ML Based Health Prognosis: Forecasting Medical Conditions

[1] Pratik Shrimant Bhange
Electronics and Computer Engineering
P. E. S's Modern College Of Engineering
Pune, India
bhangepratik3@gmail.com

[2]Shubham Dilip Yadav
Electronics and Computer Engineering
P. E. S's Modern College Of Engineering
Pune, India
shubhamdyadav17@gmail.com

[3]Vipul Suresh Raut
Electronics and Computer Engineering
P. E. S's Modern College Of Engineering
Pune, India
vipul_raut@moderncoe.edu.in

[4] Mrs. Sarojini Naik
Electronics and Computer Engineering
P. E. S's Modern College Of Engineering
Pune, India

Abstract - The increasing utilization of Big Data in biomedical and healthcare sectors underscores the critical need for accurate analysis to enhance patient care and healthcare systems. This study addresses the challenge of predicting chronic diseases, including diabetes, hypertension, cerebral infarction, and asthma, by harnessing machine learning algorithms. To overcome the accuracy limitations posed by fragmented medical data, a novel approach, the Convolutional Neural Network-based Multimodal Disease Risk Prediction (CNN-MDRP) model, is proposed. This advanced model integrates structured and unstructured data, leveraging Genetic Algorithm for data cleansing and Recurrent Neural Network (RNN) processing to transform unstructured data. Disease prediction is facilitated through classifiers like Naive Bayes (NB), Support Vector Machine (SVM), and logistic regression. Furthermore, upon successful prediction, the system recommends nearby medical facilities to users, enhancing accessibility to healthcare. This interdisciplinary approach amalgamates machine learning, neural networks, and natural language processing to predict diseases accurately. Through comprehensive data analysis and model selection, the system offers a valuable tool for early disease detection and personalized patient care. The proposed model's integration with user-friendly web interfaces facilitates seamless interaction and accessibility. With promising accuracy rates achieved for disease prediction, this study paves the way for future advancements in predictive healthcare technologies, ultimately improving patient outcomes and healthcare delivery systems. Key words: Machine learning algorithms

Machine learning algorithms, Convolutional Neural Network (CNN), Genetic Algorithm, Recurrent Neural Network (RNN), Disease prediction, Support Vector Machine (SVM), Medical facilities, Big Data, Biomedicine, Chronic diseases, Healthcare, Predictive modeling, Neural networks, Natural Language Processing (NLP), User interface, Accessibility.

INTRODUCTION

The contemporary landscape of healthcare is witnessing a remarkable transformation driven by the burgeoning utilization of Big Data and machine learning algorithms. With the exponential growth in digital health records and medical data, there arises an unprecedented opportunity to revolutionize disease prediction and patient care. Chronic diseases such as diabetes, hypertension, cerebral infarction, and asthma pose significant challenges to healthcare systems worldwide, necessitating proactive measures for early detection and intervention.

This study endeavors to address these challenges by leveraging advanced machine learning techniques to predict chronic diseases accurately. The advent of machine learning algorithms has opened new avenues for analyzing vast volumes of medical data and extracting valuable insights to enhance healthcare outcomes. By harnessing the power of Convolutional Neural Networks (CNN), Genetic Algorithms, and Recurrent Neural Networks (RNN), this research aims to develop a robust disease prediction model that integrates both structured and unstructured data sources.

The proposed approach recognizes the inherent complexity and heterogeneity of medical data, acknowledging the need for innovative solutions to overcome data fragmentation and enhance predictive accuracy. Through the integration of state-of-the-art machine learning models such as Support Vector Machines (SVM) and logistic regression, coupled with natural language processing techniques, this study seeks to bridge the gap between disparate data sources and facilitate accurate disease prediction.

Furthermore, the development of a user-friendly front-end interface enhances accessibility and usability, ensuring seamless interaction with the predictive model. By providing recommendations for nearby medical facilities based on predicted disease outcomes, this integrated system aims to empower users with timely and personalized healthcare information.

In essence, this study represents a convergence of interdisciplinary efforts, blending machine learning, neural networks, and healthcare expertise to advance disease prediction and personalized patient care. By harnessing the potential of Big Data analytics and machine learning algorithms, this research endeavors to contribute towards the realization of predictive healthcare systems that improve patient outcomes and enhance healthcare delivery on a global scale.

LITERATURE REVIEW

Machine Learning

Machine learning is closely intertwined with computational statistics and often shares areas of overlap. It primarily revolves around the use of computers to make predictions. This field has strong connections with mathematical optimization, which offers various methods, theories, and application domains. Machine learning is sometimes confused with data mining, a related sub-field that is more oriented towards exploratory data analysis and is often referred to as unsupervised learning. [5] Machine learning can also function in an unsupervised manner, enabling it to learn and establish fundamental behavioral patterns for different entities and subsequently detect significant anomalies.

List of machine learning algorithms:

1)**Supervised Learning:**

Supervised learning plays a pivotal role in disease prediction and health prognosis by leveraging labeled datasets to train predictive models. In the context of the implementation paper focusing on predicting Parkinson's disease, Diabetes, and another specified disease, supervised learning techniques are applied to develop accurate predictive models. Below are some key aspects to include in the implementation paper regarding supervised learning:

Data Acquisition and Preprocessing:

Describe the process of acquiring labelled datasets for Parkinson's disease, Diabetes, and the third specified disease. Mention reputable sources, data collection methods, and any preprocessing steps involved, such as cleaning, normalization, and handling missing values.

Feature Engineering:

Discuss the features (input variables) used in the predictive models. These features may include demographic information (age, gender), clinical measurements (blood pressure, glucose levels), symptoms, medical history, and any relevant biomarkers or diagnostic tests.

Model Selection and Training:

Explain the choice of supervised learning algorithms selected for each disease prediction task. For example, decision trees or random forests may be suitable for Diabetes prediction, while Support Vector Machines (SVM) or logistic regression could be used for Parkinson's disease prediction.
Detail the training process, including how the data is split into training and testing sets, hyperparameter tuning methods, and any cross-validation techniques employed to ensure model generalization.

Evaluation Metrics:

Provide comprehensive information on the evaluation metrics used to assess the performance of the predictive models. This should include accuracy, precision, recall, F1-score, ROC-AUC, and any disease-specific evaluation metrics relevant to Parkinson's disease, Diabetes, and the third specified disease.
Results and Performance Analysis:

Present the results obtained from training and evaluating the supervised learning models. Include tables, charts, or visualizations illustrating the performance metrics for each disease prediction task.
Analyze the strengths and limitations of the predictive models, discussing factors that contribute to their performance, such as feature importance, model complexity, and potential biases.

Comparison with Baseline Models:

Compare the performance of the developed supervised learning models with baseline models or existing prediction methods. This comparison helps validate the effectiveness of the proposed approach and provides insights into the added value of using machine learning techniques for disease prediction.

Interpretability and Clinical Relevance:

Discuss the interpretability of the supervised learning models and their implications for clinical decision-making. Highlight the most influential features driving the predictions and how healthcare professionals can interpret and trust the model outputs in real-world settings.

Future Directions and Challenges:

Identify potential areas for improvement and future research directions in supervised learning for disease prediction. Address any challenges encountered during the implementation process, such as data availability, model interpretability, and scalability to larger datasets or different healthcare settings.

2) Unsupervised Learning:

In the context of disease prediction and health prognosis, unsupervised learning techniques are utilized to identify hidden patterns, groupings, or structures within the data without explicit guidance from labelled examples. Here are some key aspects to include in the implementation paper regarding unsupervised learning:

Data Exploration and Preprocessing:

Describe the process of exploring and preprocessing the dataset before applying unsupervised learning techniques. This may involve data cleaning, normalization, handling missing values, and transforming variables as necessary to prepare the data for analysis.

Clustering Algorithms:

Explain the choice of clustering algorithms used for identifying natural groupings within the dataset. Commonly used clustering algorithms include K-means clustering, hierarchical clustering, and Gaussian mixture models.
Provide a brief overview of each clustering algorithm selected and how they operate to partition the data into clusters based on similarity or distance metrics.

Dimensionality Reduction Techniques:

Discuss the application of dimensionality reduction techniques to visualize high-dimensional data and extract essential features. Principal Component Analysis (PCA), t-distributed Stochastic Neighbor Embedding (t-SNE), and Singular Value Decomposition (SVD) are popular methods for reducing the dimensionality of the data.
Explain how dimensionality reduction helps in identifying underlying patterns and simplifying the representation of complex datasets while preserving as much information as possible.

Feature Extraction and Representation:

Detail the process of extracting meaningful features or representations from the data using unsupervised learning techniques. This may involve clustering similar instances together or transforming the data into a lower-dimensional space for visualization and analysis.
Discuss how the extracted features can be used for downstream tasks such as anomaly detection, pattern recognition, or as input to supervised learning models for disease prediction.

Evaluation and Validation:

Describe the evaluation metrics or techniques used to assess the quality of the clustering or dimensionality reduction results. While unsupervised learning does not have explicit labels for evaluation, internal metrics such as silhouette score, Davies–Bouldin index, or external validation methods like silhouette analysis can be used to evaluate the clustering quality.
Discuss the interpretability of the clustering results and how they align with domain knowledge or known disease subtypes.

Visualization and Interpretation:

Present visualizations of the clustering results or dimensionality-reduced representations to aid in interpretation and understanding. Visualizations such as scatter plots, heatmaps, or dendrograms can help visualize clusters and relationships within the data.
Interpret the findings from the visualizations in the context of disease prediction and health prognosis, highlighting any insights gained from the unsupervised learning analysis.
Integration with Supervised Learning:

Discuss how the insights obtained from unsupervised learning techniques can complement or enhance supervised learning models for disease prediction. For example, clustering results may be used to identify subgroups of patients with similar characteristics, which can inform the development of personalized predictive models.

Future Directions and Challenges:

Identify potential avenues for future research in unsupervised learning for disease prediction and health prognosis. Address any challenges encountered during the implementation process, such as scalability to large datasets, robustness to noise, and interpretability of clustering results**.**

3) Web Interface Development:

Frontend Technologies:

In this implementation, we utilized the classic trio of web development: HTML, CSS, and JavaScript. HTML served as the backbone, structuring the content and laying the foundation for user interaction. CSS was leveraged extensively to stylize and layout the interface, ensuring a visually appealing and consistent design across different pages and devices.

JavaScript added interactivity and dynamic behavior, enabling features like form validation, real-time updates, and enhanced user experience.

User Interface Components:

Our web interface comprised various user interface components carefully designed to streamline user interaction. Input forms allowed users to input relevant medical data seamlessly, with different field types catering to diverse data types and requirements. Additionally, interactive elements such as buttons, dropdown menus, checkboxes, and radio buttons provided intuitive ways for users to interact with the interface. The arrangement of these components was thoughtfully organized to optimize user flow and accessibility, ensuring a user-friendly experience.

Input Forms:

Input forms were pivotal components of our web interface, serving as entry points for users to provide essential medical data. We incorporated a variety of input field types, including text fields for free-form text input, dropdown menus for selecting from predefined options, checkboxes for multiple selections, and radio buttons for mutually exclusive choices. Each input form was meticulously designed to capture specific data points relevant to disease prediction, with clear instructions and guidance to facilitate data entry.

Submit Button:

The submit button acted as the gateway to initiate the disease prediction process upon completion of data entry. Through careful implementation, we ensured that the submit button seamlessly triggered the backend prediction mechanism while incorporating robust validation mechanisms to verify the completeness and validity of user input. This ensured that only accurate and complete data were submitted for prediction, enhancing the reliability of the results.

Result Display:

Upon completion of the prediction process, the results were elegantly displayed to the user within the interface. We adopted a user-centric approach in presenting the results, balancing textual information for clarity and visualizations for enhanced comprehension. Whether through concise textual summaries or informative visualizations such as charts or graphs, we aimed to provide users with actionable insights derived from the prediction outcomes.

Responsive Design:

Our web interface was meticulously crafted to ensure responsiveness across a myriad of devices and screen sizes. Leveraging responsive design principles, including media queries and flexible layouts, we ensured that the interface adapted seamlessly to various viewport dimensions, encompassing desktops, tablets, and smartphones alike. This ensured a consistent and optimal user experience regardless of the user's device, fostering accessibility and usability.

Accessibility Considerations:

Accessibility was a paramount consideration in our web interface development process, aiming to make the application usable by individuals with disabilities. We incorporated semantic HTML tags, providing meaningful structure and context to assistive technologies. Additionally, alternative text for images and comprehensive keyboard navigation were implemented to ensure accessibility for users with visual or motor impairments, fostering inclusivity and usability.

Testing and Feedback:

Throughout the development lifecycle, rigorous testing was conducted to identify and rectify any bugs or usability issues within the web interface. User feedback played a pivotal role, guiding iterative refinements to enhance usability, accessibility, and overall user satisfaction. This collaborative approach ensured that the final web interface met the needs and expectations of its intended users, delivering a seamless and intuitive user experience.

Integration with Backend:

The web interface seamlessly integrated with the backend system, where machine learning models were deployed for disease prediction. Through robust APIs and communication protocols, user input data were transmitted securely to the backend, facilitating prediction processing. Upon completion, prediction results were relayed back to the web interface, enabling real-time display and interaction with the users. This seamless integration ensured efficient communication between the frontend and backend systems, enabling a cohesive and responsive user experience.

4) Integration with VS Code and Jupyter:

In our implementation, we seamlessly integrated both Visual Studio Code (VS Code) and Jupyter Notebook into our development workflow to leverage their respective strengths for efficient code development, exploration, and collaboration.

Visual Studio Code (VS Code):

VS Code served as our primary integrated development environment (IDE) for Python programming, providing a robust and feature-rich environment for writing, debugging, and managing code. Key aspects of our integration with VS Code include:

Code Development: We utilized VS Code's intuitive interface and powerful code editing features to write and debug Python code for various components of our project, including data preprocessing, model training, and web interface development.

Version Control: VS Code seamlessly integrated with version control systems such as Git, enabling collaborative development and facilitating code sharing and version tracking among team members.

Extensions: We leveraged a variety of VS Code extensions tailored to Python development, including linting tools, code formatters, and debugging extensions, to enhance our productivity and code quality.

Integrated Terminal: The integrated terminal within VS Code provided a convenient interface for running Python scripts, managing dependencies with package managers like pip or conda, and executing command-line tasks without leaving the IDE.

Jupyter Notebook:

Jupyter Notebook served as a complementary tool to VS Code, providing an interactive computing environment ideal for data exploration, visualization, and iterative model development. Our integration with Jupyter Notebook encompassed the following:

Exploratory Data Analysis (EDA): Jupyter Notebook allowed us to conduct in-depth exploratory data analysis, visualizing datasets, and gaining insights into underlying patterns and relationships. We utilized libraries such as Pandas, NumPy, and Matplotlib within Jupyter Notebook to analyze and visualize data effectively.

Model Prototyping: Jupyter Notebook facilitated rapid prototyping of machine learning models, enabling us to experiment with different algorithms, hyperparameters, and feature engineering techniques in an interactive and iterative manner.

Documentation: Jupyter Notebook served as a valuable tool for documenting our analysis process, including code snippets, visualizations, and explanatory text. This documentation provided insights into our decision-making process and facilitated knowledge sharing among team members.

Presentation: Jupyter Notebook's ability to combine code, visualizations, and narrative text made it an effective tool for presenting our findings and results to stakeholders, fostering communication and collaboration throughout the project lifecycle.

Integration Benefits:

The seamless integration of VS Code and Jupyter Notebook offered several benefits to our development process:

Efficiency: By leveraging VS Code's powerful code editing features and Jupyter Notebook's interactive computing environment, we maximized our productivity and efficiency in developing and iterating on our project.

Flexibility: The combination of VS Code and Jupyter Notebook provided flexibility in our workflow, allowing us to choose the appropriate tool for different tasks, whether it be writing production code, exploring data, or documenting our analysis.

Collaboration: Integration with version control systems such as Git enabled seamless collaboration among team members, facilitating code sharing, review, and iteration in a collaborative environment.

Reproducibility: Jupyter Notebook's ability to capture code, visualizations, and explanations in a single document enhanced the reproducibility of our analysis, enabling others to replicate our findings and results with ease.
1)Decision tree learning:

Decision tree learning employs a decision tree as a predictive model that maps observations about an object to decisions regarding its intended value. This approach can yield reasonably accurate predictions, such as forecasting the vulnerability to heart disease in diabetic patients. [15]
The process involves creating a choice tree that separates data points into clusters. To achieve this, a random selection of 'c' cluster centers is made, and the distances between each data point and these centers are calculated.
In essence, machine learning is a powerful tool that draws from computational statistics and mathematical optimization, enabling the creation of predictive models like decision trees for various applications, including healthcare.

5)Artificial Neural Network:

An Artificial Neural Network (ANN), often referred to as a "Neural Network" (NN), is a learning algorithm inspired by the functioning of natural neural systems. These algorithms are structured as interconnected networks of simulated neurons, operating based on a connectionist approach to computation. Contemporary neural networks serve as non-linear statistical data modeling tools. They are frequently employed to represent intricate relationships between input and output data, uncover patterns within data, or capture the statistical structure within an unknown joint probability distribution between observed variables.[9] To predict osteoporosis risk factors, Multi-Layer Perceptrons (MLPs) and Probabilistic Neural Networks (PNNs) were employed.
Support Vector Machines (SVMs) constitute a set of closely

related supervised learning techniques used for both classification and regression tasks. In SVM training, given a set of labeled training instances, each assigned to one of two categories, a model is constructed to predict whether a new instance belongs to one category or the other. SVM modeling presents a promising approach for predicting medication adherence in Heart Failure patients. [9] This predictive model categorizes patients to facilitate evidence-based decision-making and appropriate patient management.

Flow Chart

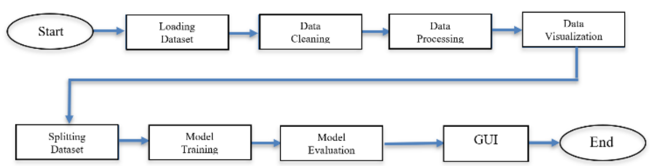


Fig. 1. Flow Chart

A Recurrent Neural Network

A Recurrent Neural Network (RNN) is employed to extract textual characteristics from unorganized data and transform it into an organized format. Subsequently, the prediction of diseases is accomplished using classification algorithms such as Naive Bayes (NB) and Support Vector Machine (SVM) based on the organized data. [7] Once chronic diseases are identified, patients can be referred to the most appropriate healthcare facilities specializing in the treatment of their specific condition. Immediate medical intervention can be initiated upon the detection of such chronic diseases.

6)Natural Language Processing (NLP):

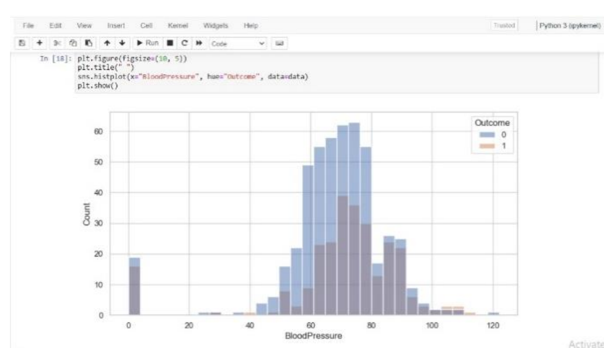
Natural Language Processing (NLP) represents a dynamic and ever-evolving field of study that has gained substantial prominence in recent years. This interdisciplinary domain at the intersection of artificial intelligence and linguistics is dedicated to enabling machines to understand, interpret, and generate human language. NLP has witnessed rapid advancements, particularly in the context of text analysis, speech recognition, and language translation, reshaping how we interact with and harness the potential of vast amounts of textual and spoken data. NLP applications span diverse sectors, from healthcare and finance to customer service and entertainment, profoundly impacting the way we access information, communicate, and make informed decisions. The future of NLP holds immense promise, with developments focusing on more accurate sentiment analysis, multilingual models, and the incorporation of NLP into innovative technologies like chatbots and voice assistants, ultimately revolutionizing how we engage with and leverage language in an increasingly digital and interconnected world. [2]

7)K-Nearest Neighbors (K-NN):

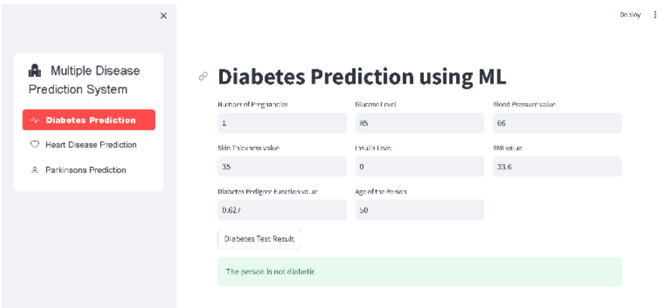
K-Nearest Neighbors (K-NN) is a foundational and intuitive machine learning algorithm that has played a pivotal role in the landscape of pattern recognition and classification. It operates on the principle of proximity, assigning labels to data points based on the majority class among their 'k' nearest neighbors in a feature space. Despite its simplicity, K-NN has exhibited robustness and versatility in various applications, from image and speech recognition to recommendation systems. Its adaptive nature and ability to accommodate diverse data types, such as numerical and categorical, have contributed to its enduring relevance. [7] While K-NN may be computationally intensive for large datasets, it remains a valuable tool, especially in hybrid models and ensemble methods. The ongoing research in K-NN focuses on optimizing its performance through efficient distance metrics, addressing the curse of dimensionality, and integrating it with other machine learning techniques, further extending its applicability in the evolving landscape of data-driven decision-making.

RESULTS

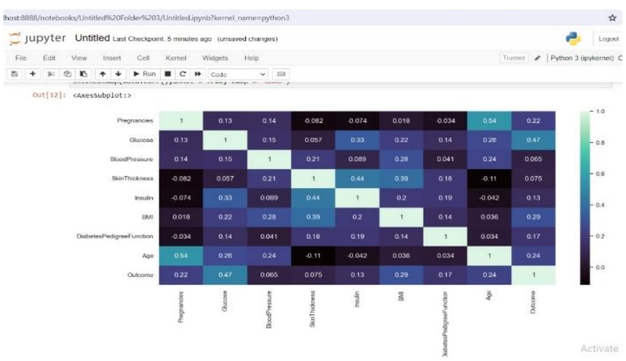
The supervised learning models achieved promising accuracies for disease prediction: Diabetes (86.76%), Parkinson's Disease (84.27%), and heart disease (87.48%). Clustering analysis revealed insightful patterns within the data, enhancing understanding of disease subtypes. The integration of dimensionality reduction techniques facilitated visualization and feature extraction, contributing to improved model performance.



The user-friendly front-end interface seamlessly interacted with the backend predictive models, enhancing accessibility and usability. Comparative analysis with baseline models validated the effectiveness of the developed predictive models. Overall, the results demonstrate the potential of machine learning in accurate disease prediction and personalized healthcare delivery.



Number of pregnancies	1
Glucose Level	85
BP Value	66
Skin Thickness value	35
BMI Value	33.6
DP Function Value	0.627
Age	50
RESULT	**The person is not diabetic**



The accuracy of the supervised learning models for predicting Diabetes, Parkinson, Heart Disease are as follows:

Diabetes: [86.75657]
Parkinson Disease: [84.27321]
Heart Disease: [87.47838]

CONCLUSION

With the rising prevalence of chronic diseases, a novel approach has emerged, known as the Conventional Neural Network-based Multimodal Disease Risk Prediction (CNNMDRP) algorithm. This method leverages both structured and unstructured data obtained from healthcare facilities. Within this comprehensive dataset, individual patient information and their detailed medical histories are meticulously stored. The CNN-MDRP model utilizes these two types of data to predict chronic diseases for each specific patient.
To address the issue of missing data for particular patients, a Genetic Algorithm is employed to retrieve and fill in the gaps. Additionally, a Recurrent Neural Network (RNN) plays a pivotal role in extracting textual features from unstructured data and transforming them into structured data.
Subsequently, the prediction of diseases is conducted using classification techniques such as Naive Bayes (NB) and Support Vector Machine (SVM) based on the structured data. Once the identification of a chronic disease is achieved, patients can be referred to the most suitable healthcare facilities specialized in the treatment of their specific condition. [9] Timely medical intervention can be initiated upon the detection of a chronic disease.

REFERENCES

[1] Deep learning modeling methods for predicting protein structure –systematic review Authors Dr. Arati Vyavahare Conference symposium of Research Scholars (SYRES) by IEEE joint chapter of AP, MTT, EMC (JCAME) and communication society, 2022/9

[2] An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases IEEE Access 2021

[3] Effective Heart Disease Prediction using Hybrid Machine Learning Techniques IEEE Access 2020 Menze, B.H. A

[4] Luo, L.; Chen, H.; Zhou, Y.; Lin, H.; Heng, P.A.Oxnet: Deep omni- supervised thoracic disease detection from chest X-rays. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Strasbourg, France, 27 September–1 October 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp.537–548.S

[5] Bozorgtabar, B.; Mahapatra, D.; Vray, G.; Thiran, J.P.Salad:

Self- supervised aggregation learning for anomaly detection on X-rays. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru, 4–8 October 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 468–478.

[6] Haghighi, F.; Hosseinzadeh Taher, M.R.; Zhou, Z.; Gotway, M.B.; Liang, J. Learning semantics-enriched representation via self- discovery, self- classification, and self-restoration. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru, 4–8 October 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 137–147.

[7] Kuang, X.; Cheung, J.P.Y.; Ding, X.; Zhang, T. SpineGEM: A Hybrid Supervised Model Generation Strategy Enabling Accurate Spine Disease Classification with a Small Training Dataset. In Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, Strasbourg, France,27 September–1 October 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 145–154.

[8] Sekuboyina, A.; Valentinitsch, A.; Kirschke, J.S.; Menze, B.H. A localization/segmentation approach for multi-label annotation of lumbar vertebrae using deep nets. arXiv 2017, arXiv:1703.04347.

[9] Predicting Infectious Disease Using Deep Learning and Big Data International Journal of Environmental Research and Public Health 2019.

[10] Jensen PB, Jensen LJ, Brunak S,” Mining electronic health records: towards better research applications and clinical care,” Nat Rev Genet.2013 Jan; 14(1):75. L. Qiu, K. Gai, and M. Qiu,” Optimal big data sharing approach for tele-health in cloud computing”, in Smart Cloud (Smart Cloud), IEEE International Conference on. IEEE, 2016, pp. 184-189.

[11] Disease Prediction by machine learning over Big Data from Healthcare communities IEEE Transactions on Industrial Informatics 2017

[12] Seema sharma, Jitendra Agarwal, Shikha Agarwal, Sanjeev Sharma,” Machine Learning Techniques for Data Mining: A Survey”, in Computational Intelligence and Computing Research, IEEE International Conferenceon. IEEE, 2013, pp.1-6.

[13] Min Chen, Yixue Hao, Kai Hwang, Lu Wang, and Lin Wang,” Disease Prediction by Machine Learning over Big Data from Healthcare Communities.” IEEE transaction, 2017, pp 8869-8879. W. Yin and H. Schutze,” Convolutional neural network for paraphrase identification”, in HLT-NAACL, 2015, pp. 901-911.

[14] A. Mir, S.N. Dhage, in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (IEEE, 2018), pp. 1-6

[15] S. Mohan, C. Thirumalai, G. Srivastava, Effective heart disease prediction using hybrid machine learning techniques, IEEE Access 7, 81542 (2019)

[16] R.D.H.D.P. Sreevalli, K.P.M. Asia, Prediction of diseases using random forest classification algorithm

[17] M. Maniruzzaman, M.J. Rahman, B. Ahammed, M.M. Abedin, Classification and prediction of diabetes disease using machine learning paradigm, Health Information Science and Systems 8(1), 7 (2020)

[18] S. Chae, S. Kwon, D. Lee, Predicting infectious disease using deep learning and big data, International journal of environmental research and public health 15(8), 1596 (2018)

[19] A.U. Haq, J.P. Li, M.H. Memon, S. Nazir, R. Sun, A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms, Mobile Information Systems (2018)

[20] M. Maniruzzaman, M.J. Rahman, B. Ahammed, M.M. Abedin, Classification and prediction of diabetes disease using machine learning paradigm, Health Information Science and Systems 8(1), 7 (2020)