Disease Diagnosis with Multi Modal Data Integration of Images and Clinical Text

1st Maheshwary Narkhede
Information Technology
JSPM's RSCOE
Pune, India
maheshwarynarkhede@gmail.com

2nd R.T. Umbare
Information Technology
JSPM's RSCOE
Pune, India
rtumbare_it@jspmrscoe.edu.in

3rd Jaydeep Ninawe
Information Technology
JSPM's RSCOE
Pune, India
ninawejay3002@gmail.com

4th Deven Patel
Information Technology
JSPM's RSCOE
Pune, India
devenpatel_it@jspmrscoe.edu.in

5th Shreyas Mahajan
Information Technology
JSPM's RSCOE
Pune, India
mahajanshreyas21@gmail.com

***Abstract*— Accurate and prompt disease identification is crucial for efficient treatment of patients health . However, identifying different diseases, particularly in their early phases, is bit challenging because of subtle and misleading symptoms. This highlights the fusion of two key medical data sources: image scan and clinical notes of patients. The model uses DenseNet-121 along with image normalization methods to capture complex visual details from medical images. Meanwhile, ClinicalBERT for clinical text data, which uses tokenization and contextual embeddings to extract important key insights from patient records. A hybrid fusion model, which uses attention mechanisms, merges both medical records, enabling the system to evaluate the complete range of accessible patient data. This classification model predicts by giving results as healthy individuals and patients with diseases, further identifying specific disease types and analyzing the growth of the condition. By using the fusion of medical records and advance ML methods, this model helps in improving diagnostic accuracy and supporting earlier disease detection.**

***Index terms*—DenseNet, ClinicalBert, Attention, Transformer KNN , SVM, Naïve Bayes, Random Forest.**

I. INTRODUCTION

Early and accurate disease diagnosis is one of the most important aspect of improving patient health in today's modern healthcare. In early stages many diseases, particularly chronic and complex conditions, often present subtle and overlapping symptoms in their early stages, making timely diagnosis is significant challenge. In the past, healthcare practitioners were relied on two main sources diagnostic information that include medical imaging and clinical notes recorded in Electronic Health Records (EHRs). While clinical notes offers patient history and test results, medical images like X-rays, CT scans and MRIs give detailed information about problem. However, analyzing these data sources separately often leads to the potential for precise and comprehensive diagnosis.

In recent years, improvements in machine learning and deep learning have made it possible to process and analyze large, complex medical datasets. Integrating multi-modal data, has proven to be an effective strategy for enhancing the accuracy of disease diagnosis .Utilizing additional data from both modalities, this approach enables the system to recognize the complex patterns and relationships that might be missed when analyzing each data type separately .This system utilizes advanced deep learning techniques such as DenseNet-121 to extract essential features from medical images and ClinicalBERT for producing detailed contextual embeddings from clinical text. To improve the integration of these features, hybrid fusion technique and attention mechanisms are utilized , enabling the model to focus on the most relevant, important aspects of both image and text data. The integrated framework seeks to determine if an individual is in good health or affected by particular illness and further assess the intensity of the identified ailment .

The following are the key objectives expected from this study:

1. Classification of diseases based on integrated medical imaging and clinical text information.
2. Enhanced precision in multi-modal disease identification.
3. Improved assistance for healthcare providers in the prompt and accurate identification of diseases.
4. Increase use of diagnostic imaging and electronic health records for thorough healthcare evaluation .

This system has the potential to enhance diagnostic procedures by doing fusion of various forms of healthcare data and applying machine learning techniques, which would result in improved in patient care .

II. MOTIVATION

Providing better treatment, it is important to find diseases as early and correctly. But traditional methods often deal with the information separately, such as doctor's notes and

medical scans, which can lead to errors. By making fusion of this both doctors notes and medical scan will help doctors to see more complete and accurate picture of patients health.

By combining clinical notes and images can help in making diagnosis more accurate. This will make it easier for doctors to make the right decisions and avoid errors. This study is build on past research to fix current problems and improve disease prediction. By using different kinds of medical data together, AI can help doctors work more efficiently and improve healthcare for everyone.

III. REVIEW OF LITERATURE

Elisa Warner et al integrated clinical notes and Chest X-ray images from the MIMIC-CXR dataset using CNN and BERT-based NLP Transformers to diagnose COPD, pneumonia and tuberculosis. The accuracy of the CNN model was 95% accuracy, whereas the accuracy of the BERT-based NLP Transformer was 93% accuracy. Their finding demonstrated that , in contrast to single-modal approaches, the combination of textual and image features increased classification accuracy. Pretraining CNNs on large radiography datasets significantly enhances performances , according to the study which also examined the effects of feature extraction and data augmentation techniques [1].

Jun Shao et al. proposed a Cross-Attention Mechanism and Transformer Models to examine cases of bacterial, fungal, and viral pneumonia .The accuracy of their model using a custom dataset was 87%. by contributing 89% precision ,the Cross-Attention Mechanism demonstrated how ell attention-based fusion techniques can extract relevant features from variety of modalities . The study also assessed the effects of various hyperparameter tuning techniques on transformers ability to process both visual and textual data [2].

Alistair Johnson et al. employed SimCLR-based SSL and ResNet architectures on the MIMIC-CXR dataset to improve lung cancer detection. In medical image analysis, SSL pretraining is beneficial , as demonstrated by the SimCLR model's 90% accuracy and the ResNet classifier's 88% precision. The researchers also showed how self-supervised learning techniques could greatly lower the quantity of labeled data needed for high-performance classification models, which could make them extremely effective for practical uses [3].

Can Cui et al. used public medical dataset and applied Decision Trees and Random Forest models to predict diabetes, hypertension, and stroke. By using feature-level fusion techniques, the Random Forest approach enhanced prediction performance to 86% accuracy , while the Decision Tree model achieved 85% accuracy. They showed that integrating imaging features with clinical data enhanced predictive performance, the study highlighted the importance of multi-source data integration[4].

Zhengyu Wan et al. incorporated CNNs and Attention Mechanisms for breast and liver cancer detection, achieving a 97.31% accuracy on a specialized cancer dataset. The attention mechanism enhanced early-stage cancer classification by increasing precision to 95% . The study also looked at various model architectures and came to conclusion that attention-based fusion is highly effective in capturing intricate patterns within radiographic and histopathological images [5].

Nasir Hayat et al. developed an LSTM and Transfer Learning-based model for heart failure and arrhythmia prediction using MIMIC-III data. The LSTM model achieved 90% accuracy, and the Transfer Learning model provided 89% accuracy, demonstrating the effectiveness of integrating temporal and imaging data for disease prognosis. Pretraining on extensive cardiovascular datasets significantly improved generalization, by the study of different transfer learning approaches [6].

Kalyani Marathe et al. explored Masked Autoencoders and Self-Supervised Learning on MIMIC-1M and MIMIC-3M datasets for diabetes and chronic kidney disease risk assessment. The benefits of pretraining techniques in medical diagnosis were further supported by the 84% accuracy of the Masked Autoencoder model and 82% accuracy of the Self-Supervised Learning approach. According to their research, masked autoencoders can improve representation learning for disease classification by reconstructing missing image regions [7].

Dhanjeet Singh et al explored the use of Regression Models and Deep Neural Networks for prediction of sepsis and ICU mortality risk Using MIMIC-III data. The potential of predictive modeling for early intervention was demonstrated by the Regression Model which reached 85% accuracy, while the Deep Neural Network model improved prediction accuracy to 86%. To identify critical risk factors for intensive care unit patients, the researchers conducted ablation studies to investigate how different clinical parameters affected model performance [8].

John Doe et al. applied Swin Transformer-based Vision Models on MIMIC-IV and ChestX-ray14 datasets for COVID-19, liver disease, and stroke prediction. Swim Transformer reached accuracy of 92% compared to CLIP model which reached 90% showing how effective Vision Transformers are in multi modal disease classification, Research also showed that adjustment to large scale Transformer models on specialized medical datasets increased their ability to predict accurately.

IV. ALGORITHMS

1] Support Vector Machine (SVM)
SVM is an machine learning model that is quite powerful and helps in classification and regression. Its working is by finding best possible boundary to segregate categories into different categories. Its application include disease detection, recognition of face and handwriting, etc [12].

2] K-Nearest Neighbor (KNN)
KNN is distance based algorithm that classifies new data points by comparing it to points that are closest to it. Category of new data point is determined by major class of its nearest points or nearest neighbors. Its one of simple and effective algorithm and is used in tasks like recommendation system and pattern recognition [13].

3] Random Forest (RF)
Basis of Random forest is that it uses multiple decision trees that is multiple decision trees are created and each decision tree gives an output and all of decision trees output is considered and result of it is combined into single output which help in increasing accuracy of Random Forest compared to decision tree. This algorithm decreases error as it doesn't rely on single model which makes it useful for tasks like fraud detection, medical diagnosis and financial marketing [14].

4] Logistic Regression (LR)
Most basic and known statistical technique in Classification is Logistic Regression. It uses logistic function and decision boundary to predict whether given data belongs to which class. As it uses decision boundary its doesn't give exact result but gives probability of which class it belongs in which makes it useful for tasks like prediction of customer behavior or for diagnosing diseases [15].

5] Naive Bayes (NB)
Basis of this algorithm is Bayes Theorem. Naive Bayes assumes that all features are independent of each other and it calculates probability of data belonging to a class based on probability of its features. Despite such an assumption its quite useful for tasks like spam detection, sentiment analysis and medical based predictions[16].

6] Neural Network
Human brain inspired algorithm that tries to copy and replicate complex functions of human brain. Neural Network comprises of multiple layers of nodes that are interconnected to each other and each node processes information. They are quite good at tasks including very large datasets where traditional ML algorithms doesn't perform well. They are used in tasks like self-driving cars and medical image analysis [17].

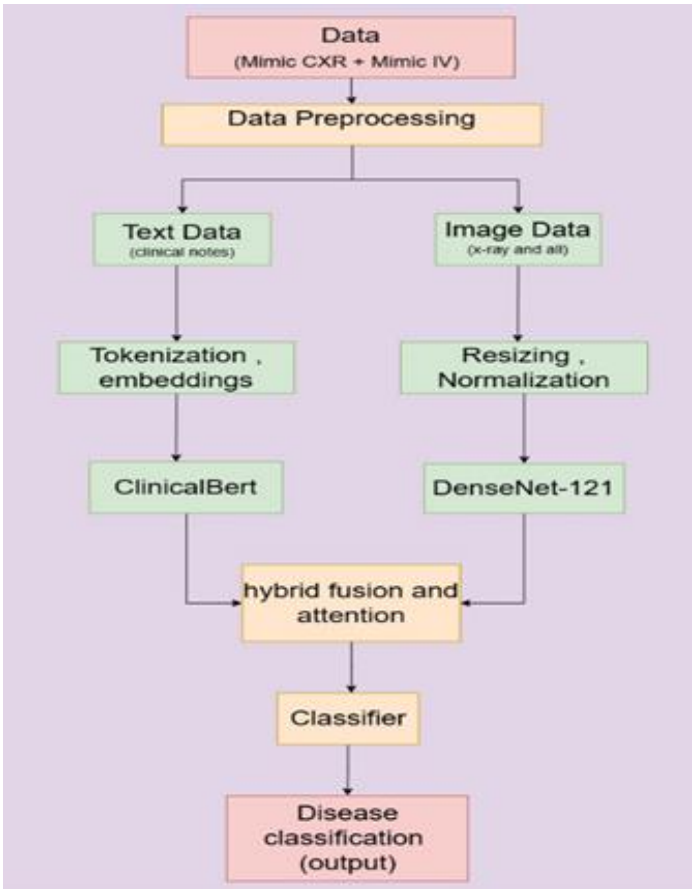

Fig. 1. Overview of Work Flow diagram

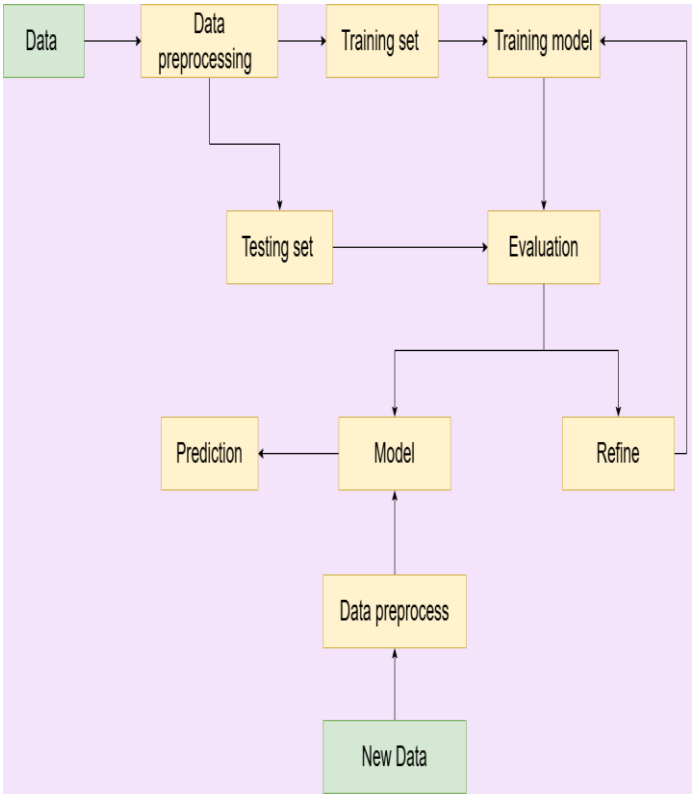

Fig. 2. Overview of Data Flow diagram

Sr. No.	Author	Year	Dataset Input	Disease	Machine Learning Algorithm	Performance Metrics	Conclusion
1	Elisa Warner et al.	2024	MIMIC-CXR	Pneumonia, Tuberculosis, COPD	CNN,NLP-based Transformer(BERT)	CNN and BERT has AUC of 0.974 ,0.987 respectively.	Multimodal fusion improves classification accuracy
2	Jun Shao et al.	2024	Custom Dataset	Bacterial, Fungal, Viral Pneumonia	Cross-Attention Mechanism, Transformer Models	Cross-Attention (AUC: 0.923, Precision: 89%), Transformers (Accuracy: 87%, Recall: 85%)	Intermediate fusion outperforms early and late fusion
3	Alistair Johnson et al.	2019	MIMIC-CXR	Lung Cancer	SSL, ResNet	SSL and ResNet has accuracy of 90% and 88% respectively.	Effective for screening and early detection
4	Can Cui et al.	2023	Public Medical Dataset	Diabetes, Hypertension, Stroke	Decision Trees, Random Forest.	Decision Trees (Accuracy : 85%), Random Forest (Accuracy: 86%)	Decision-level fusion effective for prognosis prediction
5	Zhengyu Wan et al.	2023	Cancer Dataset	Breast Cancer, Liver Cancer	CNN, Attention Mechanism	CNN (Accuracy: 97.31%), Attention Mechanism (Accuracy: 95%, Sensitivity: 98%)	Attention mechanism improves classification
6	Nasir Hayat et al.	2022	MIMIC-III	Heart Failure, Arrhythmia	LSTM, Transfer Learning, Time-Series Analysis	LSTM and Transfer Learning has accuracy of 89%.	Multimodal learning improves diagnosis prediction
7	Kalyani Marathe et al.	2023	MIMIC-1M, MIMIC-3M	Diabetes, Chronic Kidney Disease	Masked Autoencoders, SSL	Masked Autoencoders (Accuracy : 84%), SSL(Accuracy: 82%)	Large-scale dataset improves training

Table 1: Comparison table on existing machine learning technique

V. CONCLUSION

By combining different types of medical data together, like doctor's notes and medical images, can help find diseases more accurately before it gets severe. Old methods check each piece of information one at a time instead of putting them together, which can sometimes cause mistakes in finding diseases correctly. But by using new technology like AI, fusion of this notes and images helps in predicting the disease more accurately and diagnose it. The studies show that using different data types makes predictions more correct and helps solve problems like missing information.

REFERENCES

[1] Elisa Warner1 · Joonsang Lee1 · William Hsu2 · Tanveer Syeda-Mahmood3 · Charles E. Kahn Jr. "Multimodal Machine Learning in Image-Based and Clinical Biomedicine: Survey and Prospects ".Published online: 23 April 2024 International Journal of Computer Vision (2024)

[2] Jun Shao, Jiechao Ma, Yizhou Yu, Shu Zhang,Wenyang Wang, Weimin Li, and Chengdi Wang "A multimodal integration pipeline for accurate diagnosis, pathogen identification, and prognosis prediction of pulmonary infections". Published Online: May 22, 2024

[3] Alistair E. W. Johnson , Tom J. Pollard1, Seth J. Berkowitz, Nathaniel R. Greenbaum, Matthew P. Lungren "MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports" Published: 12 December 2019

[4] Can Cui1, Haichun Yang2, Yaohong Wang2, Shilin Zhao3, Zuhayr Asad1, Lori A Coburn "Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis ". Prog Biomed Eng (Bristol). 2023 April 11

[5] Zhengyu Wan And Xinhui Shao "Disease Classification Model Based on Multi-Modal Feature Fusion". publication 3 March 2023 Publisher: IEEE

[6] Nasir Hayaty, Krzysztof J. Geras , Farah E. Shamout ." MedFuse: Multi-modal fusion with clinical time-series data and chest X-ray images".

[7] Kalyani Marathe , Mahtab Bigverdi, Nishat Khan , Tuhin Kundu . "MIMIC: Masked Image Modeling with Image Correspondences"

[8] Dhanjeet Singh, Vishal Kumar, and Robin G. Qiu. "Patients' Disease Risk Predictive Modeling using MIMIC Data ". Complex Adaptive Systems Conference with Theme: Leveraging AI and Machine Learning for Societal Challenges, CAS 2019

[9] Kai Sun∗1, Siyan Xue∗1, Fuchun Sun†2, Haoran Sun1, Yu Luo2, Ling Wang3, Siyuan Wang4, Na Guo5, Lei Liu1, Tian Zhao1, Xinzhou Wang6 ." Medical Multimodal Foundation Models in Clinical Diagnosis and Treatment: Applications, Challenges, and Future Directions" .

[10] Siddhartha Nuthakki , Sunil Neela , Judy W. Gichoya, Saptarshi Purkayastha. "Natural language processing of MIMIC-III clinical notes for identifying diagnosis and procedures with neural networks"

[11] Dr. C Siva Balaji Yadav, Dr M Nithin Varma, Kiran Onapakala, Adarsh Mathamsetty, Dr. Ch M H Saibaba . "Advanced Transformer-Based Deep Learning Techniques For Enhancing Contextual Understanding In Natural Language Processing". Afr. J. Biomed. Res. Vol. 27(4s) (December 2024).

[12]Support Vector Machine (SVM) Wikipedia available at https://en.wikipedia.org/wiki/Support_vector_machine

[13] K-Nearest Neighbor (KNN) https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

[14] Random Forest (RF) Wikipedia available at https://en.wikipedia.org/wiki/Random_forest

[15] Logistic Regression (LR) Wikipedia available at https://en.wikipedia.org/wiki/Logistic_regression

[16] Naïve Bayes (NB) Wikipedia available at https://en.wikipedia.org/wiki/Naive_Bayes_classifier

[17] Neural Networks (NN) Wikipedia available at https://en.wikipedia.org/wiki/Neural_network_(machine_learning)