Chronic Disease Detection with an IoHT-Based Ensemble LSTM Framework: A Unified Approach for Enhanced Healthcare Diagnostics

Nama Singh[1], Dr. Margi Patel[2], Gaurav Kumar Gautam[3]

[1,2]Department of Computer Science and Engineering, Indore Institute of science and technology, Indore (M.P.)
[3]Department of Physical Education, Aligarh Muslim University, Aligarh (U.P.)
namasingh33@gmail.com, Margi.patel@indoreinstitute.com, gauravmtr93@gmail.com

* Corresponding Author: Nama Singh

Abstract: *This paper presents a Internet of Health Things (IoHT) framework for the detection of chronic diseases, leveraging the capabilities of IoT technologies to revolutionize healthcare monitoring, diagnostics, and treatment processes. The proposed framework is structured into three layers, including smart sensors for data collection, gateway devices for data analysis and transmission, and cloud storage for making the data accessible to medical professionals. At its core, the framework employs an ensemble Long Short-Term Memory (LSTM) model for the diagnosis of multiple chronic diseases such as kidney, heart, cancer, and diabetes, demonstrating superior performance in terms of accuracy, execution time, and model complexity when compared to existing models. The methodology encompasses data collection, minority data augmentation using Synthetic Minority Over-sampling Technique (SMOTE), rigorous data preprocessing, optimal feature extraction, and disease diagnosis through the ensemble LSTM model. The performance evaluation of the model shows notable effectiveness across different diseases, achieving an average accuracy, precision, recall, and F1-score of approximately 92-92.5%.*

Keywords: *Smart Healthcare, IoT, Disease Diagnosis, Unified Framework*.

1. Introduction

The Internet of Things (IoT) represents the transformation of physical objects into digital entities, enhancing real-world items with intelligence by embedding computing capabilities. This process enables devices, whether connected to cloud servers or not, to be integrated with backend applications for data analysis, visualization, and control. IoT facilitates communication between everyday objects using sensors, such as RFID, and actuators, to collect and transmit data over the Internet, bridging the gap between the physical and digital realms. This concept is driven by the need to manage the vast amounts of data generated by these devices, address inefficiencies in system operations, and counteract the significant energy consumption associated with digital expansion. The anticipated increase in connected devices to 100 billion by 2030 underscores the critical role of IoT in various sectors, including health informatics and telematics. In this context, IoT, coupled with advancements in computing and big data environments, is leveraged to develop AI-based disease prediction mechanisms. Despite the development of numerous IoT-based disease prediction methods, the integration of AI with IoT devices remains a relatively unexplored frontier, offering significant potential for innovation in disease prediction and other applications [1]. The problem domain highlights several critical issues within biomedical contexts and health-support systems. These challenges include the rising complexity of biomedical problems, significant hurdles in health-support systems such as inadequate medical information, preventable errors, data breaches, misdiagnosis, and delays in data transmission [2]. Additionally, current models suffer from overfitting, high computational demands, and substantial memory requirements, with a particular focus on cancer detection which limits their applicability to other diseases. The proposed solution domain aims to address these challenges by integrating IoT technology into the development of medical diagnostic models [3]. The objectives include creating an IoT-based model for remote disease diagnosis, designing a learning model that is less complex yet capable of predicting outcomes with minimal expected error, and developing a framework capable of detecting multiple diseases. IoT technology, described as having four stages, facilitates the seamless flow of information from one node to another, enabling diverse data collection and interpretation. This approach promises to enhance the efficiency and scope of disease detection and diagnosis, moving beyond single-disease frameworks to a more versatile and comprehensive health monitoring solution.

2. Literature Review

Lv et al. [1] developed a CNN-based interactive smart healthcare prediction and evaluation model (SHPE model) with an accuracy of 82.4%. Ferdousi et al. [2] proposed a machine learning framework for early risk prediction of diabetes using wearable IoT sensor data, achieving 94% accuracy with the Random Tree algorithm. Muthu et al. [3] introduced a system using Generalize approximate Reasoning base Intelligence Control (GARIC) with regression rules and deep learning mechanisms for patient data processing from IoT. Umar et al. [4] enhanced EEG data classification for cognitive healthcare using pre-trained CNN models, demonstrating the efficacy of end-to-end learning. Liyakathunisa, et al. [5] suggested a method for monitoring healthcare using sensor and IoT devices, employing intelligent algorithms for faster analysis and better treatment recommendations. Manogaran et al. [6] proposed a scalable machine learning technique for identifying DNA range changes using Bayesian HMM and GM clustering. Pham et al. [7] presented DeepCare, an RNN architecture for predicting future medical risks with improved accuracy over traditional models. Mukati et al. [8] discussed the benefits of IoT in healthcare, highlighting its potential in managing various medical conditions and improving healthcare performance. Sykes et al. [9] found that 86% of COVID-19 patients reported residual symptoms, suggesting the biopsychosocial effects may play a significant role in Long-COVID. Kovács et al. [10] reviewed the effectiveness of chest CT scans for COVID-19 detection, emphasizing its potential compared to RT-PCR. Rhee et al. [11] summarized evidence on SARS-CoV-2 infectivity, indicating that the virus is most contagious around symptom onset with decreasing infectivity thereafter. Otoom et al. [12]proposed an IoT framework for real-time symptom data collection and analysis for early COVID-19 case identification, with five algorithms showing over 90% accuracy. Tuli et al. [13] applied an ML-based model to predict COVID-19 growth, showing improved accuracy with the Generalized Inverse Weibull distribution. Kumar et al. [14] discussed using IoT to monitor health and minimize COVID-19 spread, emphasizing the technology's potential in pandemic management. Vedaei et al. [15] proposed a framework combining IoT nodes, smartphone apps, and fog-based ML for real-time health monitoring and infection risk prediction, highlighting the importance of maintaining physical distance. Venkatrao and Kareemulla [16] proposed a hybrid deep learning approach for chronic kidney disease classification. Sánchez [17] presented a data-driven approach for chronic kidney disease diagnosis.

3. Overview of Internet of Healthcare Things (IoHT)

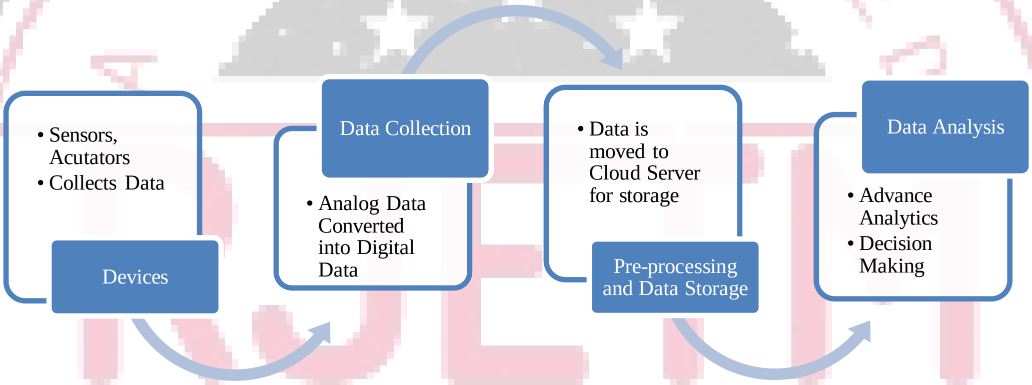

Fig.1. Stages of IoHT Data Analysis

The process for disease detection in the context of the Internet of Health Things (IoHT) involves a four-step data handling procedure, aimed at improving healthcare outcomes through advanced IoT technologies, as in fig 1. The steps are as follows:

- Data Collection: Utilizes a variety of devices such as sensors, actuators, monitors, detectors, and camera systems to gather data relevant to healthcare.
- Data Conversion: The collected data, which is initially in analog form, undergoes conversion into digital format for easier processing and analysis.
- Data Pre-processing and Storage: The digital data is then pre-processed, standardized, and transferred to data centers or the cloud for storage.
- Data Management and Analysis: Finally, the data is managed and analyzed using Advanced Analytics Applications, enabling effective decision-making and insights that can lead to improved healthcare outcomes.

The IoHT redefines healthcare by ensuring better care, improved treatment outcomes, reduced medical costs, optimized processes and workflows, enhanced performance, and a superior patient experience.

4. Proposed Methodology

The proposed IoHT framework for disease detection is structured into three main layers, fig 2:

- Layer 1: Comprises smart sensors and information related to patients, hospitals, or insurance companies, among others.
- Layer 2: Involves gateway or switching devices responsible for analyzing and transmitting the collected data.
- Layer 3: Data analyzed by medical professionals is stored in the cloud, making the framework robust, flexible, and bidirectional.

This systematic approach leverages IoT technologies to enhance healthcare monitoring, diagnostics, and treatment, demonstrating a comprehensive and scalable model for future health technologies.

In this model, a unified framework is presented for chronic disease detection using ensemble LSTM model. The entire working is composed of following steps:

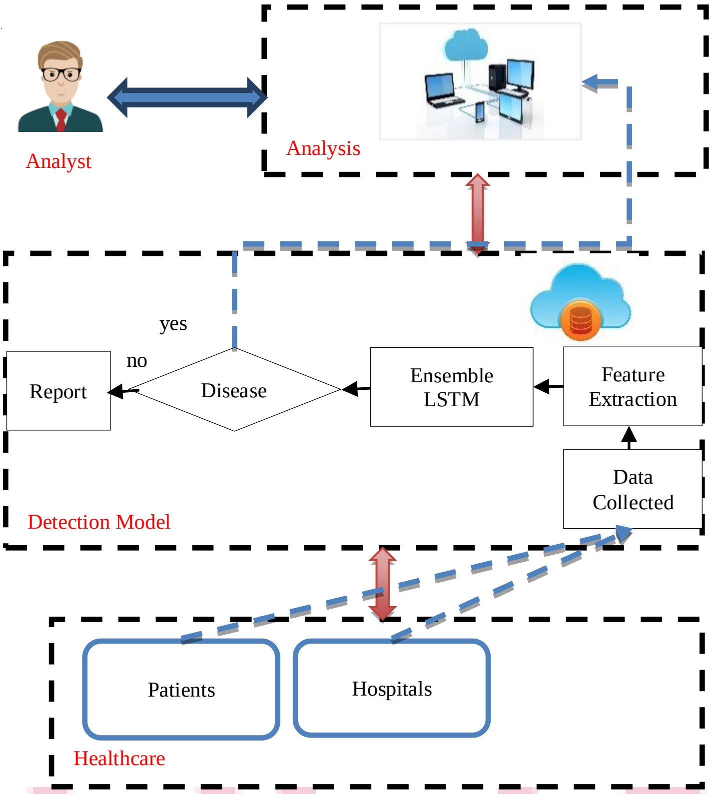

Fig. 2. Proposed Model

Data Collection from different sources: In this step, the data is collected from various publicly available repositories, such as Kaggle [18]-[21]. This ensures a comprehensive and diverse dataset for a robust analysis, encompassing multiple chronic disease such as kidney [18], diabetes [19], heart [20], and cancer [21].

Minority Data Augmentation: This step is performed to augment new data using Synthetic Minority Over-sampling Technique (SMOTE). This is a method designed to combat class imbalance in datasets by generating synthetic samples of the minority class to achieve a balance with the majority class. Unlike simple replication, SMOTE uses the k-nearest neighbors algorithm to create new instances that combine features of existing minority samples and their nearest neighbors, thereby avoiding overfitting and enhancing model generalizability. This process not only increases the quantity of minority class data but also diversifies it, making the dataset more representative and improving the performance of machine learning models on imbalanced datasets. SMOTE is described as a "balancer," emphasizing its utility in creating a more equitable class distribution for better analytical outcomes.

Pre-processing: During pre-processing following steps are performed as presented in fig 3.

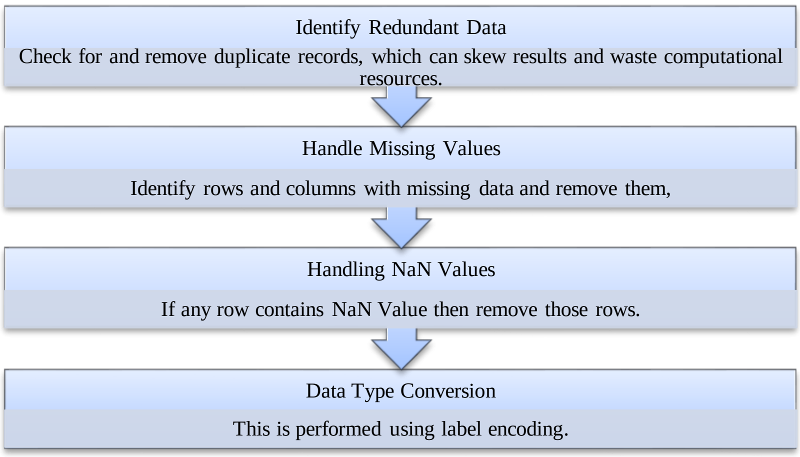

Fig. 3. Data Preprocessing

In the pre-processing step, the focus is on refining the dataset to ensure its suitability for machine learning. This involves identifying and eliminating duplicate records to prevent skewed results. Rows and columns with missing values are removed to maintain data integrity, with special attention to rows containing NaN values, which are also discarded.

Additionally, data types are standardized, converting features into their appropriate formats by applying label encoding to categorical variables, to facilitate the model's ability to accurately interpret and learn from the data.
In this step, the label encoding method is used to convert the filtered data after statistical evaluation so that they can result in better prediction results. This method is adopted because it provides better prediction results of machine learning with those data that show no relationship to each other.
Its operation can be described using a simple mathematical representation.
Let's consider a set of categorical labels $L = \{l_1, l_2, \ldots, l_n\}$, where each l_i is a unique categorical label in the dataset. The Label Encoder performs the following steps:

- Sorting and Indexing: It first sorts the unique labels in the dataset. This sorting is typically in alphabetical order (for string labels) or in ascending order (for numeric labels). Let $S = \{s_1, s_2, \ldots, s_n\}$ be the sorted set of unique labels.
- Mapping to Integers: Each sorted label si is then mapped to a unique integer. The mapping function is defined as:
- Where $f(l_i) = j$, where $s_j = l_i$
- Here, j is the index of label l_i in the sorted set S, starting from 0.
- Encoding: For each label l_i in the dataset, the Label Encoder replaces it with $f(l_i)$, the corresponding integer value.
- Mathematically, the encoding of a label li can be seen as finding its index in the sorted array of unique labels.

Optimal Feature extraction using Feature Importance: The working methodology of decision tree based optimal feature selection to identify the most important features for a predictive model.
Disease Diagnosis using Ensemble LSTM: In this work, a ensemble LSTM model is presented for diagnosis of different diseases as kidney, cancer, diabetes, heart, and normal. The model depicted in fig 4 designed for disease classification. It follows these steps:

- Disease-Related Features: The input is divided into four different categories of disease-related features, specifically for heart disease, cancer, kidney disease, and diabetes.
- Feature Reshaping: Each category of disease features undergoes a reshaping process. This is likely meant to transform the data into a format suitable for processing by the next stage in the model. Reshaping could involve changing the dimensionality of the data to fit the input requirements of the LSTM networks.
- LSTM Networks: Each reshaped feature set is fed into its own Long Short-Term Memory (LSTM) network. LSTM networks are a type of recurrent neural network (RNN) that are well-suited for sequential data and are capable of learning long-term dependencies. They are commonly used in time-series analysis, natural language processing, and other applications where data is temporally correlated.
- Ensemble Result: The outputs of the individual LSTM networks are then combined in an ensemble manner. Ensemble methods typically involve combining the predictions of multiple models to improve accuracy and reduce the likelihood of overfitting. This ensemble result encapsulates the predictions from all four LSTM networks.
- Disease Classification: Finally, the ensemble result is used for disease classification. The system classifies the input data into one of the diseases based on the learned features and patterns from the LSTM networks.

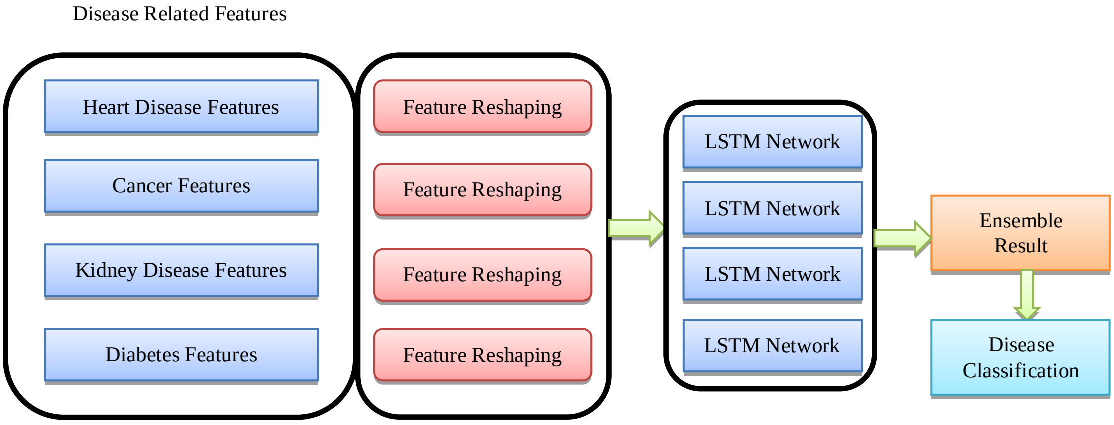

Fig. 4. Architecture of Ensemble LSTM

5. Results and Discussion

Following Performance parameters are used:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$F1\_score = \frac{(2 * Precision * Recall)}{(Precision + Recall)}$$

Where TP, FN, FP and TN represent the number of true positives, false negatives, false positives and true negatives, respectively.
The table 1 presents a performance evaluation of a diagnostic model across four different diseases: Kidney, Heart, Cancer, and Diabetes, using metrics such as Accuracy, Precision, Recall, and F1-Score. The model shows good performance in diagnosing Kidney and Cancer with nearly perfect scores of 99% and 98% across all metrics, respectively. However, it has slightly lower effectiveness for Heart disease, with scores around 89-90%, and the lowest performance is observed in Diabetes diagnosis. Overall, the model achieves an average performance of approximately 92-92.5% across all metrics, indicating a high level of diagnostic accuracy and reliability across a range of diseases.

Table 1. Performance Evaluation

Disease Type	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Kidney	99	99	99	99
Heart	89	90	89	90
Cancer	98	98	98	98
Diabetes	82	83	83	83
Average	92	92.5	92	92.5

Table 2 provides a comparative analysis of different diagnostic models. The first existing model [16], utilizes Deep Learning (DL) and achieves an accuracy of 99.19% with an execution time of approximately 9 seconds, but is noted for its high complexity and susceptibility to overfitting. The second Existing model [17], employs an Xgboost Classifier, achieving a slightly lower accuracy of 98% also characterized by high complexity and overfitting issues. In contrast, the proposed model adopts an Ensemble Long Short-Term Memory (LSTM) approach, which not only covers multiple disease types but also outperforms the existing models in accuracy at 99.54%, while reducing execution time to around 7 seconds. Remarkably, the proposed model is distinguished by its low complexity and the absence of overfitting, indicating a more efficient and reliable diagnostic tool compared to the existing models.

Table 2. Comparative State-of-Art

	Existing [16]	**Existing [17]**	**Proposed**
Algorithm	DL	Xgboost Classifier	Ensemble LSTM
Disease Type	Single (Kidney)	Single (Kidney)	Multiple
Accuracy	99.19%	98%	99.54%
Execution Time	~9 sec	-	~7 Sec
Complexity	High	High	Low
Overfitting	Yes	Yes	No

6. Conclusion

The development and implementation of the proposed IoHT framework for disease detection signify a substantial advancement in healthcare technology. By integrating smart sensors, data analysis gateways, and cloud storage, the framework ensures robustness, flexibility, and bidirectionality in healthcare data management. The ensemble LSTM model at the heart of this framework offers a comprehensive solution for chronic disease detection, outperforming existing models in accuracy, efficiency, and simplicity. Through innovative data collection, augmentation, preprocessing, and optimal feature extraction techniques, the model demonstrates high diagnostic accuracy across multiple diseases, addressing the challenges of overfitting and model complexity seen in previous approaches. The success of this framework heralds a new era in the application of IoHT technologies for enhancing healthcare outcomes, providing a scalable and efficient model for future advancements in medical diagnostics and treatment planning.

References

[1] Z Lv, Z Yu, S Xie, A Alamri, "Deep learning-based smart predictive evaluation for interactive multimedia-enabled smart healthcare", ACM Transactions on Multimedia Computing, 2022.
[2] R. Ferdousi, M. A. Hossain and A. E. Saddik, "Early-Stage Risk Prediction of Non-Communicable Disease Using Machine Learning in Health CPS," in IEEE Access, vol. 9, pp. 96823-96837, 2021, doi: 10.1109/ACCESS.2021.3094063.
[3] Muthu, B., Sivaparthipan, C.B., Manogaran, G. et al. IOT based wearable sensor for diseases prediction and symptom analysis in healthcare sector. Peer-to-Peer Netw. Appl. 13, 2123–2134 (2020). https://doi.org/10.1007/s12083-019-00823-2
[4] Amin SU et al (2019) Cognitive smart healthcare for pathology detection and monitoring. IEEE Access 7:10745–10753 27.
[5] Syed L et al. (2019) Data science algorithms and techniques for smart healthcare using IoT and big data analytics. Smart Techniques for a Smarter Planet. Springer, Cham: 211–241.
[6] Manogaran G et al (2018) Machine learning based big data processing framework for cancer diagnosis using hidden Markov model and GM clustering. Wirel Pers Commun 102(3):2099–2116.
[7] P. Nguyen, T. Tran, N. Wickramasinghe, and S. Venkatesh, "Deepr: A convolutional net for medical records," IEEE J. Biomed. Health Inform., vol. 21, no. 1, pp. 22–30, Jan. 2017.

[8] Mukati, Naveen, et al. "Healthcare assistance to COVID-19 patient using internet of things (IoT) enabled technologies." Materials today: proceedings 80 (2023): 3777-3781.

[9] Sykes, Dominic L., et al. "Post-COVID-19 symptom burden: what is long-COVID and how should we manage it?." Lung 199 (2021): 113-119.

[10] Kovács, Anita, et al. "The sensitivity and specificity of chest CT in the diagnosis of COVID-19." European Radiology 31 (2021): 2819-2824.

[11] Rhee, Chanu, et al. "Duration of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infectivity: when is it safe to discontinue isolation?." Clinical infectious diseases 72.8 (2021): 1467-1474.

[12] Otoom, Mwaffaq, et al. "An IoT-based framework for early identification and monitoring of COVID-19 cases." Biomedical signal processing and control 62 (2020): 102149.

[13] Tuli, Shreshth, et al. "HealthFog: An ensemble deep learning based Smart Healthcare System for Automatic Diagnosis of Heart Diseases in integrated IoT and fog computing environments." Future Generation Computer Systems 104 (2020): 187-200.

[14] Kumar, Krishna, Narendra Kumar, and Rachna Shah. "Role of IoT to avoid spreading of COVID-19." International Journal of Intelligent Networks 1 (2020): 32-35.

[15] Vedaei, Seyed Shahim, et al. "COVID-SAFE: An IoT-based system for automated health monitoring and surveillance in post-pandemic life." IEEE access 8 (2020): 188538-188551.

[16] K. Venkatrao and S. Kareemulla, "HDLNET: A Hybrid Deep Learning Network Model With Intelligent IOT for Detection and Classification of Chronic Kidney Disease," in IEEE Access, vol. 11, pp. 99638-99652, 2023, doi: 10.1109/ACCESS.2023.3312183.

[17] P. A. Moreno-Sánchez, "Data-Driven Early Diagnosis of Chronic Kidney Disease: Development and Evaluation of an Explainable AI Model," in IEEE Access, vol. 11, pp. 38359-38369, 2023, doi: 10.1109/ACCESS.2023.3264270.

[18] Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., & Johannes, R.S. (1988). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. *In Proceedings of the Symposium on Computer Applications and Medical Care* (pp. 261--265). IEEE Computer Society Press.

[19] https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease

[20] Alizadehsani, R., Roshanzamir, M., Abdar, M. et al. A database for using machine learning and data mining techniques for coronary artery disease diagnosis. Sci Data 6, 227 (2019). https://doi.org/10.1038/s41597-019-0206-3.

[21] https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer