vol. 9, pp. 82 146–82 168, 2021.

[32] L. Aziz, M. S. B. Haji Salam, U. U. Sheikh, and S. Ayub, "Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review," *IEEE Access*, vol. 8, pp. 170 461–170 495, 2020.

[33] W. Cai and Z. Wei, "PiiGAN: Generative adversarial networks for pluralistic image inpainting," *IEEE Access*, vol. 8, pp. 48 451–48 463, 2020.

[34] H. Alhichri, A. S. Alswayed, Y. Bazi, N. Ammour, and N. A. Alajlan, "Classification of remote sensing images using EfficientNet-B3 CNN model with attention," *IEEE Access*, vol. 9, pp. 14 078–14 094, 2021.

[35] G. Huo, Z. Wu, and J. Li, "Underwater object classification in sidescan sonar images using deep transfer learning and semisynthetic training data," *IEEE Access*, vol. 8, pp. 47 407–47 418, 2020.

[36] A. Abdollahi, B. Pradhan, and A. Alamri, "VNet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data," *IEEE Access*, vol. 8, pp. 179 424–179 436, 2020.

[37] A. Jalal, A. Ahmed, A. A. Rafique, and K. Kim, "Scene semantic recognition based on modified fuzzy c-mean and maximum entropy using object-to-object relations," *IEEE Access*, vol. 9, pp. 27 758–27 772, 2021.

[38] B. Cui, X. Chen, and Y. Lu, "Semantic segmentation of remote sensing images using transfer learning and deep convolutional neural network with dense connection," *IEEE Access*, vol. 8, pp. 116 744–116 755, 2020.

[39] X. Li, H. Su, and G. Liu, "Insulator defect recognition based on global detection and local segmentation," *IEEE Access*, vol. 8, pp. 59 934–59 946, 2020.

[40] I. Ahmed, M. Ahmad, F. A. Khan, and M. Asif, "Comparison of deep-learning-based segmentation models: Using top view person images," *IEEE Access*, vol. 8, pp. 136 361–136 373, 2020.

[41] H. Li, L. Deng, C. Yang, J. Liu, and Z. Gu, "Enhanced YOLO v3 tiny network for real-time ship detection

from visual image," *IEEE Access*, vol. 9, pp. 16 692–16 706, 2021.

[42] A. Abdollahi, B. Pradhan, S. Gite, and A. Alamri, "Building footprint extraction from high resolution aerial images using generative adversarial network (GAN) architecture," *IEEE Access*, vol. 8, pp. 209 517–209 527, 2020.

[43] G.-S. Xia, W. Yang, J. Delon, Y. Gousseau, H. Sun, and H. Maître, "Structural high-resolution satellite image indexing," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 38, 07 2010.

[44] V. Mnih, "Machine learning for aerial image labeling," Ph.D. dissertation, University of Toronto, CAN, 2013, aAINR96184.

[45] ISDNNSF. (2019) Remote sensing image scene classification. Accessed: insert the date you accessed this website. [Online]. Available: http://rscup.bjxintong.com.cn/

[46] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[47] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft coco: Common objects in context," 2014. [Online]. Available: https://arxiv.org/abs/1405.0312

[48] M. Everingham, S. Eslami, L. Van Gool *et al.*, "The pascal visual object classes challenge: A retrospective," *International Journal of Computer Vision*, vol. 111, pp. 98–136, 2015. [Online]. Available: https://doi.org/10.1007/s11263-014-0733-5

[49] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 6, pp. 1452–1464, 2018.

[50] G. Morales, G. Kemper, G. Sevillano, D. Arteaga, I. Ortega, and J. Telles, "Automatic segmentation of mauritia flexuosa in unmanned aerial vehicle (uav) imagery using deep learning," *Forests*, vol. 9, no. 12, 2018. [Online]. Available: https://www.mdpi.com/1999-4907/9/12/736

[51] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 270–279. [Online]. Available: https://doi.org/10.1145/1869790.1869829

[52] E. Othman, Y. Bazi, F. Melgani, H. Alhichri, N. Alajlan, and M. Zuair, "Domain adaptation network for cross-scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4441–4456, 2017.

[53] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 11, pp. 2321–2325, 2015.

[54] Q. Wang, S. Liu, J. Chanussot, and X. Li, "Scene classification with recurrent attention of vhr remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1155–1167, 2019.

[55] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.

[56] Y. Liu, J. Yao, X. Lu, M. Xia, X. Wang, and Y. Liu, "Roadnet: Learning to comprehensively analyze road networks in complex urban scenes from high-resolution remotely sensed images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2043–2056, 2019.

[57] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation," in *Computer Vision – ECCV 2006*, A. Leonardis, H. Bischof, and A. Pinz, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 1–15.

[58] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," in *2004 Conference on Computer Vision and Pattern Recognition Workshop*, 2004, pp. 178–178.

[59] Z. Shao, W. Wu, Z. Wang, W. Du, and C. Li, "Seaships: A large-scale precisely annotated dataset for ship detection," *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2593–2604, 2018.

[60] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, "SUN database: Large-scale scene recognition from abbey to zoo," *IEEE*, Jun. 2010. [Online]. Available: https://doi.org/10.1109/cvpr.2010.5539970

[61] A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood, and A. Anwar, "TON_IoT telemetry dataset: A new generation dataset of IoT and IIoT for data-driven intrusion detection systems," *IEEE Access*, vol. 8, pp. 165 130–165 150, 2020.

[62] S. Zavrak and M. Iskefiyeli, "Anomaly-based intrusion detection from network flow features using variational autoencoder," *IEEE Access*, vol. 8, pp. 108 346–108 358, 2020.

[63] G. Karatas, O. Demir, and O. K. Sahingoz, "Increasing the performance of machine learning-based IDSs on an imbalanced and up-to-date dataset," *IEEE Access*, vol. 8, pp. 32 150–32 162, 2020.

[64] S. Manimurugan, S. Al-Mutairi, M. M. Aborokbah, N. Chilamkurti, S. Ganesan, and R. Patan, "Effective attack detection in internet of medical things smart environment using a deep belief neural network," *IEEE Access*, vol. 8, pp. 77 396–77 404, 2020.

[65] M. Wang, K. Zheng, Y. Yang, and X. Wang, "An explainable machine learning framework for intrusion detection systems," *IEEE Access*, vol. 8, pp. 73 127–73 141, 2020.

[66] I. Ullah and Q. H. Mahmoud, "Design and development of a deep learning-based model for anomaly detection in IoT networks," *IEEE Access*, vol. 9, pp. 103 906–103 926, 2021.

[67] M. Bagaa, T. Taleb, J. B. Bernabe, and A. Skarmeta, "A machine learning security framework for iot systems," *IEEE Access*, vol. 8, pp. 114 066–114 077, 2020.

[68] L. Liu, P. Wang, J. Lin, and L. Liu, "Intrusion detection of imbalanced network traffic based on machine learning and deep learning," *IEEE Access*, vol. 9, pp. 7550–7563, 2021.

[69] G. Andresini, A. Appice, N. D. Mauro, C. Loglisci, and D. Malerba, "Multi-channel deep feature learning for intrusion detection," *IEEE Access*, vol. 8, pp. 53 346–53 359, 2020.

[70] A. Samy, H. Yu, and H. Zhang, "Fog-based attack detection framework for internet of things using deep learning," *IEEE Access*, vol. 8, pp. 74 571–74 585, 2020.

[71] L. Vu, Q. U. Nguyen, D. N. Nguyen, D. T. Hoang, and E. Dutkiewicz, "Deep transfer learning for IoT attack detection," *IEEE Access*, vol. 8, pp. 107 335–107 344, 2020.

[72] A. R. Gad, A. A. Nashat, and T. M. Barkat, "Intrusion detection system using machine learning for vehicular ad hoc networks based on ToN-IoT dataset," *IEEE Access*, vol. 9, pp. 142 206–142 217, 2021.

[73] F. Khan, C. Ncube, L. K. Ramasamy, S. Kadry, and Y. Nam, "A digital DNA sequencing engine for ransomware detection using machine learning," *IEEE Access*, vol. 8, pp. 119 710–119 719, 2020.

[74] M. Skelton, "Ransomware json dataset," 2017. [Online]. Available: https://github.com/codingo/Ransomware-Json-Dataset/blob/master/README.md

[75] CIC. (2018) Cse-cic-ids2018 — canadian institute for cybersecurity — unb. Accessed: insert the date you accessed this website. [Online]. Available: https://www.unb.ca/cic/datasets/ids-2018.html

[76] KDD. (n.d.) Kdd cup databases. Accessed: insert the date you accessed this website. [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup

[77] I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proceedings of the 4th International Conference on Information Systems Security and Privacy*. SCITEPRESS - Science and Technology Publications, 2018.

[78] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*. IEEE, Jul. 2009.

[79] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, "Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-IoT dataset," *Future Gener. Comput. Syst.*, vol. 100, pp. 779–796, Nov. 2019.

[80] H. Kang, D. H. Ahn, G. M. Lee, J. Do Yoo, K. H. Park, and H. K. Kim, "IoT network intrusion dataset," 2019.

[81] H. Hindy, E. Bayne, M. Bures, R. Atkinson, C. Tachtatzis, and X. Bellekens, "Machine learning based IoT intrusion detection system: An MQTT case study (MQTT-IoT-IDS2020 dataset)," 2020.

[82] S. Garcia, A. Parmisano, and M. J. Erquiaga, "IoT-23: A labeled dataset with malicious and benign IoT network traffic," 2020.

[83] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *2015 Military Communications and Information Systems Conference (MilCIS)*. IEEE, Nov. 2015.

[84] A. Verma and V. Ranga, "Evaluation of network intrusion detection systems for RPL based 6LoWPAN networks in IoT," *Wirel. Pers. Commun.*, vol. 108, no. 3, pp. 1571–1594, Oct. 2019.

[85] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breitenbacher, and Y. Elovici, "N-BaIoT—Network-Based detection of IoT botnet attacks using deep autoencoders," *IEEE Pervasive Comput.*, vol. 17, no. 3, pp. 12–22, Jul. 2018.

[86] N. Moustafa, "ToN_IoT datasets," 2019. [Online]. Available: https://research.unsw.edu.au/projects/toniot-datasets

[87] A. H. Bukhari, M. A. Z. Raja, M. Sulaiman, S. Islam, M. Shoaib, and P. Kumam, "Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting," *IEEE Access*, vol. 8, pp. 71 326–71 338, 2020.

[88] T. M. Alam, K. Shaukat, I. A. Hameed, S. Luo, M. U. Sarwar, S. Shabbir, J. Li, and M. Khushi, "An investigation of credit card default prediction in the imbalanced datasets," *IEEE Access*, vol. 8, pp. 201 173–201 198, 2020.

[89] I.-C. Yeh and C.-H. Lien, "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2473–2480, Mar. 2009.

[90] U. Grömping. (2019, 11) South German Credit Data: Correcting a Widely Used Data Set. [Online]. Available: http://www1.beuth-hochschule.de/FB_II/reports/Report-2019-004.pdf

[91] A. Dal Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4915–4928, Aug. 2014.

[92] R. Ghorbani and R. Ghousi, "Comparing different resampling methods in predicting students' performance using machine learning techniques," *IEEE Access*, vol. 8, pp. 67 899–67 911, 2020.

[93] Z. Kastrati, A. S. Imran, and A. Kurti, "Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs," *IEEE Access*, vol. 8, pp. 106 799–106 810, 2020.

[94] P. Cortez and A. Silva, "Using data mining to predict secondary school student performance," *EUROSIS*, 01 2008. [Online]. Available: http://www3.dsi.uminho.pt/pcortez/student.pdf

[95] G. Elia, G. Solazzo, G. Lorenzo, and G. Passiante, "Assessing learners' satisfaction in collaborative online courses through a big data approach," *Computers in Human Behavior*, vol. 92, pp. 589–599, Mar. 2019. [Online]. Available: https://doi.org/10.1016/j.chb.2018.04.033

[96] Fernandez, "Google-news-200d - dataset by jaredfern," 2017. [Online]. Available: https://data.world/jaredfern/google-news-200d

[97] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, and A. Joulin, "Advances in pre-training distributed word representations," *arXiv (Cornell University)*, Dec 2017. [Online]. Available: http://arxiv.org/abs/1712.09405

[98] Wikipedia, "English wikipedia." [Online]. Available: https://dumps.wikimedia.org/enwiki/latest/

[99] Koehn, "Statistical and neural machine translation." [Online]. Available: https://statmt.org/

[100] L. Han, "Umbc webbase corpus," 2013. [Online]. Available: https://ebiquity.umbc.edu/resource/html/id/351/UMBC-webbase-corpus

[101] Graff, David and Cieri, Christopher, "English gigaword," 2003. [Online]. Available: https://catalog.ldc.upenn.edu/LDC2003T05

[102] S. Nagel, "Common crawl," Jun 2017. [Online]. Available: https://commoncrawl.org/2017/06/

[103] M. A. Albreem, A. H. A. Habbash, A. M. Abu-Hudrouss, and S. S. Ikki, "Overview of precoding techniques for massive MIMO," *IEEE Access*, vol. 9, pp. 60 764–60 801, 2021.

[104] S. H. Rafi, Nahid-Al-Masood, S. R. Deeba, and E. Hossain, "A short-term load forecasting method using integrated CNN and LSTM network," *IEEE Access*, vol. 9, pp. 32 436–32 448, 2021.

[105] Z. Qu, C. Hou, C. Hou, and W. Wang, "Radar signal intra-pulse modulation recognition based on convolutional neural network and deep q-learning network," *IEEE Access*, vol. 8, pp. 49 125–49 136, 2020.

[106] M. Zhang, J. Li, Y. Li, and R. Xu, "Deep learning for short-term voltage stability assessment of power systems," *IEEE Access*, vol. 9, pp. 29 711–29 718, 2021.

[107] K. Nisar, Z. Sabir, M. A. Zahoor Raja, A. A. Ag. Ibrahim, F. Erdogan, M. R. Haque, J. J. P. C. Rodrigues, and D. B. Rawat, "Design of morlet wavelet neural network for solving a class of singular pantograph nonlinear differential models," *IEEE Access*, vol. 9, pp. 77 845–77 862, 2021.

[108] A. A. Abdelhamid, E.-S. M. El-Kenawy, B. Alotaibi, G. M. Amer, M. Y. Abdelkader, A. Ibrahim, and M. M. Eid, "Robust speech emotion recognition using CNN+LSTM based on stochastic fractal search optimization algorithm," *IEEE Access*, vol. 10, pp. 49 265–49 284, 2022.

[109] Y. Cimtay, E. Ekmekcioglu, and S. Caglar-Ozhan, "Cross-subject multimodal emotion recognition based on hybrid fusion," *IEEE Access*, vol. 8, pp. 168 865–168 878, 2020.

[110] M. Khateeb, S. M. Anwar, and M. Alnowami, "Multi-Domain feature fusion for emotion classification using DEAP dataset," *IEEE Access*, vol. 9, pp. 12 134–12 142, 2021.

[111] S. R. Livingstone and F. A. Russo, "The ryerson Audio-Visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in north american english," *PLoS One*, vol. 13, no. 5, p. e0196391, May 2018.

[112] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, "A database of german emotional speech," in *Interspeech 2005*. ISCA: ISCA, Sep. 2005.

[113] P. Jackson and S. ul haq. (2011, 04) Surrey audio-visual expressed emotion (savee) database. [Online]. Available: http://kahlan.eps.surrey.ac.uk/savee/

[114] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "IEMOCAP: interactive emotional dyadic motion capture database," *Lang. Resour. Eval.*, vol. 42, no. 4, pp. 335–359, Dec. 2008.

[115] E. Ekmekcioglu and Y. Cimtay, "Loughborough university multimodal emotion dataset-2," Mar. 2021.

[116] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis ;using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012.

[117] B. K. Savas and Y. Becerikli, "Real time driver fatigue detection system based on multi-task ConNN," *IEEE Access*, vol. 8, pp. 12 491–12 498, 2020.

[118] Q. Zhang, X. Chang, and S. B. Bian, "Vehicle-damage-detection segmentation algorithm based on improved mask RCNN," *IEEE Access*, vol. 8, pp. 6997–7004, 2020.

[119] C. Huang, X. Wang, J. Cao, S. Wang, and Y. Zhang, "HCF: A hybrid CNN framework for behavior detection of distracted drivers," *IEEE Access*, vol. 8, pp. 109 335–109 349, 2020.

[120] M. Gjoreski, M. Z. Gams, M. Lustrek, P. Genc, J.-U. Garbas, and T. Hassan, "Machine learning and end-to-end deep learning for monitoring driver distractions from physiological and visual signals," *IEEE Access*, vol. 8, pp. 70 590–70 603, 2020.

[121] M. Kostinger, P. Wohlhart, P. M. Roth, and H. Bischof, "Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization," in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. IEEE, Nov. 2011. [Online]. Available: https://doi.org/10.1109/iccvw.2011.6130513

[122] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, "YawDD," in *Proceedings of the 5th ACM Multimedia Systems Conference*. New York, NY, USA: ACM, Mar. 2014.

[123] C.-H. Weng, Y.-H. Lai, and S.-H. Lai, "Driver drowsiness detection via a hierarchical temporal deep belief network," in *Computer Vision – ACCV 2016 Workshops*, ser. Lecture notes in computer science. Cham: Springer International Publishing, 2017, pp. 117–133.

[124] "State farm distracted driver detection," 2014. [Online]. Available: https://www.kaggle.com/c/state-farm-distracted-driver-detection

[125] I. Pavlidis, M. Dcosta, S. Taamneh, M. Manser, T. Ferris, R. Wunderlich, E. Akleman, and P. Tsiamyrtzis, "Dissecting driver behaviors under cognitive, emotional, sensorimotor, and mixed stressors," *Sci. Rep.*, vol. 6, no. 1, May 2016.

[126] M. K. Elhadad, K. F. Li, and F. Gebali, "Detecting misleading information on COVID-19," *IEEE Access*, vol. 8, pp. 165 201–165 215, Sep. 2020.

[127] M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, M. A. Bencherif, and M. A. Mekhtiche, "Hand gesture recognition for sign language using 3DCNN," *IEEE Access*, vol. 8, pp. 79 491–79 509, 2020.

[128] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed, A. Mehmood, and M. T. Sadiq, "Document-level text classification using single-layer multisize filters convolutional neural network," *IEEE Access*, vol. 8, pp. 42 689–42 707, 2020.

[129] F. Alzamzami, M. Hoda, and A. El Saddik, "Light gradient boosting machine for general sentiment classification on short texts: A comparative evaluation," *IEEE Access*, vol. 8, pp. 101 840–101 858, 2020.

[130] A. Yousaf, M. Umer, S. Sadiq, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, "Emotion recognition by textual tweets classification using voting classifier (LR-SGD)," *IEEE Access*, vol. 9, pp. 6286–6295, 2021.

[131] A. Abdul Aziz and A. Starkey, "Predicting supervise machine learning performances for sentiment analysis using contextual-based approaches," *IEEE Access*, vol. 8, pp. 17 722–17 733, 2020.

[132] H. A. Almuzaini and A. M. Azmi, "Impact of stemming and word embedding on deep learning-based arabic text categorization," *IEEE Access*, vol. 8, pp. 127 913–127 928, 2020.

[133] T. M. Ghanim, M. I. Khalil, and H. M. Abbas, "Comparative study on deep convolution neural networks DCNN-based offline arabic handwriting recognition," *IEEE Access*, vol. 8, pp. 95 465–95 482, 2020.

[134] S. Y. Arafat and M. J. Iqbal, "Urdu-text detection and recognition in natural scene images using deep learning," *IEEE Access*, vol. 8, pp. 96 787–96 803, 2020.

[135] C. for Language Engineering. (2010) Valid ligatures of urdu. [Online]. Available: https://www.cle.org.pk/software/ling_resources/UrduLigatures.htm

[136] S. P. Agency. Spa. [Online]. Available: http://www.spa.gov.sa/

[137] Google. Fact check tools. [Online]. Available: https://toolbox.google.com/factcheck/about#fce-included

[138] T. Zia, M. P. Akhter, and Q. Abbas, "Comparative study of feature selection approaches for urdu text categorization," *Malaysian Journal of Computer Science*, vol. 28, 2015.

[139] CS2R, Center of Smart Robotics Research CS2R, "Arabic sign language," 2022.

[140] T. Shanableh, K. Assaleh, and M. Al-Rousan, "Spatio-temporal feature-extraction techniques for isolated gesture recognition in arabic sign language," *IEEE Trans. Syst. Man Cybern. B Cybern.*, vol. 37, no. 3, pp. 641–650, Jun. 2007.

[141] A. M. Martinez, R. B. Wilbur, R. Shay, and A. C. Kak, "Purdue RVL-SLLL ASL database for automatic recognition of american sign language," in *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*. IEEE Comput. Soc, 2003.

[142] M. Sharjeel, R. M. A. Nawab, and P. Rayson, "COUNTER: corpus of urdu news text reuse," *Lang. Resour. Eval.*, vol. 51, no. 3, pp. 777–803, Sep. 2017.

[143] R. Abaalkhail, F. Alzamzami, S. Aloufi, R. Alharthi, and A. E. Saddik, "Affectional ontology and multimedia dataset for sentiment analysis," in *Lecture Notes in Computer Science*, ser. Lecture notes in computer science. Cham: Springer International Publishing, 2018, pp. 15–28.

[144] B. Ghaddar and J. Naoum-Sawaya, "High dimensional data classification and feature selection using support

vector machines," *Eur. J. Oper. Res.*, vol. 265, no. 3, pp. 993–1004, Mar. 2018.

[145] A. Chouigui, O. B. Khiroun, and B. Elayeb, "ANT corpus: An arabic news text collection for textual classification," in *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*. IEEE, Oct. 2017.

[146] M. Pechwitz, H. El Abed, and V. Märgner, "Handwritten arabic word recognition using the IFN/ENIT-database," in *Guide to OCR for Arabic Scripts*. London: Springer London, 2012, pp. 169–213.

[147] B. Zhou, A. Lapedriza, A. Torralba, and A. Oliva, "Places: An image database for deep scene understanding," *J. Vis.*, vol. 17, no. 10, p. 296, Aug. 2017.

[148] Stanford University, "WebBase," http://dbpubs.stanford.edu:8091/~testbed/doc2/WebBase/, n.d.

[149] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics—a comprehensive review," *IEEE Access*, vol. 8, pp. 29 857–29 881, 2020.

[150] H.-Y. Chen and C.-H. Lee, "Vibration signals analysis by explainable artificial intelligence (XAI) approach: Application on bearing faults diagnosis," *IEEE Access*, vol. 8, pp. 134 246–134 256, 2020.

[151] G. Hua, L. Zhu, J. Wu, C. Shen, L. Zhou, and Q. Lin, "Blockchain-based federated learning for intelligent control in heavy haul railway," *IEEE Access*, vol. 8, pp. 176 830–176 839, 2020.

[152] C. W. R. University. (2021, Aug) Bearing data center. [Online]. Available: https://engineering.case.edu/bearingdatacenter/download-data-file

[153] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification," *PHME CONF*, vol. 3, no. 1, Jul. 2016.

[154] NASA. Intelligent systems division. [Online]. Available: https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/

[155] M. Goyal, A. Oakley, P. Bansal, D. Dancey, and M. H. Yap, "Skin lesion segmentation in dermoscopic images with ensemble deep learning methods," *IEEE Access*, vol. 8, pp. 4171–4181, 2020.

[156] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, M. Imran, and M. Shoaib, "A deep learning model based on concatenation approach for the diagnosis of brain tumor," *IEEE Access*, vol. 8, pp. 55 135–55 144, 2020.

[157] E.-S. M. El-Kenawy, A. Ibrahim, S. Mirjalili, M. M. Eid, and S. E. Hussein, "Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images," *IEEE Access*, vol. 8, pp. 179 317–179 335, Sep. 2020.

[158] A. Ishaq, S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques," *IEEE Access*, vol. 9, pp. 39 707–39 716, 2021.

[159] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning," *IEEE Access*, vol. 8, pp. 114 822–114 832, 2020.

[160] C. Iwendi, S. Khan, J. H. Anajemba, A. K. Bashir, and F. Noor, "Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model," *IEEE Access*, vol. 8, pp. 28 462–28 474, 2020.

[161] A. M. Shaker, M. Tantawi, H. A. Shedeed, and M. F. Tolba, "Generalization of convolutional neural networks for ECG classification using generative adversarial networks," *IEEE Access*, vol. 8, pp. 35 592–35 605, 2020.

[162] R. A. Welikala, P. Remagnino, J. H. Lim, C. S. Chan, S. Rajendran, T. G. Kallarakkal, R. B. Zain, R. D. Jayasinghe, J. Rimal, A. R. Kerr, R. Amtha, K. Patil, W. M. Tilakaratne, J. Gibson, S. C. Cheong, and S. A. Barman, "Automated detection and classification of oral lesions using deep learning for early detection of oral cancer," *IEEE Access*, vol. 8, pp. 132 677–132 693, 2020.

[163] M. A. Khan, M. S. Sarfraz, M. Alhaisoni, A. A. Albesher, S. Wang, and I. Ashraf, "StomachNet: Optimal deep learning features fusion for stomach abnormalities classification," *IEEE Access*, vol. 8, pp. 197 969–197 981, 2020.

[164] L. Wei, K. Ding, and H. Hu, "Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network," *IEEE Access*, vol. 8, pp. 99 633–99 647, 2020.

[165] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "DeTrac: Transfer learning of class decomposed medical images in convolutional neural networks," *IEEE Access*, vol. 8, pp. 74 901–74 913, 2020.

[166] N. M. Aboelenein, P. Songhao, A. Koubaa, A. Noor, and A. Afifi, "HTTU-net: Hybrid two track u-net for automatic brain tumor segmentation," *IEEE Access*, vol. 8, pp. 101 406–101 415, 2020.

[167] D. A. Zebari, D. Q. Zeebaree, A. M. Abdulazeez, H. Haron, and H. N. A. Hamed, "Improved threshold based and trainable fully automated segmentation for breast cancer boundary and pectoral muscle in mammogram images," *IEEE Access*, vol. 8, pp. 203 097–203 116, 2020.

[168] N. E. M. Khalifa, M. H. N. Taha, D. Ezzat Ali, A. Slowik, and A. E. Hassanien, "Artificial intelligence technique for gene expression by tumor RNA-seq data: A novel optimized deep learning approach," *IEEE Access*, vol. 8, pp. 22 874–22 883, 2020.

[169] H. Kaushik, D. Singh, M. Kaur, H. Alshazly, A. Zaguia, and H. Hamam, "Diabetic retinopathy diagnosis from fundus images using stacked generalization of deep models," *IEEE Access*, vol. 9, pp. 108 276–108 292, 2021.

[170] S. B. Shuvo, S. N. Ali, S. I. Swapnil, M. S. Al-Rakhami, and A. Gumaei, "CardioXNet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings," *IEEE Access*, vol. 9, pp. 36 955–36 967, 2021.

[171] S. Murugan, C. Venkatesan, M. G. Sumithra, X.-Z. Gao, B. Elakkiya, M. Akila, and S. Manoharan, "DEMNET: A deep learning model for early diagnosis of alzheimer diseases and dementia from MR images," *IEEE Access*, vol. 9, pp. 90 319–90 329, 2021.

[172] M. Khushi, K. Shaukat, T. M. Alam, I. A. Hameed, S. Uddin, S. Luo, X. Yang, and M. C. Reyes, "A comparative performance analysis of data resampling methods on imbalance medical data," *IEEE Access*, vol. 9, pp. 109 960–109 975, 2021.

[173] Z. Huang, X. Zhu, M. Ding, and X. Zhang, "Medical image classification using a light-weighted hybrid neural network based on PCANet and DenseNet," *IEEE Access*, vol. 8, pp. 24 697–24 712, 2020.

[174] G. Vrbancic and V. Podgorelec, "Transfer learning with adaptive fine-tuning," *IEEE Access*, vol. 8, pp. 196 197–196 211, 2020.

[175] J. Wang, L. Yang, Z. Huo, W. He, and J. Luo, "Multi-Label classification of fundus images with EfficientNet," *IEEE Access*, vol. 8, pp. 212 499–212 508, 2020.

[176] M. Ali, S. O. Gilani, A. Waris, K. Zafar, and M. Jamil, "Brain tumour image segmentation using deep networks," *IEEE Access*, vol. 8, pp. 153 589–153 598, 2020.

[177] D. Singh, V. Kumar, M. Kaur, M. Y. Jabarulla, and H.-N. Lee, "Screening of COVID-19 suspected subjects using multi-crossover genetic algorithm based dense convolutional neural network," *IEEE Access*, vol. 9, pp. 142 566–142 580, 2021.

[178] Y. Wang, W. Yue, X. Li, S. Liu, L. Guo, H. Xu, H. Zhang, and G. Yang, "Comparison study of radiomics and deep learning-based methods for thyroid nodules classification using ultrasound images," *IEEE Access*, vol. 8, pp. 52 010–52 017, 2020.

[179] Y. Yari, T. V. Nguyen, and H. T. Nguyen, "Deep learning applied for histological diagnosis of breast cancer," *IEEE Access*, vol. 8, pp. 162 432–162 448, 2020.

[180] X. Xu, S. Jeong, and J. Li, "Interpretation of electrocardiogram (ECG) rhythm by combined CNN and BiLSTM," *IEEE Access*, vol. 8, pp. 125 380–125 388, 2020.

[181] L. Ibrahim, M. Mesinovic, K.-W. Yang, and M. A. Eid, "Explainable prediction of acute myocardial infarction using machine learning and shapley values," *IEEE Access*, vol. 8, pp. 210 410–210 417, 2020.

[182] M. A. Awal, M. Masud, M. S. Hossain, A. A.-M. Bulbul, S. M. H. Mahmud, and A. K. Bairagi, "A novel bayesian optimization-based machine learning framework for COVID-19 detection from inpatient facility data," *IEEE Access*, vol. 9, pp. 10 263–10 281, Jan. 2021.

[183] R. Ashraf, M. A. Habib, M. Akram, M. A. Latif, M. S. A. Malik, M. Awais, S. H. Dar, T. Mahmood, M. Yasir, and Z. Abbas, "Deep convolution neural network for big data medical image classification," *IEEE Access*, vol. 8, pp. 105 659–105 670, 2020.

[184] J. J. Bird, J. Kobylarz, D. R. Faria, A. Ekart, and E. P. Ribeiro, "Cross-domain MLP and CNN transfer learning for biological signal processing: EEG and EMG," *IEEE Access*, vol. 8, pp. 54 789–54 801, 2020.

[185] K. D. McCay, E. S. L. Ho, H. P. H. Shum, G. Fehringer, C. Marcroft, and N. D. Embleton, "Abnormal infant movements classification with deep learning on pose-based features," *IEEE Access*, vol. 8, pp. 51 582–51 592, 2020.

[186] W. Mardini, M. M. Bani Yassein, R. Al-Rawashdeh, S. Aljawarneh, Y. Khamayseh, and O. Meqdadi, "Enhanced detection of epileptic seizure using EEG signals in combination with machine learning classifiers," *IEEE Access*, vol. 8, pp. 24 046–24 055, 2020.

[187] S. Mehmood, T. M. Ghazal, M. A. Khan, M. Zubair, M. T. Naseem, T. Faiz, and M. Ahmad, "Malignancy detection in lung and colon histopathology images using transfer learning with class selective image processing," *IEEE Access*, vol. 10, pp. 25 657–25 668, 2022.

[188] R. Man, P. Yang, and B. Xu, "Classification of breast cancer histopathological images using discriminative patches screened by generative adversarial networks," *IEEE Access*, vol. 8, pp. 155 362–155 377, 2020.

[189] S. Basheer, S. Bhatia, and S. B. Sakri, "Computational modeling of dementia prediction using deep neural network: Analysis on OASIS dataset," *IEEE Access*, vol. 9, pp. 42 449–42 462, 2021.

[190] Y. Wang, B. Lei, A. Elazab, E.-L. Tan, W. Wang, F. Huang, X. Gong, and T. Wang, "Breast cancer image classification via multi-network features and dual-network orthogonal low-rank learning," *IEEE Access*, vol. 8, pp. 27 779–27 792, 2020.

[191] MedRxiv, "Medrxiv." [Online]. Available: https://www.medrxiv.org/

[192] N. L. of Medicine, "Medpix." [Online]. Available: https://medpix.nlm.nih.gov/home

[193] ——, "Home." [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/

[194] RaedioPaedia, "Covid-19." [Online]. Available: https://radiopaedia.org/articles/covid-19-4?lang=us

[195] [Online]. Available: https://www.biorxiv.org/

[196] J. Suckling, J. Parker, D. Dance, and et al., "Mammographic image analysis society (mias) database v1.21," Dataset, 2015. [Online]. Available: https://www.repository.cam.ac.uk/handle/1810/250394

[197] M. A. Guevara Lopez, N. Posada, D. Moura, R. Pollán, J. Franco-Valiente, C. Ortega, M. Del Solar, G. Díaz-Herrero, I. Ramos, J. Loureiro, T. Fernandes, and B. Araújo, "Bcdr: A breast cancer digital repository," pp. 1065–1066, 01 2012.

[198] S. Hassan, "Mit automated auscultation system," 06 2005. [Online]. Available: https://dspace.mit.edu/handle/1721.1/18018

[199] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," 2017.

[200] T. Mendonca, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "PH^2 - a dermoscopic image database for research and benchmarking," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, Jul. 2013.

[201] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang, and Q. Feng, "Enhanced performance of brain tumor classification via tumor region augmentation and partition," *PLoS One*, vol. 10, no. 10, p. e0140381, Oct. 2015.

[202] S. G. Armato, 3rd, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, E. J. R. Van Beeke, D. Yankelevitz, A. M. Biancardi, P. H. Bland, M. S. Brown, R. M. Engelmann, G. E. Laderach, D. Max, R. C. Pais, D. P. Y. Qing, R. Y. Roberts, A. R. Smith, A. Starkey, P. Batrah, P. Caligiuri, A. Farooqi, G. W. Gladish, C. M. Jude, R. F. Munden, I. Petkovska, L. E. Quint, L. H. Schwartz, B. Sundaram, L. E. Dodd, C. Fenimore, D. Gur, N. Petrick, J. Freymann, J. Kirby, B. Hughes, A. V. Casteele, S. Gupte, M. Sallamm, M. D. Heath, M. H. Kuhn, E. Dharaiya, R. Burns, D. S. Fryd, M. Salganicoff, V. Anand, U. Shreter, S. Vastagh, and B. Y. Croft, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, Feb. 2011.

[203] T. Ahmad, A. Munir, S. H. Bhatti, M. Aftab, and M. A. Raza, "Survival analysis of heart failure patients: A case study," *PLoS One*, vol. 12, no. 7, p. e0181001, Jul. 2017.

[204] M. Combalia, N. C. F. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, C. Carrera, A. Barreiro, A. C. Halpern, S. Puig, and J. Malvehy, "BCN20000: Dermoscopic lesions in the wild," 2019.

[205] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. Data*, vol. 5, no. 1, p. 180161, Aug. 2018.

[206] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, May 2001.

[207] M. A. Khan, M. Rashid, M. Sharif, K. Javed, and T. Akram, "Classification of gastrointestinal diseases of stomach from WCE using improved saliency-based method and discriminant features selection," *Multimed. Tools Appl.*, vol. 78, no. 19, pp. 27 743–27 770, Oct. 2019.

[208] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, "Kvasir-SEG: A segmented polyp dataset," 2019.

[209] K. Pogorelov, K. R. Randel, C. Griwodz, S. L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D.-T. Dang-Nguyen, M. Lux, P. T. Schmidt, M. Riegler, and P. Halvorsen, "KVASIR," in *Proceedings of the 8th ACM on Multimedia Systems Conference*. New York, NY, USA: ACM, Jun. 2017.

[210] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Comput. Med. Imaging Graph.*, vol. 43, pp. 99–111, Jul. 2015.

[211] J. Silva, A. Histace, O. Romain, X. Dray, and B. Granado, "Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 9, no. 2, pp. 283–293, Mar. 2014.

[212] J. Bernal, J. Sánchez, and F. Vilariño, "Towards automatic polyp detection with a polyp appearance model," *Pattern Recognit.*, vol. 45, no. 9, pp. 3166–3182, Sep. 2012.

[213] D. Gutman, N. C. F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC)," 2016.

[214] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *AJR Am. J. Roentgenol.*, vol. 174, no. 1, pp. 71–74, Jan. 2000.

[215] J. N. Kather, C.-A. Weis, F. Bianconi, S. M. Melchers, L. R. Schad, T. Gaiser, A. Marx, and F. G. Zöllner, "Multi-class texture analysis in colorectal cancer histology," *Sci. Rep.*, vol. 6, no. 1, Sep. 2016.

[216] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M.-A. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, Ç. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H.-C. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[217] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: toward a full-field digital mammographic database," *Acad. Radiol.*, vol. 19, no. 2, pp. 236–248, Feb. 2012.

[218] S. K. Saha, B. Fernando, J. Cuadros, D. Xiao, and Y. Kanagasingam, "Automated quality assessment of colour fundus images for diabetic retinopathy screening in telemedicine," *J. Digit. Imaging*, vol. 31, no. 6, pp. 869–878, Dec. 2018.

[219] Yaseen, G.-Y. Son, and S. Kwon, "Classification of heart sound signal using multiple features," *Appl. Sci. (Basel)*, vol. 8, no. 12, p. 2344, Nov. 2018.

[220] C. Liu, D. Springer, Q. Li, B. Moody, R. A. Juan, F. J. Chorro, F. Castells, J. M. Roig, I. Silva, A. E. W. Johnson, Z. Syed, S. E. Schmidt, C. D. Papadaniil, L. Hadjileontiadis, H. Naseri, A. Moukadem, A. Dieterlen, C. Brandt, H. Tang, M. Samieinasab, M. R. Samieinasab, R. Sameni, R. G. Mark, and G. D. Clifford, "An open access database for the evaluation of heart sound algorithms," *Physiol. Meas.*, vol. 37, no. 12, pp. 2181–2213, Dec. 2016.

[221] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of heart sound recordings by a duration-dependent hidden markov model," *Physiol. Meas.*, vol. 31, no. 4, pp. 513–529, Apr. 2010.

[222] C. D. Papadaniil and L. J. Hadjileontiadis, "Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 4, pp. 1138–1152, Jul. 2014.

[223] H. Naseri and M. R. Homaeinezhad, "Detection and boundary identification of phonocardiogram sounds using an expert frequency-energy based metric," *Ann. Biomed. Eng.*, vol. 41, no. 2, pp. 279–292, Feb. 2013.

[224] A. Moukadem, A. Dieterlen, N. Hueber, and C. Brandt, "Localization of heart sounds based on s-transform and radial basis function neural network," in *IFMBE Proceedings*, ser. IFMBE proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 168–171.

[225] H. Tang, T. Li, Y. Park, and T. Qiu, "Separation of heart sound signal from noise in joint cycle frequency-time-frequency domains based on fuzzy detection," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 10, pp. 2438–2447, Oct. 2010.

[226] M. Samieinasab and R. Sameni, "Fetal phonocardiogram extraction using single channel blind source separation," in *2015 23rd Iranian Conference on Electrical Engineering*. IEEE, May 2015.

[227] Dubey, "Alzheimer's dataset (4 class of images)," Dec 2019. [Online]. Available: https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images

[228] U. of South Florida, "Ddsm: Digital database for screening mammography." [Online]. Available: http://www.eng.usf.edu/cvprg/mammography/database.html

[229] U. of Peking, "Odir-2019," 2019. [Online]. Available: https://odir2019.grand-challenge.org/dataset/

[230] Società Italiana di Radiologia Medica e Interventistica (SIRM), "COVID-19 Database," Available at: https://www.sirm.org/category/senza-categoria/covid-19/, 2020.

[231] J. C. Monteral, "COVID-Chestxray Database," Available at: https://github.com/ieee8023/covid-chestxray-dataset, 2020.

[232] T. R. Muhammad, E. H. Chowdhury, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. Al-Emadi, and M. B. I. Reaz, "COVID-19 CHEST X-RAY DATABASE," Available at: https://www.kaggle.com/tawsifurrahman/covid19-radiography-database, 2020.

[233] C. Imaging, "This is a thread of COVID-19 CXR (all SARS-CoV2 PCR+) from my hospital (Spain). I hope it could help," Available at: https://threadreaderapp.com/thread/1243928581983670272.html, 2020.

[234] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2097–2106.

[235] P. Mooney, "Chest X-Ray Images (Pneumonia)," Available at: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia, 2018.

[236] C. Health and B. Health, "Coronavirus disease 2019 (covid-19) clinical data repository," Accessed from https://covidclinicaldata.org/., 2020.

[237] B. T. Wyman, D. J. Harvey, K. Crawford, M. A. Bernstein, O. Carmichael, P. E. Cole, P. K. Crane, C. DeCarli, N. C. Fox, J. L. Gunter, D. Hill, R. J. Killiany, C. Pachai, A. J. Schwarz, N. Schuff, M. L. Senjem, J. Suhy, P. M. Thompson, M. Weiner, C. R. Jack, Jr, and Alzheimer's Disease Neuroimaging Initiative, "Standardization of analysis sets for reporting results from ADNI MRI data," *Alzheimers. Dement.*, vol. 9, no. 3, pp. 332–337, May 2013.

[238] C. S. Zhu, P. F. Pinsky, B. S. Kramer, P. C. Prorok, M. P. Purdue, C. D. Berg, and J. K. Gohagan, "The prostate, lung, colorectal, and ovarian cancer screening trial and its associated research resource," *J. Natl. Cancer Inst.*, vol. 105, no. 22, pp. 1684–1693, Nov. 2013.

[239] National Lung Screening Trial Research Team, "Data from the national lung screening trial (NLST)," 2013.

[240] P. Leavey, A. Sengupta, D. Rakheja, O. Daescu, H. B. Arunachalam, and R. Mishra, "Osteosarcoma UT southwestern/UT dallas for viable and necrotic tumor assessment," 2019.

[241] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki, M. Prastawa, E. Alberts, J. Lipkova, J. Freymann, J. Kirby, M. Bilello, H. Fathallah-Shaykh, R. Wiest, J. Kirschke, B. Wiestler, R. Colen, A. Kotrotsou, P. Lamontagne, D. Marcus, M. Milchenko, A. Nazeri, M.-A. Weber, A. Mahajan, U. Baid, E. Gerstner, D. Kwon, G. Acharya, M. Agarwal, M. Alam, A. Albiol, A. Albiol, F. J. Albiol, V. Alex, N. Allinson, P. H. A. Amorim, A. Amrutkar, G. Anand, S. Andermatt, T. Arbel, P. Arbelaez, A. Avery, M. Azmat, P. B., W. Bai, S. Banerjee, B. Barth, T. Batchelder, K. Batmanghelich, E. Battistella, A. Beers, M. Belyaev, M. Bendszus, E. Benson, J. Bernal, H. N. Bharath, G. Biros, S. Bisdas, J. Brown, M. Cabezas, S. Cao, J. M. Cardoso, E. N. Carver, A. Casamitjana, L. S. Castillo, M. Catà, P. Cattin, A. Cerigues, V. S. Chagas, S. Chandra, Y.-J. Chang, S. Chang, K. Chang, J. Chazalon, S. Chen, W. Chen, J. W. Chen, Z. Chen, K. Cheng, A. R. Choudhury, R. Chylla, A. Clérigues, S. Colleman, R. G. R. Colmeiro, M. Combalia, A. Costa, X. Cui, Z. Dai, L. Dai, L. A. Daza, E. Deutsch, C. Ding, C. Dong, S. Dong, W. Dudzik, Z. Eaton-Rosen, G. Egan, G. Escudero, T. Estienne, R. Everson, J. Fabrizio, Y. Fan, L. Fang, X. Feng, E. Ferrante, L. Fidon, M. Fischer, A. P. French, N. Fridman, H. Fu, D. Fuentes, Y. Gao, E. Gates, D. Gering, A. Gholami, W. Gierke, B. Glocker, M. Gong, S. González-Villá, T. Grosges, Y. Guan, S. Guo, S. Gupta, W.-S. Han, I. S. Han, K. Harmuth, H. He, A. Hernández-Sabaté, E. Herrmann, N. Himthani, W. Hsu, C. Hsu, X. Hu, X. Hu, Y. Hu, Y. Hu, R. Hua, T.-Y. Huang, W. Huang, S. Van Huffel, Q. Huo, V. Hv, K. M. Iftekharuddin, F. Isensee, M. Islam, A. S. Jackson, S. R. Jambawalikar, A. Jesson, W. Jian, P. Jin, V. J. M. Jose, A. Jungo, B. Kainz, K. Kamnitsas, P.-Y. Kao, A. Karnawat, T. Kellermeier, A. Kermi, K. Keutzer, M. T. Khadir, M. Khened, P. Kickingereder, G. Kim, N. King, H. Knapp, U. Knecht, L. Kohli, D. Kong, X. Kong, S. Koppers, A. Kori, G. Krishnamurthi, E. Krivov, P. Kumar, K. Kushibar, D. Lachinov, T. Lambrou, J. Lee, C. Lee, Y. Lee, M. Lee, S. Lefkovits, L. Lefkovits, J. Levitt, T. Li, H. Li, W. Li, H. Li, X. Li, Y. Li, H. Li, Z. Li, X. Li, Z. Li, X. Li, W. Li, Z.-S. Lin, F. Lin, P. Lio, C. Liu, B. Liu, X. Liu, M. Liu, J. Liu, L. Liu, X. Llado, M. M. Lopez, P. R. Lorenzo, Z. Lu, L. Luo, Z. Luo, J. Ma, K. Ma, T. Mackie, A. Madabushi, I. Mahmoudi, K. H. Maier-Hein, P. Maji, C. P. Mammen, A. Mang, B. S. Manjunath, M. Marcinkiewicz, S. McDonagh, S. McKenna, R. McKinley, M. Mehl, S. Mehta, R. Mehta, R. Meier, C. Meinel, D. Merhof, C. Meyer, R. Miller, S. Mitra, A. Moiyadi, D. Molina-Garcia, M. A. B. Monteiro, G. Mrukwa, A. Myronenko, J. Nalepa, T. Ngo, D. Nie, H. Ning, C. Niu, N. K. Nuechterlein, E. Oermann, A. Oliveira, D. D. C. Oliveira, A. Oliver, A. F. I. Osman, Y.-N. Ou, S. Ourselin, N. Paragios, M. S. Park, B. Paschke, J. G. Pauloski, K. Pawar, N. Pawlowski, L. Pei, S. Peng, S. M. Pereira, J. Perez-Beteta, V. M. Perez-Garcia, S. Pezold, B. Pham, A. Phophalia, G. Piella, G. N. Pillai, M. Piraud, M. Pisov, A. Popli, M. P. Pound, R. Pourreza, P. Prasanna, V. Prkovska, T. P. Pridmore, S. Puch, É. Puybareau, B. Qian, X. Qiao, M. Rajchl, S. Rane, M. Rebsamen, H. Ren, X. Ren, K. Revanuru, M. Rezaei, O. Rippel, L. C. Rivera, C. Robert, B. Rosen, D. Rueckert, M. Safwan, M. Salem, J. Salvi, I. Sanchez, I. Sánchez, H. M. Santos, E. Sartor, D. Schellingerhout, K. Scheufele, M. R. Scott, A. A. Scussel, S. Sedlar, J. P. Serrano-Rubio, N. J. Shah, N. Shah, M. Shaikh, B. U. Shankar, Z. Shboul, H. Shen, D. Shen, L. Shen, H. Shen, V. Shenoy, F. Shi, H. E. Shin, H. Shu, D. Sima, M. Sinclair, O. Smedby, J. M. Snyder, M. Soltaninejad, G. Song, M. Soni, J. Stawiaski, S. Subramanian, L. Sun, R. Sun, J. Sun, K. Sun, Y. Sun, G. Sun, S. Sun, Y. R. Suter, L. Szilagyi, S. Talbar, D. Tao, D. Tao, Z. Teng, S. Thakur, M. H. Thakur, S. Tharakan, P. Tiwari, G. Tochon, T. Tran, Y. M. Tsai, K.-L. Tseng, T. A. Tuan, V. Turlapov, N. Tustison, M. Vakalopoulou, S. Valverde, R. Vanguri, E. Vasiliev, J. Ventura, L. Vera, T. Vercauteren, C. A. Verrastro, L. Vidyaratne, V. Vilaplana, A. Vivekanandan, G. Wang, Q. Wang, C. J. Wang, W. Wang, D. Wang, R. Wang, Y. Wang, C. Wang, G. Wang, N. Wen, X. Wen, L. Weninger, W. Wick, S. Wu, Q. Wu, Y. Wu, Y. Xia, Y. Xu, X. Xu, P. Xu, T.-L. Yang, X. Yang, H.-Y. Yang, J. Yang, H. Yang, G. Yang, H. Yao, X. Ye, C. Yin, B. Young-Moxon, J. Yu, X. Yue, S. Zhang, A. Zhang, K. Zhang, X. Zhang, L. Zhang, X. Zhang, Y. Zhang, L. Zhang, J. Zhang, X. Zhang, T. Zhang, S. Zhao, Y. Zhao, X. Zhao, L. Zhao, Y. Zheng, L. Zhong, C. Zhou, X. Zhou, F. Zhou, H. Zhu, J. Zhu, Y. Zhuge, W. Zong, J. Kalpathy-Cramer, K. Farahani, C. Davatzikos, K. van Leemput, and B. Menze, "Identifying

the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," 2018.

[242] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. A. Emadi, M. B. I. Reaz, and M. T. Islam, "Can AI help in screening viral and COVID-19 pneumonia?" *IEEE Access*, vol. 8, pp. 132 665–132 676, 2020.

[243] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, Jul. 2016.

[244] Y.-G. Kim, D. Shin, M. Y. Park, S. Lee, M. S. Jeon, D. Yoon, and R. W. Park, "ECG-ViEW II, a freely accessible electrocardiogram database," *PLoS One*, vol. 12, no. 4, p. e0176222, Apr. 2017.

[245] J. J. Bird, L. J. Manso, E. P. Ribeiro, A. Ekart, and D. R. Faria, "A study on mental state classification using EEG-based brain-machine interface," in *2018 International Conference on Intelligent Systems (IS)*. IEEE, Sep. 2018.

[246] J. J. Bird, J. Kobylarz, D. R. Faria, A. Ekart, and E. P. Ribeiro, "Cross-domain MLP and CNN transfer learning for biological signal processing: EEG and EMG," *IEEE Access*, vol. 8, pp. 54 789–54 801, 2020.

[247] N. Hesse, C. Bodensteiner, M. Arens, U. G. Hofmann, R. Weinberger, and A. Sebastian Schroeder, "Computer vision for medical infant motion analysis: State of the art and RGB-D data set," in *Lecture Notes in Computer Science*, ser. Lecture notes in computer science. Cham: Springer International Publishing, 2019, pp. 32–49.

[248] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state," *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, vol. 64, no. 6 Pt 1, p. 061907, Dec. 2001.

[249] A. A. Borkowski, M. M. Bui, L. B. Thomas, C. P. Wilson, L. A. DeLand, and S. M. Mastorides, "Lung and colon cancer histopathological image dataset (LC25000)," 2019.

[250] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *J. Cogn. Neurosci.*, vol. 19, no. 9, pp. 1498–1507, Sep. 2007.

[251] G. Aresta, T. Araújo, S. Kwok, S. S. Chennamsetty, M. Safwan, V. Alex, B. Marami, M. Prastawa, M. Chan, M. Donovan, G. Fernandez, J. Zeineh, M. Kohl, C. Walz, F. Ludwig, S. Braunewell, M. Baust, Q. D. Vu, M. N. N. To, E. Kim, J. T. Kwak, S. Galal, V. Sanchez-Freire, N. Brancati, M. Frucci, D. Riccio, Y. Wang, L. Sun, K. Ma, J. Fang, I. Kone, L. Boulmane, A. Campilho, C. Eloy, A. Polónia, and P. Aguiar, "BACH: Grand challenge on breast cancer histology images," *Med. Image Anal.*, vol. 56, pp. 122–139, Aug. 2019.

[252] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)," 2019. [Online]. Available: https://arxiv.org/abs/1902.03368

[253] C. Ferles, "Cancer types: Rna sequencing values from tumor samples/tissues," 2018. [Online]. Available: https://data.mendeley.com/datasets/sf5n64hydt/1

[254] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Ng, "Reading digits in natural images with unsupervised feature learning," *NIPS*, 01 2011.

[255] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," in *2008 8th IEEE International Conference on Automatic Face &amp Gesture Recognition*. IEEE, Sep. 2008. [Online]. Available: https://doi.org/10.1109/afgr.2008.4813399

[256] E. Schoitsch, "Autonomous vehicles and automated driving status, perspectives and societal impact," *Information technology, society and economy strategic cross-influences (IDIMT2016)*, vol. 45, no. 1, pp. 405–424, 2016.

[257] D. Tomar, "An emotion detection system based on multi least squares twin support vector machine," *Adv. Artif. Intell.*, 2014.