StreamTag: A Platform for Flexible Tagged Data Management

David Díaz-Jiménez[1][a], Francisco Mata-Mata[1][b], José L. López[1][c], Luis G. Pérez-Cordón[1][d], José-María Serrano[1][e], Carmen Martínez-Cruz[2][f], Juana M. Morcillo-Martínez[3][g], Ángeles Verdejo-Espinosa[4][h], Juan C. Cuevas-Martínez[5][i], Raquel Viciana-Abad[5][j], Pedro J. Reche-López[5][k], José M. Pérez-Lorenzo[5][l], Juan F. Gaitán-Guerrero[1][m] and Macarena Espinilla[1][n]

[1] *Department of Computer Science, University of Jaén, 23071, Jaén, Spain*

[2] *Department of Languages and Computer Systems, University of Granada, 18071, Granada, Spain*

[3] *Psycology Department, Faculty of Social Work, University of Jaén, 23071 Jaén, Spain*

[4] *Electrical Engineering Department, University of Jaén, 23071 Jaén, Spain*

[5] *Telecommunication Engineering Department, University of Jaén, 23071 Jaén, Spain*

Keywords: Human Activity Recognition, Data Labeling, Healthcare Monitoring, Internet of Things, Artificial Intelligence, Personalized Care, Nursing Home Monitoring.

Abstract: This paper presents StreamTag, a platform designed for the efficient management of labeled data in healthcare environments, particularly for activity recognition systems in residential and nursing home settings. Human Activity Recognition (HAR) is crucial for monitoring patient behaviors and supporting personalized care, and this field has evolved significantly with advances in IoT and AI. StreamTag integrates a flexible data labeling structure and a modular architecture, enabling data collection, labeling, and secure management of activity data. The system leverages non-relational databases for scalable data handling, along with secure protocols to ensure data integrity and privacy. This work examines existing approaches in HAR, including data-driven, knowledge-based, and hybrid models, and situates StreamTag as a versatile solution that combines flexible user-controlled labeling with high adaptability for diverse healthcare contexts. Future directions are suggested for enhancing system functionality and integration with more advanced analytical tools.

1 INTRODUCTION

Healthcare stands as a pivotal area within artificial intelligence (AI), where the potential for innovation and impact is immense (Yu et al., 2018; Davenport and Kalakota, 2019; Alowais et al., 2023). As AI applications continue to emerge, they address a range of objectives in health—from patient monitoring and diagnostics (Kumar et al., 2023) to personalized treatment plans—transforming the landscape of healthcare practices (Johnson et al., 2021). However, the effectiveness of these AI applications heavily relies on the availability and quality of data, particularly those datasets where specific types of events are labeled. Accurate event labeling is necessary for training AI models to recognize patterns, predict outcomes, and ultimately support clinical decision-making (Miller and Brown, 2017; Duan et al., 2019).

In many healthcare environments, event labeling has traditionally been managed through manual processes, such as written logs or entries in physical notebooks. While these conventional methods have served

[a] https://orcid.org/0000-0003-1791-4258
[b] https://orcid.org/0000-0001-6099-0016
[c] https://orcid.org/0000-0003-2583-8638
[d] https://orcid.org/0000-0002-0753-6460
[e] https://orcid.org/0000-0001-5046-0724
[f] https://orcid.org/0000-0002-8117-0647
[g] https://orcid.org/0000-0002-5271-6145
[h] https://orcid.org/0000-0002-7998-553X
[i] https://orcid.org/0000-0003-3749-5986
[j] https://orcid.org/0000-0003-2545-7229
[k] https://orcid.org/0000-0002-5417-3551
[l] https://orcid.org/0000-0002-5286-8026
[m] https://orcid.org/0009-0007-6872-1401
[n] https://orcid.org/0000-0003-1118-7782

Díaz-Jiménez, D., Mata-Mata, F., López, J. L., Pérez-Cordón, L. G., Serrano, J.-M., Martínez-Cruz, C., Morcillo-Martínez, J. M., Verdejo-Espinosa, Á., Cuevas-Martínez, J. C., Viciana-Abad, R., Reche-López, P. J., Pérez-Lorenzo, J. M., Gaitán-Guerrero, J. F. and Espinilla, M.
StreamTag: A Platform for Flexible Tagged Data Management.
DOI: 10.5220/0013278600003938

In *Proceedings of the 11th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2025)*, pages 233-239
ISBN: 978-989-758-743-6; ISSN: 2184-4984

a purpose, they fall short in terms of efficiency, accuracy, and scalability, particularly when handling large volumes of data in real-time. For applications that require prompt and continuous data input, such as patient activity monitoring or health event tracking, manual systems often introduce delays and inconsistencies. These limitations underscore the need for tools capable of automating data labeling and centralizing information storage, ensuring instant access and seamless data flow within healthcare systems.

In response to these challenges, Stream Tag was developed as an mobile application designed to meet the specific data management needs of healthcare providers. This tool offers a solution for real-time event labeling, addressing the inefficiencies and inaccuracies of traditional data entry methods. Stream Tag's core functionality revolves around its ability to instantly label and transmit events to a centralized server, thereby enabling a real-time flow of critical information. This feature is particularly advantageous in dynamic healthcare settings, such as hospitals and nursing homes, where up-to-date information can make a significant difference in patient care and clinical decision-making.

The application provides a interface that allows healthcare staff to quickly label events such as patient activities, medication administration, or health incidents, eliminating the delays typically associated with manual data entry. It is designed to work with existing healthcare systems, providing a user-friendly experience that requires minimal training. Each event, once labeled, is securely transmitted to a dedicated server which makes it readily available to authorized personnel across the network.

In addition, the system's flexibility allows users to define a set of predefined specific labels and activities relevant to their context, creating a customized data environment that reflects the unique needs of each facility. In elderly care settings, for instance, the staff can track residents' daily activities or health events with ease, supporting continuous monitoring. By replacing manual records with a digital, centralized platform, Stream Tag reduces the administrative burden on healthcare providers, allowing them to focus more on direct patient care.

The structure of the paper is as follows: in Section 1, we present the general context, motivation, and objectives of the research on labeled data management in healthcare environments. Section 2 provides a perspective of current artificial intelligence use, labeling solutions and relevant technologies in the field of data labeling and management. Section 3 describes the architecture of the StreamTag platform, detailing its main components and their interactions for secure and efficient data handling. In Section 4, we explain the functionality of each view and the configuration options available to users. Finally, Section 5 presents the main contributions of StreamTag and outlines future challenges and potential improvements.

2 RELATED WORKS

In the field of healthcare and well-being, Data Acquisition and Management has become essential in order to monitor and understand user behavior patterns, especially in residential and home care contexts. Research in human activity recognition (HAR) (Jobanputra et al., 2019) has advanced significantly in recent years, enabled by technologies such as the Internet of Things (IoT) (Laghari et al., 2021; Mouha et al., 2021; Bhuiyan et al., 2021) and artificial intelligence (AI) (Haug and Drazen, 2023; Zhang and Lu, 2021). HAR focuses on identifying and classifying human activities using data collected from sensors, such as accelerometers and gyroscopes, found in portable or fixed devices (Ramanujam et al., 2021; Dang et al., 2020), while another approaches make use of RGB cameras, thermal cameras an so on (Ke et al., 2013; Shaikh and Chai, 2021; Dang et al., 2020). These type of systems allows movements and behaviors to be analyzed, generating data that can be labeled and subsequently used to build models to detect activities in real time. This is especially useful in the healthcare context, where accurate activity monitoring can help prevent incidents, assess patient condition and support personalized care (Johnson et al., 2021).

The application of AI in HAR has improved the accuracy and efficiency in the identification of activities. However, these methods often require labeled data in order to train the models properly. The quality of this data is essential, as the correct functioning of the models to perform accurate classification depends heavily on properly defined labels. In this sense, platforms that allow flexible and controlled labeling are relevant, as they ensure the system's adaptability for different contexts and users.

The ability to personalize and adapt HAR platforms is a crucial component in the usability of these technologies in healthcare environments. On the one hand, there are solutions for tagging data based on collaborative effort (Chang et al., 2017; Wang et al., 2012; Huang and Zhao, 2024), although this type of solution achieves its goal, it requires a large number of people depending on the data to be tagged. On the other hand are platforms that allow automatic data labeling (Ratner et al., 2017; Wu et al., 2021; Dong et al., 2014), but rely on the use of machine learning models to label the data. Although depending on the data to be labeled there may be greater or lesser accuracy in the final results. In the case of StreamTag, an interface has been developed that allows selecting predetermined activities or creating and tagging customized activities according to the context, a functionality that facilitates the adaptation of the system to various situations and allows activities to be recorded in a semiautomatic-way. The need to store data in an efficient and scalable manner has led to the adoption of non-relational databases, such as MongoDB, which allow data to be stored without a rigid structure, adapting to changes in storage requirements and the volume of activity data that is continuously generated. This type of database is for this reason ideal for systems such as StreamTag, which need to handle heterogeneous information and fast queries without compromising storage flexibility.

3 ARCHITECTURE

This section presents the architecture of StreamTag, detailing its various components.

Firstly, as illustrated in Figure 1, there are different layers that interact with each other, specifically the application layer, a reverse proxy, an API, and the database. The foundational component of this system is the application. Both the development of the application and the other essential elements within the system are carefully guided by a set of structured guidelines and a set of considerations. One of these is the decision to proceed with native development for the Android platform. This choice primarily stems from Android's significant market share, which surpasses that of other mobile operating systems, further motivating this selection are practical factors such as the need for accelerated development cycles suited to internal projects and the straightforward deployment capabilities that Android provides. The application can be distributed directly to devices, circumventing the lengthy approval processes often required by other platforms. The activities detected by the system are securely stored locally on the device. This data remains on the device until the user manually activates the transmission process, at which point the information is sent to the server for further processing or analysis.

The system's next component is the reverse proxy, which is implemented using Nginx. Nginx is a web server configured to handle the redirection of application requests to the backend API, optimizing request management. Beyond its redirecting function, the reverse proxy increase the security by enforcing TLS encryption standards, specifically versions 1.2 and 1.3, which facilitate secure data exchange and ensure confidentiality. Moreover, Nginx is enhanced with protective rules against potential saturation attacks and unsupported methods in the API that could introduce vulnerabilities, such as DELETE requests.

The API layer, constructed using the FastAPI framework in Python, serves as the interface through which the application interacts with the system's database. FastAPI is selected here for its efficiency in handling professional-grade API development. Within this layer, a series of endpoints has been designed to enable data access for the application. To maintain secure access, the API relies on JSON Web Tokens (JWT) for authentication. Each time a user logs in, a JWT is generated and must accompany each request's headers, ensuring that access to API endpoints is authenticated and secure.

The data storage layer leverages MongoDB, a non-relational database. This choice is justified by the system's lack of complex relational data requirements, thereby eliminating the need for cross-queries and making MongoDB an optimal solution. The data storage setup is centralized, housing user details and default application data. Should alternative databases be required, other MongoDB-based instances can be seamlessly integrated by updating connection credentials. However, it is essential to note that user access credentials will continue to be centralized in the main database to maintain consistency and control.

Listing 1: Sample Document Actitity Structure.

```
{
  "_id": {
    "$oid": "660
       cf794bdc4eab8625dfd4e"
  },
  "user": "llopez",
  "date_init": {
    "$numberLong":
       "1712124641448"
  },
  "date_end": {
    "$numberLong":
       "1712125839665"
  },
  "tag": "Eating",
  "additional_info": "Coffe
     Breakfast"
}
```

- **_id:** A unique identifier for each document, represented by an ObjectId. The `$oid` key designates the format used by MongoDB.

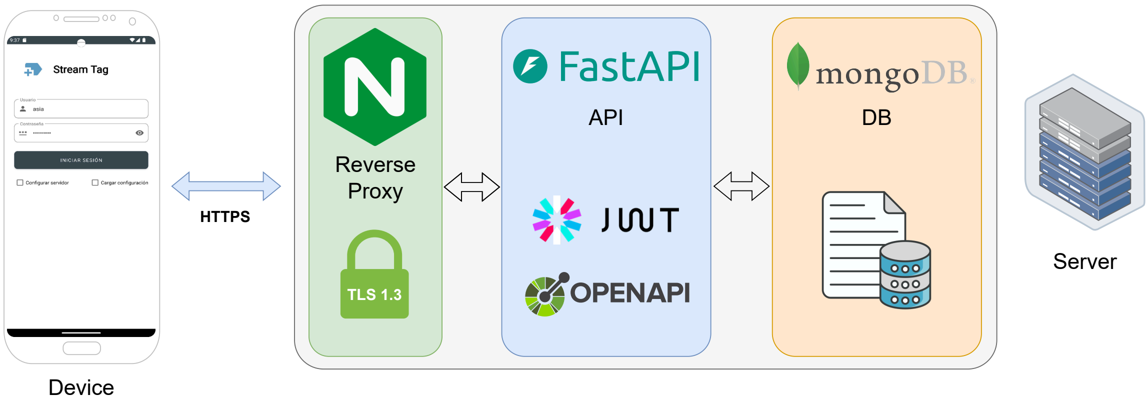

Figure 1: System components.

Figure 2: API endpoints.

- **user:** Stores the username associated with the activity. In this example, the user is `"llopez"`, linking the document to a specific individual.
- **date_init** and **date_end:** Timestamps indicating the start and end times of the activity. These are stored as 64-bit integers (`$numberLong`) for precision. These fields allow for calculating the duration of the activity.
- **tag:** Represents the type of activity. In this example, `"Eating"` classifies the document as related to food intake, enabling easy categorization and filtering.
- **additional_info:** Provides additional context for the activity. Here, it contains `"Coffe Breakfast"`, indicating the specific meal.

Additionally store user authentication data, including personal and security information.

Listing 2: Sample Document User Structure.

```
{
  "_id": "650
      d55ec652c99880a8c54f9",
  "username": "llopez",
  "full_name": "Jose Luis Lopez
      Ruiz",
  "email": "llopez@ujaen.es",
  "hashed_password": "
      $argon2id$v=19$m=65536,t=3
      ,p=4$SMn5vxei9B4DIMS4t3ZO
      ...",
  "disabled": false
}
```

It should be noted that the information is encrypted at the disk level, rather than at the database level, as this functionality is reserved for premium versions of mongoDB.

4 SYSTEM

In this section, information on the use of the Stream Tag platform is provided. The functionality of the StreamTag platform is divided into several views, each designed to allow for easy user interaction and efficient management of labeled data. These views provide both default options and advanced configurations, allowing for a personalized and optimized experience for the healthcare environment in nursing homes or any environment requiring activity monitoring and labeling. The main sections of the application and their functionalities are described below.

4.1 Login

Access to the system starts with the login screen, Figure 3, where the user is presented with the StreamTag logo and two fields to enter his credentials (username and password). This interface also includes a login button and two additional options: configuration of advanced parameters and loading the last server configuration. These options allow customizing the startup process, ensuring that the system adapts to the specific conditions of use, such as the security requirements of the data server.

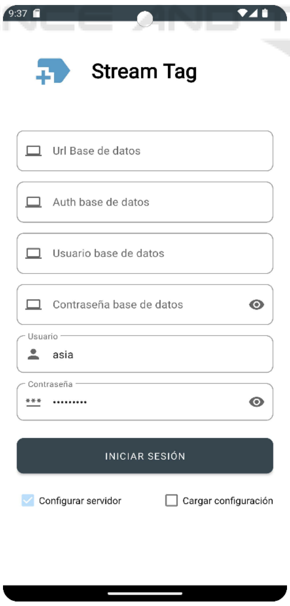

Figure 3: Login with additional options.

The user can access the system by entering their name and password and pressing the login button. To optimize the experience, StreamTag automatically remembers the credentials entered at the last login, which streamlines future logins. If the user selects the "Configure Server" option, a set of advanced options are displayed that allow the user to define the database URL for storing the labels, along with authentication settings and user-specific access data. This modular and configurable system allows data storage to operate in an isolated and secure manner, guaranteeing the integrity and privacy of the information generated and complying with security standards in the handling of sensitive data.

4.2 Main View

After logging in, the user accesses the main view Figure 4, an interface divided into three functional sections.

Figure 4: Main view.

4.2.1 Tag Upload

The first section contains a button that allows uploading the generated labels to the server, along with a text field where the user can specify the name of the collection in which the data will be saved. This name can be defined manually or, if the user leaves the field empty, it will be generated automatically based on the user's name and the current date. This functionality ensures that data is stored in an orderly manner and can be easily retrieved for further analysis, optimizing data management on the server.

4.2.2 Activity Management

In the second section, the user can select and incorporate specific activities from a predefined list. This list, defined on the server, is adapted according to the specific configuration of the database, showing only the authorized activities for each user or context. The user can add an individual activity using the “Add” button or load a set of preconfigured activities using the “Add full day” button. This configuration capability is particularly useful in contexts where recurring or full-day activities are required, thus simplifying interaction and reducing the time required to define activities.

4.2.3 Viewing and Managing Activities

The third section displays the activities that the user has added in the current session. Each activity appears with its name and specific buttons to start, stop or delete it, allowing precise control over the recording of each activity. Additionally, each activity includes a text field where the user can add complementary details, such as the type of medication administered or any relevant observations. The list of activities can be reordered with a long press and a drag-and-drop motion, which facilitates organization and management according to the user’s priorities or context of use.

4.3 Custom View

The custom view, Figure 5, offers an alternative to the main view, providing greater flexibility in activity management. This view differs in two fundamental aspects:

4.3.1 Dynamic Activity Input Field

Instead of a predefined list, this view features a text entry field that allows the user to enter specific activities not included in the predefined set. This functionality is ideal for users who need to record unusual or unique activities, expanding the application’s customization and adaptability possibilities of the application.

4.3.2 Detailed Activity Configuration

Unlike the main view, the custom view does not include an “Add Full Day” button, allowing the user to design and manage activities in a more detailed and individualized way. This is useful in scenarios where a high level of specificity in activity recording is required, as each entry can be manually adjusted to the user’s needs.

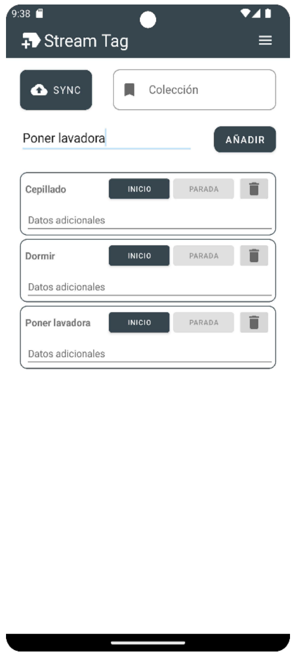

Figure 5: Custom view.

5 CONCLUSIONS

The primary problem addressed in this work is the growing need to manage labeled data in the healthcare sector, particularly in hospital residency settings, where structured access to historical and categorized information is essential for quality care and clinical decision-making. The lack of an efficient infrastructure to collect, organize, and use these data in a secure and scalable way presents a considerable challenge in these environments.

As a solution, an architecture has been developed, consisting of a native application, a reverse proxy, a modern API, and a non-relational database. This platform enables structured management of labeled data, facilitating access and analysis for the creation of detailed health profiles and the identification of patterns in residents’ health status. Integration capability with IoT devices also optimizes automated data collection, benefiting healthcare staff by providing constant access to organized and categorized information, thus improving operational efficiency and data security.

To assess the impact and expand the utility of the platform, future work should focus on conducting studies on clinical and operational effectiveness. These studies would provide objective metrics on its contribution to healthcare delivery and organization. Additionally, usability tests aimed at users with minimal training would help ensure that the system is accessible and easy to use, confirming its applicability across various clinical settings and levels of staff training.

ACKNOWLEDGMENTS

This result has been partially supported by grant PID2021-127275OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by "ERDF A way of making Europe", grant PDC2023-145863-I00 funded by MICIU/AEI/10.13039/501100011033 and by "European Union NextGenerationEU/PRTR", and grant M.2 PDC_000756 funded by Consejería de Universidad, Investigación e Innovación and by ERDF Andalusia Program 2021-2027.

REFERENCES

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H. A., et al. (2023). Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC medical education*, 23(1):689.

Bhuiyan, M. N., Rahman, M. M., Billah, M. M., and Saha, D. (2021). Internet of things (iot): A review of its enabling technologies in healthcare applications, standards protocols, security, and market opportunities. *IEEE Internet of Things Journal*, 8(13):10474–10498.

Chang, J. C., Amershi, S., and Kamar, E. (2017). Revolt: Collaborative crowdsourcing for labeling machine learning datasets. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 2334–2346, New York, NY, USA. Association for Computing Machinery.

Dang, L. M., Min, K., Wang, H., Piran, M. J., Lee, C. H., and Moon, H. (2020). Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognition*, 108:107561.

Davenport, T. and Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future healthcare journal*, 6(2):94–98.

Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., and Zhang, W. (2014). Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, page 601–610, New York, NY, USA. Association for Computing Machinery.

Duan, Y., Edwards, J., and Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of big data - evolution, challenges and research agenda. *Int. J. Inf. Manag.*, 48:63–71.

Haug, C. J. and Drazen, J. M. (2023). Artificial intelligence and machine learning in clinical medicine, 2023. *New England Journal of Medicine*, 388(13):1201–1208.

Huang, Q. and Zhao, T. (2024). Data collection and labeling techniques for machine learning. *arXiv preprint arXiv:2407.12793*.

Jobanputra, C., Bavishi, J., and Doshi, N. (2019). Human activity recognition: A survey. *Procedia Computer Science*, 155:698–703.

Johnson, K. B., Wei, W.-Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., Zhao, J., and Snowdon, J. L. (2021). Precision medicine, ai, and the future of personalized health care. *Clinical and translational science*, 14(1):86–93.

Ke, S.-R., Thuc, H. L. U., Lee, Y.-J., Hwang, J.-N., Yoo, J.-H., and Choi, K.-H. (2013). A review on video-based human activity recognition. *Computers*, 2(2):88–131.

Kumar, Y., Koul, A., Singla, R., and Ijaz, M. F. (2023). Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *Journal of ambient intelligence and humanized computing*, 14(7):8459–8486.

Laghari, A. A., Wu, K., Laghari, R. A., Ali, M., and Khan, A. A. (2021). A review and state of art of internet of things (iot). *Archives of Computational Methods in Engineering*, pages 1–19.

Miller, D. and Brown, E. (2017). Artificial intelligence in medical practice: The question to the answer? *The American journal of medicine*, 131 2:129–133.

Mouha, R. A. R. A. et al. (2021). Internet of things (iot). *Journal of Data Analysis and Information Processing*, 9(02):77.

Ramanujam, E., Perumal, T., and Padmavathi, S. (2021). Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review. *IEEE Sensors Journal*, 21(12):13029–13040.

Ratner, A. J., Bach, S. H., Ehrenberg, H. R., and Ré, C. (2017). Snorkel: Fast training set generation for information extraction. In *Proceedings of the 2017 ACM International Conference on Management of Data*, SIGMOD '17, page 1683–1686, New York, NY, USA. Association for Computing Machinery.

Shaikh, M. B. and Chai, D. (2021). Rgb-d data-based action recognition: a review. *Sensors*, 21(12):4246.

Wang, J., Kraska, T., Franklin, M. J., and Feng, J. (2012). Crowder: crowdsourcing entity resolution. *Proc. VLDB Endow.*, 5(11):1483–1494.

Wu, R., Sakala, P., Li, P., Chu, X., and He, Y. (2021). Demonstration of panda: a weakly supervised entity matching system. *Proc. VLDB Endow.*, 14(12):2735–2738.

Yu, K.-H., Beam, A. L., and Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature biomedical engineering*, 2(10):719–731.

Zhang, C. and Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, 23:100224.