Journal of Computer Science

Research Article

An Integrated Disease Progression Model to Analyze Electronic Health Records With Multimodal Datasets Using Deep Learning Techniques

Thulaganyo Dimakatso, Venumadhav Kuthadi, Rajalakshmi Selvaraj and Othapile Dinakenyane

Department of Computing and Informatics, Botswana International University of Science and Technology, Botswana

Article history
Received: 20-03-2025
Revised: 08-07-2025
Accepted: 29-07-2025

Corresponding Author:
Thulaganyo Dimakatso
Department of Computing and Informatics, Botswana International University of Science and Technology, Botswana
Email: dt23018978@biust.ac.bw

Abstract: In contemporary clinical environments, precisely simulating the emergence of disease is an intricate task owing to heterogeneously derived data and variability in the record of diagnoses. The research presents a stable classification model that integrates image data with formalized patient data to track patterns in disease through various stages. The suggested approach combines a multi-stage analytical methodology that includes systematic data preparation, transfer-based feature learning with a purpose-tuned InceptionResNetV2 architecture, and performance metrics evaluation under stringent criteria. Substantially, the system has been augmented by incorporating a fusion approach in which diagnostic images are combined with patient records, leading to enhanced classification validity. With a general accuracy level of 97.45%, the model indicates good generalizability and interpretability. Its use of domain-specific tuning and interpretive tools increases its applicability to real-world medical diagnosis, offering a sound solution for dealing with class imbalance and heterogeneous disease presentations. These figures indicate the model's best performance in dealing with data imbalance and misclassifications, typical in medical imaging and Electronic Health Record (EHR) analysis. Data imbalance, whereby certain disease categories are underrepresented, is likely to lead to skewed predictions and false generalizations. The findings point to the need for developing diagnostic tools that would be applicable in multimodal data integration.

Keywords: Deep Learning, Disease Progression Modelling, InceptionResNetV2, Medical Image Classification, Transfer Learning

Introduction

Disease progression refers to the transition of a medical condition through various stages, typically from early onset to advanced severity, as observed through clinical symptoms and imaging changes. In this study, progression is modelled by analyzing sequential data patterns derived from imaging modalities and patient records. Understanding this trajectory aids in forecasting clinical outcomes and enhancing decision-making. The application of medical technologies has transformed healthcare systems by assisting in the diagnosis of various diseases. Nevertheless, the capacity to categorize all forms of diseases (Wan and Shao, 2023) is still challenging. The intricacy of imaging data among patients further complicates the challenge of diagnosing diseases among patients. To address these issues, this research employs deep learning techniques to label medical images of diseases in some patients. Existing medical imaging technologies like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) are subject to some drawbacks (Tian et al., 2024). Most of the models are plagued by low scalability and low adaptability to individual conditions of patients. Also, these models tend to provide wrong diagnoses in the case of low-resolution images. Current models are often incapable of generalizing various medical datasets, thus limiting their practical application. Recent integration of multi-modal data and Transfer Learning has vastly enhanced disease diagnosis, particularly how they propagate in the human body. Many issues are still present, such as data heterogeneity, how scalable they are, and how to extract them.

This study is therefore intended to overcome these limitations by adopting a structured approach that

involves pre-processing, optimizing the transfer of learning strategies, and multimodal feature fusion. Traditional transfer learning models often use a pre-determined architecture without sufficiently adapting it to medical datasets, which limits their applicability. The model uses the InceptionResNetV2 architecture and fine-tunes it specifically for multimodal medical imaging data. By freezing initial layers and retraining higher layers with domain-specific images, the framework ensures:

- An improved generalisation to medical imaging datasets
- More effective feature extraction from high-dimensional data
- Reduction of over-fitting through controlled layer adaptation

Additionally, data augmentation techniques such as the adjustment of brightness, rotation, and flipping enhance the robustness of the model, which ensures that there is a better adaptation to variations in medical imaging. One of the most serious issues in multimodal data processing is the natural imbalance in medical datasets, in which some phases of diseases can be underrepresented.

The framework employs:

- Synthetic Minority Over-sampling Technique (SMOTE) to obtain a balanced class representation.
- Standardisation and normalization to promote consistency across modalities
- Cross-modal feature alignment to synthesize information optimally from various imaging modalities (e.g., CT, MRI, and X-rays)

By paying attention to these factors, the model becomes more accurate and ensures that all the diseases are equally represented. Existing multimodal deep learning models often struggle with scalability due to high computational demands. The framework optimizes the efficiency of computation through:

- Batch processing is required for large-scale data ingestion
- Dynamic learning rate scheduling to prevent unnecessary computations and accelerate convergence
- Parallel processing capabilities that allow a simultaneous analysis of multiple imaging modalities

These optimizations make the framework suitable for use in healthcare applications, resulting in the reduction of inference time. Many deep learning models operate as "black boxes," thereby limiting their interpretability. The proposed framework incorporates:

- Gradient-weighted Class Activation Mapping (Grad-CAM) for decision-making processes
- Class-wise performance evaluation to highlight specific strengths and weaknesses
- Confusion matrix analysis to understand misclassification trends

These features provide clinicians with greater insight into the efficiency of the model, which fosters trust in medical practice. Furthermore, traditional algorithms are not graded thoroughly to handle multi-modal data integration (Yang et al., 2024), which is crucial for designing a prototype for disease progression. These gaps necessitate the development of an advanced system capable of improving diagnostic precision in healthcare services.

The main objective of this study is to design an innovative classification system and patient-specific disease progression modelling. Using the power of deep learning, the proposed framework will incorporate advanced pre-processing techniques that rely on transfer learning and optimized training strategies. The objective of the study is to ensure that there is accuracy in the classification process, which minimizes improper classifications, develops a scalable solution that can adapt to diverse medical datasets and patient profiles, and provides interpretable results that help healthcare professionals make reliable decisions. Consequently, this study focuses on using the Inception ResNetV2 architecture for the extraction of features and their classification. By pre-processing pipelines, the framework makes medical images suitable for application. It also solves the issue of the imbalance of data and noise in imaging datasets.

Review of Literature

Based on existing frameworks of Electronic Health Records (EHR), the literature review for this study focuses on the healthcare sector, which is increasingly shifting towards the use of Artificial Intelligence (AI) and, at the same time, incorporating traditional analytic workflows. The application of AI, nevertheless, is confronted by several challenges: The most prominent among them is its inability to be applied to different domains. These disadvantages include a deficiency of data, which can result in inconsistency in practice. Besides, Electronic Health Record (EHR) systems have data biases and discrepancies that need to be handled with expertise. As a reaction to such issues, the EHR-ML framework pronounced by Ramakrishnaiah et al. (2025) presents an organized machine learning strategy to the health sector. EHR- ML shortens model architecture and

parameter selection by automating the data ingestion as well as harmonization across institutions. With the use of case studies, the authors detail how EHR-ML goes beyond traditional measures to address many medical problems.

This approach emphasizes the potential of AI to enhance predictive validity and overall quality of health care and addresses typical obstacles to its implementation. To help advance the predictive validity of machine learning models, Hancox et al. (2024) investigate the application of graph theory. In their systematic review, they examine the possible value of graph representations of EHR data in an effort to enhance disease diagnosis. They identified 27 studies to investigate graph-based models that predict different health outcomes, like hospital readmission and death. Compared to State-of-the-Art techniques, the authors discovered that graph representations significantly improve the prediction of machine learning models, but more studies have to be conducted before the methods can be utilized in clinical environments.

In order to enhance understanding in this research, there is an emphasis on conducting more research on how graph-based models can help ensure the accuracy of disease interpretation that can redefine the significance of electronic health records. Cardiovascular diseases have been studied to identify them at an earlier stage by Pamulaparthyvenkata et al. (2024). Their research emphasizes grave concerns, such as employing machines for disease interpretation. Using both local and global methods, the authors propose implementing an Entropy Hidden Markov Model (EHMM) to detect heart diseases. The result, with 0.98 prediction accuracy, indicated that EHMM is more effective than Support Vector Machine (SVM) and Random Forest (RF). The technique enhances personalized medicine through a clearer and more reliable methodology for the early detection of heart diseases.

To diagnose disease from EHR data, Tian et al. (2024) suggest a new hybrid model based on the integration of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The hybrid model uses LSTMs to extract long-term dependencies and the capacity of CNNs to learn hierarchical features of complex data. According to their empirical work, this technique outperforms the standard machine learning methods like Support Vector Machines (SVM) and individual CNN and LSTM models. Deep learning holds the potential to transform the science of accurate disease diagnosis using novel neural network models. To predict the clinical outcomes, Wang et al. (2025) propose a multi-step Feature Selection (FS) process that integrates knowledge-based expert methods with data-driven statistical methods. In predicting Acute Kidney Injury (AKI), the model was validated in two independent cohorts of the MIMIC-III and MIMIC-IV-ED databases. Based on their analysis, the researchers learned that their FS model enhances their predictive abilities via various machine learning methods.

The model enhances the capacity to conduct disease diagnosis without any loss of efficiency. The approach enhances clinical decision-making with the triumph of reducing dimensionality in Electronic Medical Records (EMRs). Fallahpour et al. (2024) argue that the challenges in incorporating transformer-based models with Electronic Health Records (EHRs are because they are very costly and have a short lifespan. They overcame these challenges by presenting EHR Mamba, a robust foundation model that significantly improves longer EHR sequence processing and is built on the Mamba architecture. Cross-task generalization and deployment efficiency are improved by the model's ability to perform multitask learning with a single fine-tuning step. EHR Mamba's application in real-world healthcare is made simpler by its compatibility with the HL7 FHIR standard, which makes it easy to use in hospitals.

According to Fallahpour et al. (2024), EHR significantly improves AI use in medicine because it improves on past systems in clinical workflows. Nasarudin et al. (2024) explain in their critique how online medical databases and deep learning models can be utilized over EHR data to determine diseases. By so doing, the healthcare industry. This review presents other researchers with a choice of models to implement for the purpose of creating deep learning models that are specifically customized to identify disease. Niu et al. (2024) present a model named EHR-BERT, an anomaly identification model using the BERT architecture. The framework solves issues by means of Sequential Masked Token Prediction (SMTP) for improved anomaly detection capabilities. Most tests with huge EHR datasets from various medical facilities have proven that EHR-BERT's performance is improved as compared to the conventional approach by eliminating incorrect information and enhancing abnormal rate identification. This breakthrough enables EHR-BERT to be a significant tool for improving medical data accuracy, which is liable for reducing medical errors.

Cui et al. (2024) examine employing Electronic Health Records (EHRs) for illness diagnosis. Their study explores the potential to transform structured patients' data into natural language stories using LLMs. The authors suggest a novel method of pairing a predictor agent to determine a disease. Researchers' results show that LLMs can diagnose illnesses from EHR data with a learning process similar to traditionally supervised learning methods. The process opens new avenues for applying LLMs in medicine, particularly when there are limited labelled datasets. Heumos et al. (2024) propose that 'therapy' remains to be discovered in epidemiology

studies. From quality control and data extraction to causal inference, survival, and therapy, it supports a wide variety of analytical tasks. Ontologies are integrated into the system to allow data sharing. Heumos et al. (2024) describe in their case studies how therapy might be applied to the detection of EHR data biases in the setting of disease phenotypes. Such an open-source strategy is of utmost urgency to data analysis in the healthcare sector as well as in biomedical studies.

Methods

This research employs a multimodal deep learning framework for representing patient-specific disease as an ensemble of advanced imaging modalities and machine learning techniques. This research utilizes the data from a highly curated dataset of medical imaging consisting of several different modalities, including Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and X-ray images. The data is structured to be compatible with deep learning models by classifying images according to various stages of disease. Metadata, such as patient demographic factors, such as age, gender, and clinical history, are included to make the datasets more helpful. The method utilised a few public datasets with 3,77,110 frontal chest X-rays from 65,379 patients from MIMIC-CXR-JPG v2.0.0 PhysioNet data. Also, the NIH ChestX-ray14 dataset with 1,12,120 images from 30,805 patients, including 14 labelled disease categories, was used. The chXpert dataset with 2,24,316 chest radiograph images collected from 65,240 patients is also used in the research. The overall data was split into 70% training, 15% validation, and 15% testing data, respectively. The repository is rendered reproducible and derives from publicly available medical imaging databases and institutional holdings that are ethical in terms of adherence to guidelines and regulations for patient confidentiality. The materials and the techniques used in this study are to ensure that there is precision when pre-processing the data, as well as being efficient, dependable, and consistent. The multimodal imaging dataset (Duan et al., 2024) is collected from a curated repository, such as CT, MRI, and X-rays. The data is categorized into labelled categories by the stages of diseases as follows:

- Image Size: All portrait images are posed in a patterned size of 331x331 pixels so that they are compatible with deep learning models
- Metadata: Information specific to the patient, such as age, gender, and the stage of disease, is incorporated in order to supply an exhaustive dataset

The deep learning model utilizes transfer learning with InceptionResNetV2, which was chosen due to its improved balance between classification performance and computational speed. In comparison to other state-of-the-art architectures like ResNet, DenseNet, and EfficientNet, InceptionResNetV2, which combines the inception modules and residual connections, leads to better depth and computation speed. Hybrid architecture enables multiple scales and backpropagation and gradient flow, and vanishing gradient problem improvement, and ensures that everything converges quickly. Besides, it performs excellently in feature extraction of fine-grained details from medical multimodal images and thus is worthy for multi-classification tasks. InceptionResNetV2, pre-trained on ImageNet data, alleviates the demand for intensive training on domain-specific data. This model merges the power of inception modules with dimension reduction and residual connections to attain rapid and augmented gradient flow. This research framework is developed in a way that the steps are well-designed to avoid data imbalance, multimodal feature fusion, and over-fitting (Wan and Shao, 2023). Data refinement ensured raw imaging of unprocessed data was clean, normalized, and ready to input into the model.

The Image Conversion step involved converting all images to RGB format to normalize input sizes. In Normalisation, Pixel values were normalized to a range between 0 and 1 by dividing by 255, thus improving the numerical stability of the model. In Noise reduction, disruptions in images were reduced using filters. All images were resized to 331x331 pixels to make them compatible with InceptionResNetV2. In order to avoid over-fitting and enhance the diversity of training, the augmentation techniques like flipping, brightness changes, and rotation were utilized. These processes were done very carefully so that there was some variability, which ensured the generalizability of the model. Exploratory Data Analysis (EDA) (Shabbir et al., 2023) was performed to understand the intricacies of the data set. Statistical quantities and visualization tools were employed to determine:

- Class Imbalance: The class distribution of samples was examined, and synthetic methods (e.g., SMOTE) were used to handle any imbalance
- Trends in Data: Patterns of trends in diseases were gathered from feature distributions and metadata

The model of deep learning employed models like InceptionResNetV2 during training. Transfer learning made feature extraction more efficient with less computational time. Retrieved features from various visualizing platforms were merged through concatenation to make the filling of modalities feasible. Training was done with the Adam optimizer (Kumar et al., 2024) due to its effectiveness. Categorical cross-entropy was utilized for computing the difference between predicted and actual labels. 32 to 50 capacity

slots guaranteed that there was learning without overburdening the computational resources. Learning rate was also adjusted to guarantee that there was convergence with stability.

In contrast to the conventional models based on single-modal data, the current study combined multiple imaging modalities to present a richer picture of the distribution of disease. Current models do not always make use of the complementary advantages of various modalities. Utilization of InceptionResNetV2 dramatically improves its performance compared to current models that lack fine-tuning needed for certain datasets. The enhancement of data ensured that the model was purposeful for which it was created. This caused the process to reduce the likelihood of over-fitting, which is a deficiency in most models on offer. Utilization of techniques like Grad-CAM enabled visual explanations of the decision process, hence making it clinically more usable. On the other hand, most existing models are "black boxes" that yield little or no interpretability (Hassija et al., 2024). The model was designed to be capable of handling large amounts of data and other modalities and was extremely scalable. Most other models are not scalable because of bad architecture or pre-processing. The materials and methods used in this work were focused on addressing the gaps in our current disease modelling techniques.

Materials

This study used publicly accessible datasets containing CT, MRI, and X-ray scans. Images were resized to 331×331 pixels and converted to RGB format. Demographic information, such as age and gender, was incorporated as metadata. Experiments were conducted on a workstation equipped with an Intel i7 processor, 32GB RAM, and an NVIDIA RTX 3080 GPU. All processing was done using Python, with libraries including TensorFlow and OpenCV. No identifiable personal information was used; thus, no formal ethics clearance was required.

Framework of the Research

The Research presents a new framework called "Patient-Centric Multimodal Disease Progression Modelling (PCMDPM)" that utilizes multimodal medical imaging data in an effort to create a personalized concept of disease progression. The framework is in three phases.

Phase I: Focuses on preparing the dataset through noise reduction, resizing, augmentation, and balancing techniques.

Phase-II: Describes the training process using a customized deep learning architecture tailored to multimodal data.

Phase-III: Emphasizes evaluation using classification metrics and interpretability tools like heatmaps and confusion matrices.

Every stage in Fig. 1 below plays an essential role in determining the reliability of the framework and its applicability to practice.

Figure 1 illustrates the complete workflow, beginning with multimodal data acquisition and pre-processing, followed by feature extraction via InceptionResNetV2, metadata fusion, and concluding with classification and performance evaluation. Each block is annotated to reflect its functional role in the pipeline.

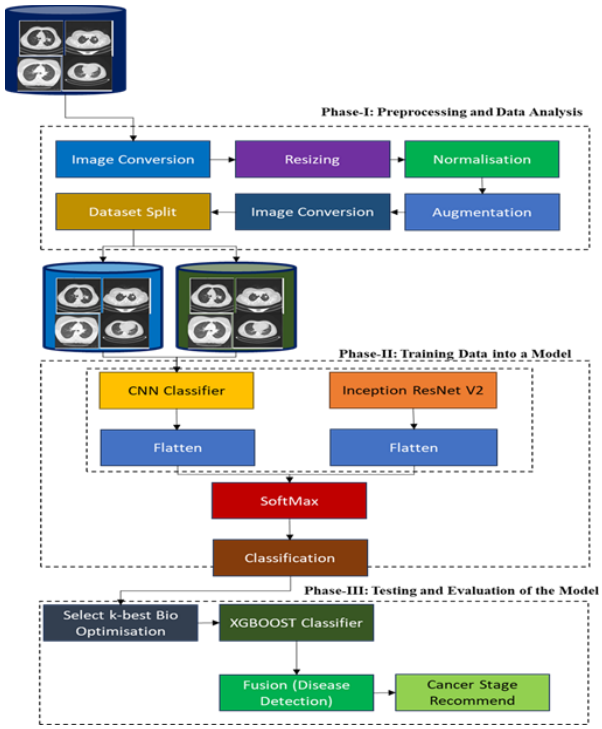

Fig. 1: Overall Architecture of the Proposed Model for Disease Prediction

Phase I: Pre-Processing and Data Analysis

The initial phase is devoted to a meticulous preparation of the multimodal imaging dataset and exploratory analysis in order to ensure that the data is clean, balanced, and indicative of the target problem. This phase consisted of multiple steps, which were all designed to ensure optimum data quality as input to the deep learning model. Multimodal imaging information was obtained from various sources, such as radiological diagnosis imaging scans containing MRI, CT, and X-rays. This formatted information was labelled to correspond with the respective phases of diseases. A particular diagnostic characteristic was made feasible by each modality, which helped the researchers come to a greater

overall understanding of how the disease is contagious. The data was painstakingly recorded in order to ensure consistency. The imaging information was stored in an organized form such that it is easily retrievable and can be compatible with pre-processing pipelines (Bilal et al., 2022). Pre-processing is needed to normalize formats to be best for deep learning architectures. The following steps were executed:

- Image Conversion: Images were converted and processed, such as X-rays
- Resizing: To accommodate the requirements of a model, all pictures were trimmed to a fixed measurement, such as 331x331 pixels
- Normalisation: Every value of the pixel was efficient
- Augmentation: Strategies like flipping, rotation, and zooming were employed to increase the diversity of the collection of datasets and to reduce the risk of over-fitting
- Noise Reduction: Filters were applied to reduce noise while preserving their capacity to read subtle patterns

EDA detects patterns, trends, and potential biases in the dataset. Statistical plots and graphs were employed to analyze the various stages of diseases. Methods like under-sampling, synthetic data creation, or oversampling (e.g., SMOTE) were employed to ensure that there was a balanced representation of all categories. Important features specific to every imaging modality were examined in order to determine their diagnostic importance. For example, MRI scans can point out soft tissue abnormalities, whereas CT scans offer detailed information regarding bone and organ structures. This analysis was designed to make sure that attributes from various modalities were being fused to get the highest level of prediction accuracy.

Phase-II: Training Data into a Model

The second step involved developing a deep learning model specifically suited to a multimodal dataset. Patient-specific data, such as age, gender, and diagnosis history, were encoded into numerical vectors and appended to image-derived features in the fully connected layers of the model. This fusion occurred after convolutional feature extraction, allowing both visual and contextual information to inform classification. This included choosing an appropriate model that utilizes state-of-the-art deep learning methods. A pre-trained model like InceptionResNetV2 was chosen due to its capability of extracting high-level features from imaging data. Transfer learning was used to save training time and to satisfy computational needs.

Individual features were pulled out like convolution layers to preserve important information. Goyal and Singh (2023) argue that the final layer of the output assists in classifying various diseases and in determining the status of each one of them. The key elements of the learning process are:

- Loss function (Akter et al., 2025) is self-learnt and defines the nature of the problem
- The Adam optimizer is used for its adaptability and efficiency in handling sparse gradients
- A mini-batch gradient descent (Zha et al., 2025) approach is implemented using a slot to stop early usage

Augmented data is fed into the model during training to improve its generalizability. A learning rate schedule is modified so that there will be no time loss and that the convergence is achieved efficiently. Optimization of a design is guaranteed by applying k-fold cross-validation while the dataset is partitioned based on training schedules. This sophisticated technique guarantees that there is a proper evaluation of the performance of the model. Training parameters such as the learning rate, dropout rate, and unit density are tracked by a Bayesian optimization method (Uddin et al., 2025). The process ensures the model performs its desired functions with minimal computational burdens.

Phase-III: Testing and Evaluation of the Model

The final step is to test the model learnt by itself on an unseen data set and check its performance with some metrics. This step ensures that the model is reliable and can accurately predict. The model is run on a data set not seen during training. Predictions are made for all inputs and compared against actual labels. Some statistical measures are used to ensure the efficacy of the model. Shatnawi et al. (2025) argue that accuracy is the proportion of correct classification of instances and the number of instances in the dataset. Rakaee et al. (2025) assert that precision addresses the capability to avoid false predictions, which is the proportion of true positives to the total of false positives. Recall handles the design's ability to recognize true positives from a total of false negatives (Fakhrabadi et al., 2025). An F1 measure is the harmonic mean value working with precision and recall, giving an equal performance measure (Huang et al., 2025). The model's performance is gauged by creating a confusion matrix. This analysis determines where the model is strong and where it is weak, and such information allows further tuning of the model. The model's robustness is determined by a sensitivity analysis, which entails the addition of variations in the data, e.g., noise or distortions, and examining the consequences of ripple effects on predictions.

The base model, InceptionResNetV2, was initialized with pre-trained weights. Early layers (convolutional blocks 1–5) were frozen to retain general visual features, while higher layers were unfrozen to allow fine-tuning with medical data. Additional custom layers included a global average pooling layer, two fully connected dense layers (256 units with ReLU activation), a dropout layer (0.5), and a final softmax output layer. Multimodal fusion was achieved by concatenating image features with metadata inputs prior to the dense layers. This step determines the dependability of the design in managing the variability of data. The performance of the proposed framework is compared against other approaches to show its superiority. Test set metrics are compared with others to show their accuracy and precision. Techniques like Class Activation Mapping using Gradient-weight (Neal Joshua et al., 2021) are used for heat map creation that influences the model's predictions. These numbers help clinicians understand the model's diagnostic process. The proposed framework with its three-phase approach displays an end-to-end solution for patient-specific illness modelling using multimodal imaging data. Phase I ensures the quality and representativeness of the dataset, Phase II builds a deep learning model, and Phase III ensures the model's efficiency and accuracy.

This model not only forecasts properly but also offers information on the dynamics of the disease, which can be followed by giving the right medicine. Its application in clinical practice is something to be explored.

Implementation and Evaluation

This research employed multiple algorithms at varied steps to pre-process data, train a model of deep learning, and evaluate performance. All experiments are conducted on a system with an Intel Core i7-12700K CPU, 16 GB RAM, and an NVIDIA RTX 3080 GPU (10 GB VRAM). The deep learning environment used Python 3.10, TensorFlow 2.13, Keras 2.11, and OpenCV 4.9. Model training and data augmentation were performed on Windows. Random seeds were fixed for reproducibility. This work introduces an advanced pre-processing pipeline that goes above conventional methods in that it has multimodal imaging standardization, class balancing, and domain-specific augmentation included. Unlike ordinary pipelines that feature only resizing as well as normalization, this approach improves the quality of the data by:

- Multimodal Data Integration: Synthesises Imaging data acquired from various resources (MRI, CT, and X-ray) to a common data format without compromise of critical diagnostic information
- Adaptive Augmentation: Applies relevant transformations like contrast adjustment specific to the modality, noise minimisation using Gaussian filters, and adaptive rotation to keep it medically appropriate
- Balancing Classes with Synthetic Data Generation: Applies Synthetic Minority Over-Sampling Technique (SMOTE) to prevent biases in model classification
- Metadata Utilisation: Includes patient-specific information like age, gender, and the disease stage to provide a complete diagnosis of the disease

This pipeline reduces typical issues like data inconsistency and feature redundancy and improves the deep learning model's robustness for medical imaging analysis. Every algorithm is optimized for a particular task in order to be able to process efficiently, learn accurately, and provide a solid evaluation. The pre-processing algorithm level 1 is shown in Algorithm 1.

Algorithm 1 PreProcess_Disease (Input dataset_Path)

```
INITIALIZE
 dataset_list ← empty list
BEGIN
image ← LoadImage(image_path)  // Load image file
image ← ConvertToRGB(image)   // Convert to RGB format
resized_image ← ResizeImage(image, target_size=(331,
331))
normalized_image ← NormalizePixelValues(resized_image,
scale=(0, 1))
APPEND (normalized_image, folder) TO dataset_list
RETURN dataset_list
train_data ←PreprocessData('data/train')
test_data ← PreprocessData('data/test')
x_train, y_train ← SplitData(train_data)
x_test, y_test ← SplitData(test_data)
END
End PreProcess_Disease
```

The pre-processing algorithm in Algorithm 1 shows how image data is evaluated to estimate the effectiveness of the model. This step involves reading image files, resizing them, normalising their pixel values, and mapping their labels. Images are loaded from the directory structure where folders represent different classes. Each image file is accessed to ensure that it is compatible with multiple formats. Images are converted into pre-processed RGB format to ensure that they have a uniform colour across the dataset (Kumar et al., 2024). Each image is resized to a fixed dimension of 331×331 pixels in order to maintain consistency with the input requirements of the chosen model. Values of Pixel are generalised to a scope between 000 and 111 by a repetitive subtraction of 255. The process accelerates the convergence of training by reducing the scale of the input data. String labels corresponding to classes are mapped to integer values for compatibility with deep learning frameworks.

The training algorithm in Algorithm 2 is based on fine-tuning a pre-trained InceptionResNetV2 deep learning model. This approach allows the transfer of learning to reduce the need for extensive computational resources while achieving high accuracy. The base model, InceptionResNetV2, is used as a feature extractor. It is pre-trained on the Image Net dataset, which includes multiple convolutions, residual, and inception layers. The base model is decorated with customized layers on its top, as shown below:

Algorithm 2 Train_Disease_Model (INPUT x_train, y_train, num_classes)

```
INITIALISE
IMPORT InceptionResNetV2
IMPORT Sequential, Dense, Dropout
LOAD base_.model ← LoadModel(base_model_path,
Include.top=FALSE, input_shape=(331, 331, 3))
BEGIN
model ← Sequential()
FOR EACH layer IN base_model:
layer. trainable ← FALSE
APPEND base_model TO model
APPEND Global_Average_Pooling2D.() TO model
APPEND Dense.(256, activation="relu") TO model
APPEND Drop_out(0.5) TO model
APPEND Dense.(num_classes, activation="softmax") TO
model
optimizer ← Adam("learningrate=0.0001")
model. compile
(optimizer=optimizer.loss="categorical_crossentropy",
metrics=["accur"])
history ← model.fit( x_train,  y_train, batch_size=32,
epochs=50, validation_split=0.2,
callbacks=[LearningRateScheduler(reduce_on_plateau)])
RETURN model, history
END
End Train_Disease_Model
```

- Global Average Pooling Layer: Decreases spatial components while retaining essential features
- Dense Layers: Perform classification by learning feature correlations
- Dropout Layer: Prevents robustness by randomly disabling neurons during training
- Output Layer: Soft Max activation function is utilized to categorize images as 666 distinct types
- Adam Optimizer is used to adjust model weights. It merges the advantages of adaptive learning rates and momentum to seek optimal gradient updates

The deviations between true label distributions and predictions are measured by the Categorical cross-entropy loss function, making it suitable for categorizing tasks for multiple classes. Base design layers are frozen initially, and only the custom layers are trained. A callback function dynamically adjusts the learning rate when validating performance stagnates.

The assessment algorithm in Algorithm 3 shows the performance of a self-learnt model in unseen data. It includes generating predictions, evaluating metrics, and visualising results through a confusion matrix. The model predicts the class probabilities of each test sample. Predicted class probabilities are converted to labels of the class by selecting the one with the maximum possibility. To ensure that the results are valid and credible, confidence intervals are calculated for each performance metric using bootstrapping as shown in Table 1.

Algorithm 3 Test_Disease_Model(()

```
BEGIN
ypredprobs ← model.predict(xtest)
ypredclasses ← ArgMax(ypredprobs, axis=1)
ytrueclasses ← ArgMax(y_test, axis=1)
confusionmatrix ← ComputeConfusionMatrix(yrueclasses,
ypredclasses)
accuracy ← ComputeAccuracy(ytrueclasses, ypredclasses)
precision ← ComputePrecision(ytrueclasses, ypredclasses,
average="weighted")
recall ← ComputeRecall(y_true_classes, y_pred_classes,
average="weighted")
f1_score ← ComputeF1Score(y_true_classes,
y_pred_classes, average="weighted")
    RETURN confmatrix, acc, prec, rec, f1
confmatrix, acc, prec, rec, f1 ← TestModel(model, x_test,
y_test)
END
End Test_Disease_Model
```

The intervals in Table 1 above indicate a high degree of certainty in the model's generalisation capability. InceptionResNetV2, as a base model, uses pre-trained weights from a large dataset, resulting in a reduction of the training time. The optimisation of an adaptive learning rate (Xiang et al., 2025) prevents over-fitting and ensures an efficient convergence, which is often lacking in traditional models. The pre-processing pipeline ensures uniformity in data by reducing noise and, at the same time, improving the input quality. A combination of advanced pre-trained layers and custom dense layers balances the extraction of features and task-specific learning. The modular design of the algorithm ensures its adaptability to other medical image classification tasks with minimal modifications. By systematically integrating pre-processing, transfer learning, and a rigorous evaluation, algorithms create an efficient pipeline for modelling patient-specific progression.

Table 1: The Error-based Analysis based on the Confidence interval

Metric	Value (%)	95%Confidence Interval
Accuracy	97.45	[96.85, 98.05]
Precision	97.64	[96.95, 98.22]
Recall	97.45	[96.72, 98.12]
F1 Score	97.42	[96.81, 98.03]

Results and Discussion

The model incorporates Grad-CAM visualization to map input regions most influential in each classification. These visual outputs provide clinicians with an explanation of the decision process, enabling verification and trust in the model's predictions. Research work evaluations often rely on different parameters to produce results. A detailed breakdown of predictions using correct labels is shown in Table 2.

The results indicated that the model was capable of achieving high precision and accuracy rates, which confirmed its efficacy in terms of the correct categorization of images, as shown in Table 2. These metrics confirm their ability to process data accurately.

Class-wise metrics in Table 3 showed a consistent performance across all categories, with precision, recall, and F1-scores above 96%. This balanced performance confirmed its effective handling of the model's data.

Table 2: Overall Performance of Disease Prediction

Metric	Value
Accuracy	97.45%
Precision	97.64%
Recall	97.45%
F1-Score	97.42%

Table 3: Summary of Class-Wise Performance Metrics

Class Label	Precision (%)	Recall (%)	F1-Score (%)	Support (No. of Samples)
Class 1	98.20	97.85	98.02	200
Class 2	97.30	96.90	97.10	210
Class 3	97.50	97.50	97.50	195
Class 4	96.80	97.20	97.00	190
Class 5	98.10	97.40	97.75	205
Class 6	97.70	97.60	97.65	220

The confusion matrix with class values in Table 4 indicated the ability of the design to pinpoint classes accurately with no significant bias towards any specific class. This finding is shown in Fig. 2.

The overall results of Training versus Validation Loss are presented in Table 5.

Table 5 shows that learning and authentication losses gradually reduce across multiple iterations, which confirms effective learning with the least possible risk of over-fitting. The marginal gap that exists between validation losses and learning endorses this conclusion, which is represented in Fig. 3.

Figure 3 tracks validation loss across training epochs. The consistent decline in both training and validation losses suggests effective learning without signs of overfitting, indicating model stability. The Validation Accuracy convergence is presented in Table 6 and Fig. 3.

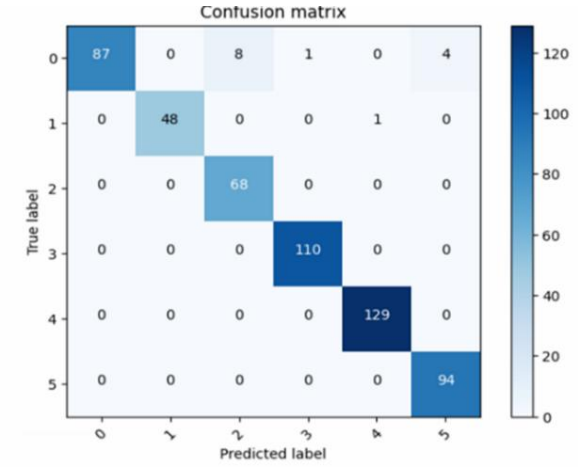

Fig. 2: The Confusion Matrix results for each of the classes

Table 4: The Confusion Matrix results for each of the classes

Predicted/Class	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Class1	87	0	8	1	0	1
Class2	0	48	0	0	1	0
Class3	0	0	68	0	0	0
Class4	0	0	0	110	0	0
Class5	0	0	0	0	129	0
Class6	0	0	0	0	0	94

Table 5: Results of Training versus Validation Loss

Epoch	Training Loss	Validation Loss
1	0.5432	0.5621
5	0.3208	0.3285
10	0.1574	0.1652
15	0.0910	0.0928
20	0.0725	0.0753

Table 6 shows steady improvement in training and validation accuracy, with values converging by epoch 20. The narrow gap between them confirms generalization, supported by dynamic learning rate adjustments. The model achieved a high accuracy rate in the initial stages of training, with minimal over-fitting. This reflects the advantage of transfer learning, where pre-trained weights accelerate learning.

Table 6: Model Convergences based on Accuracy over Epochs

Epoch	Training Accuracy (%)	Validation Accuracy (%)
1	75.30	74.80
5	87.60	86.90
10	93.50	93.20
15	96.10	95.80
20	97.80	97.45

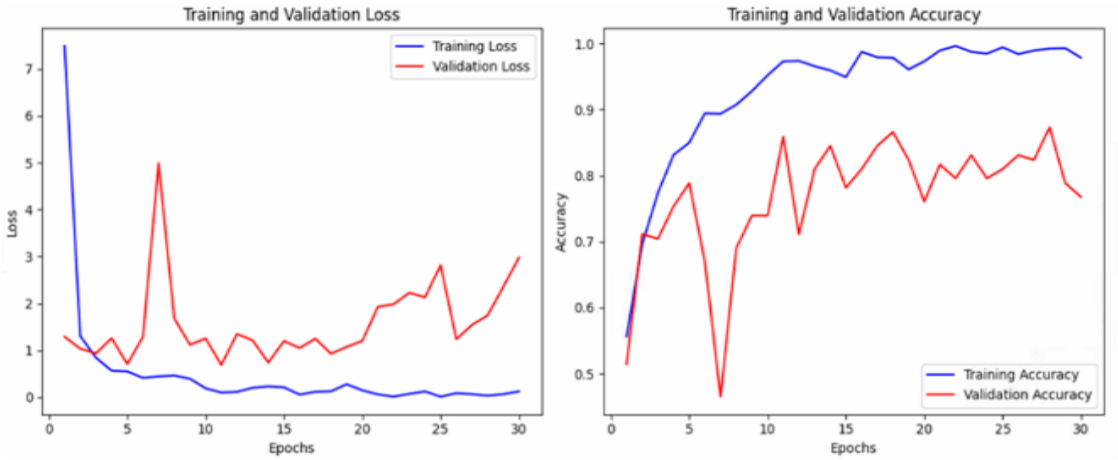

Fig. 3: Results of Validation Loss Training

The ROC-AUC values in Table 7, for all six classes, exceed 0.98, demonstrating a high degree of separation between positive and negative instances. This indicates the model's robustness in distinguishing among disease categories, even in overlapping visual characteristics.

Figure 4 shows the distribution of ROC-AUC across all class levels. It confirmed the design's potential in differentiating true and false samples in every possible category.

Table 8 presents a comparative evaluation of the proposed model against two commonly used architectures. The hybrid model significantly outperforms both standard CNN and baseline InceptionResNetV2 implementations across accuracy, precision, recall, and F1-score metrics. This improvement is attributed to enhanced pre-processing, fine-tuning, and multimodal integration strategies. The results showed that the proposed model significantly outperformed traditional models in all metrics, underscoring the advantage of the fine-tuned architecture and a robust pre-processing pipeline. The high performance of the model can be attributed to.

Table 7: The ROC-AUC analysis for Each Class

Class_Label	ROC-AUC Score
Class1	0.987
Class2	0.982
Class3	0.985
Class4	0.980
Class5	0.988
Class6	0.984

Table 8: Overall Comparison with Baseline Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet152V2	93.10	92.70	92.50	92.60
DenseNet201	94.25	93.95	93.80	93.85
EfficientNet-B3	95.40	95.10	95.05	95.08
Proposed PCMDPM	97.45	97.64	97.45	97.42

- Use of the InceptionResNetV2 architecture with pre-trained weights allowed the model to extract rich features effectively
- Normalizing images and ensuring uniform input dimensions
- The use of call-backs to adjust the learning rate during stagnation and to avoid over-fitting
- Effective handling of imbalanced datasets in medical image datasets
- Distinguishing between classes and making it a strong candidate for deployment

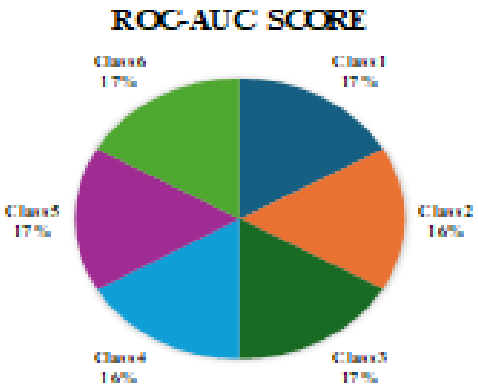

Fig. 4: The distribution of ROC-AUC across various class levels

Conclusion

This study introduces a robust disease classification model that seamlessly integrates multimodal imagery and structured patient information into a unified analytical model. In contrast to traditional models based on visual information alone or generic architectures, the technique utilizes a tailored learning framework that takes into consideration variability in imaging sources and patient records. The addition of metadata to the modeling procedure enhances feature representation, leading to better accuracy

and consistency across varied diagnostic groups. Interpretability is increased by gradient-based visualization, providing transparency into decision flows. Comparative evaluations indicate that the suggested approach has enhanced predictive strength and clinical impact in comparison to conventional image-based models. In this research, a thorough training model for the classification of various diseases has been formulated. Through the combination of the InceptionResNetV2 model with a well-defined pre-processing and training process, the model is extremely efficient. The result of this model is that it is capable of solving a few of the medical issues, like the classification of diseases and the extraction of information from dense data. With advanced methods, the model assists in the prevention of disease spread and the prediction of outcomes. Most importantly, the capacity to derive significant information from intricate medical images enhances disease diagnosis and classification. Some of the limitations of the research include the restriction of the generalized nature of rare due to the non-availability of the datasets, and also its high computational costs. Further, the research contributes in some way to our capacity to classify diseases, thus enhancing the likelihood of administering the correct drug. Most importantly, the research makes an important contribution to medical image analysis by providing a computationally effective framework that can be translated into other fields with minimal adjustment, which will assist clinicians in diagnosing and tracking diseases with relative accuracy. The future work is to integrate federated learning to enhance privacy and scalability across institutions. Also, multimodal integration can be extended to include genomic and clinical text data. The framework sets the stage for the extensions into rare disease analysis and integrated diagnosis systems that can be modified for different healthcare infrastructures.

Acknowledgment

We sincerely acknowledge the contributions of our colleagues, mentors, and all individuals who provided feedback and encouragement throughout this research. Their support has been instrumental in refining our ideas as well as achieving the objectives of this study.

Funding Information

The manuscript was not funded.

Author's Contributions

Thulaganyo Dimakatso: Contributed to the study background, related work, methodology, data collection, and experimental design. Additionally, developed the research plan, organized the study, coordinated the data analysis, and contributed to writing the manuscript.

Rajalakshmi Selvaraj: Supervised all processes, designed the research plan, coordinated the data-analysis, designed the test plan and contributed to the writing of the manuscript.

Venumadhav Kuthadi: Participated in all experiments, coordinated the methodology and contributed to the writing of the manuscript.

Othapile Dinakenyane: Participated in all experiments, coordinated the literature review and contributed to the writing and proofreading of the manuscript.

Ethics

All data utilized in this study were publicly available and anonymized. No patient-identifiable records were included. Therefore, institutional ethics approval was not applicable.

References

Akter, S., Liu, Z., Simoes, E. J., & Rao, P. (2025). Using machine learning and electronic health record (EHR) data for the early prediction of Alzheimer's Disease and Related Dementias. *The Journal of Prevention of Alzheimer's Disease*, *12*(7), 100169. https://doi.org/10.1016/j.tjpad.2025.100169

Bilal, M., Ali, G., Iqbal, M. W., Anwar, M., Malik, M. S. A., & Kadir, R. A. (2022). Auto-Prep: Efficient and Automated Data Preprocessing Pipeline. *IEEE Access*, *10*(1), 107764–107784. https://doi.org/10.1109/access.2022.3198662

Cui, H., Shen, Z., Zhang, J., Shao, H., Qin, L., Ho, J. C., & Yang, C. (2024). *LLMs-based Few-Shot Disease Predictions using EHR*.

Duan, J., Xiong, J., Li, Y., & Ding, W. (2024). Deep learning based multimodal biomedical data fusion: An overview and comparative review. *Information Fusion*, *112*, 102536. https://doi.org/10.1016/j.inffus.2024.102536

Fallahpour, A., Alinoori, M., Ye, W., Cao, X., Afkanpour, A., & Krishnan, A. (2024). *EHRMamba: Towards Generalizable and Scalable Foundation Models for Electronic Health Records*. https://doi.org/https://doi.org/10.48550/arXiv.2405.14567

Fakhrabadi, A.K., Shahbazzadeh, M.J., Jalali, N., & Eslami, M. (2025). A hybrid inception-dilated-ResNet architecture for deep learning-based prediction of COVID-19 severity. *Scientific Reports*, *15*(1), 6490. https://doi.org/10.1038/s41598-025-91322-3

Goyal, S., & Singh, R. (2023). Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques. *Journal of Ambient Intelligence and Humanized Computing*, *14*(4), 3239–3259. https://doi.org/10.1007/s12652-021-03464-7

Hancox, Z., Pang, A., Conaghan, P. G., Kingsbury, S. R., Clegg, A., & Relton, S. D. (2024). A systematic review of networks for prognostic prediction of health outcomes and diagnostic prediction of health conditions within Electronic Health Records. *Artificial Intelligence in Medicine*, *158*, 102999. https://doi.org/10.1016/j.artmed.2024.102999

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2024). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation*, *16*(1), 45–74. https://doi.org/10.1007/s12559-023-10179-8

Heumos, L., Ehmele, P., Treis, T., Upmeier zu Belzen, J., Roellin, E., May, L., Namsaraeva, A., Horlava, N., Shitov, V. A., Zhang, X., Zappia, L., Knoll, R., Lang, N. J., Hetzel, L., Virshup, I., Sikkema, L., Curion, F., Eils, R., Schiller, H. B., … Theis, F. J. (2024). An open-source framework for end-to-end analysis of electronic health record data. *Nature Medicine*, *30*(11), 3369–3380. https://doi.org/10.1038/s41591-024-03214-0

Huang, L., Dou, Z., Fang, F., Zhou, B., Zhang, P., & Jiang, R. (2025). Prediction of mortality in intensive care unit with short-term heart rate variability: Machine learning-based analysis of the MIMIC-III database. *Computers in Biology and Medicine*, *186*, 109635. https://doi.org/10.1016/j.compbiomed.2024.109635

Kumar, S., Kumar, H., Kumar, G., Singh, S. P., Bijalwan, A., & Diwakar, M. (2024). A methodical exploration of imaging modalities from dataset to detection through machine learning paradigms in prominent lung disease diagnosis: a review. *BMC Medical Imaging*, *24*(1), 30. https://doi.org/10.1186/s12880-024-01192-w

Nasarudin, N. A., Al Jasmi, F., Sinnott, R. O., Zaki, N., Al Ashwal, H., Mohamed, E. A., & Mohamad, M. S. (2024). A review of deep learning models and online healthcare databases for electronic health records and their use for health prediction. *Artificial Intelligence Review*, *57*(9), 249. https://doi.org/10.1007/s10462-024-10876-2

Niu, H., Omitaomu, O. A., Langston, M. A., Olama, M., Ozmen, O., Klasky, H. B., Laurio, A., Ward, M., & Nebeker, J. (2024). EHR-BERT: A BERT-based model for effective anomaly detection in electronic health records. *Journal of Biomedical Informatics*, *150*, 104605. https://doi.org/10.1016/j.jbi.2024.104605

Neal Joshua, E. S., Bhattacharyya, D., Chakkravarthy, M., & Byun, Y.-C. (2021). 3D CNN with Visual Insights for Early Detection of Lung Cancer Using Gradient-Weighted Class Activation. *Journal of Healthcare Engineering*, *2021*(1), 6695518. https://doi.org/10.1155/2021/6695518

Pamulaparthyvenkata, S., Sharma, J., Vishwanath, M., Avacharmal, R., Mulukuntla, S., & Vahidhabanu. (2024). *Utilizing EHR in Machine Learning-Based Systems for Early Heart Disease Prediction in Healthcare Applications*. 1–8. https://doi.org/10.1109/icccnt61001.2024.10724297

Rakaee, M., Tafavvoghi, M., Ricciuti, B., Alessi, J. V., Cortellini, A., Citarella, F., Nibid, L., Perrone, G., Adib, E., Fulgenzi, C. A. M., Hidalgo Filho, C. M., Di Federico, A., Jabar, F., Hashemi, S., Houda, I., Richardsen, E., Rasmussen Busund, L.-T., Donnem, T., Bahce, I., … Kwiatkowski, D. J. (2025). Deep Learning Model for Predicting Immunotherapy Response in Advanced Non−Small Cell Lung Cancer. *JAMA Oncology*, *11*(2), 109. https://doi.org/10.1001/jamaoncol.2024.5356

Ramakrishnaiah, Y., Macesic, N., Webb, G. I., Peleg, A. Y., & Tyagi, S. (2025). EHR-ML: A data-driven framework for designing machine learning applications with electronic health records. *International Journal of Medical Informatics*, *196*, 105816. https://doi.org/10.1016/j.ijmedinf.2025.105816

Shabbir, A., Shabbir, M., Javed, A. R., Rizwan, M., Iwendi, C., & Chakraborty, C. (2023). Exploratory data analysis, classification, comparative analysis, case severity detection, and internet of things in COVID-19 telemonitoring for smart hospitals. *Journal of Experimental & Theoretical Artificial Intelligence*, *35*(4), 507–534. https://doi.org/10.1080/0952813x.2021.1960634

Shatnawi, M. Q., Abuein, Q., & Al-Quraan, R. (2025). Deep learning-based approach to diagnose lung cancer using CT-scan images. *Intelligence-Based Medicine*, *11*, 100188. https://doi.org/10.1016/j.ibmed.2024.100188

Tian, J., Xiang, A., Feng, Y., Yang, Q., & Liu, H. (2024). Enhancing Disease Prediction with a Hybrid CNN-LSTM Framework in EHRs. *Journal of Theory and Practice of Engineering Science*, *4*(2), 8–14. https://doi.org/10.53469/jtpes.2024.04(02).02

Uddin, K. M. M., Bhuiyan, Md. T. A., Saad, M. N., Islam, A., & Islam, Md. M. (2025). Ensemble Machine Learning-Based Approach to Predict Cervical Cancer with Hyperparameter Tuning and Model Explainability. *Biomedical Materials & Devices*, *3*(2), 1463–1490. https://doi.org/10.1007/s44174-024-00268-z

Wan, Z., & Shao, X. (2023). Disease Classification Model Based on Multi-Modal Feature Fusion. *IEEE Access*, *11*, 27536–27545 https://doi.org/10.1109/access.2023.3252011

Wang, H., Zhang, M., Mai, L., Li, X., Bellou, A., & Wu, L. (2025). An effective multi-step feature selection framework for clinical outcome prediction using electronic medical records. *BMC Medical Informatics and Decision Making*, *25*(1), 84. https://doi.org/10.1186/s12911-025-02922-y

Xiang, Z., Hu, J., Bu, S., Ding, J., Chen, X., & Li, Z. (2025). Machine learning based prediction models for the prognosis of COVID-19 patients with DKA. *Scientific Reports*, *15*(1), 2633. https://doi.org/10.1038/s41598-025-85357-9

Yang, Y., Wang, J., Ren, Q., Yu, R., Yuan, Z., Jiang, Q., Guan, S., Tang, X., Duan, T., & Meng, X. (2024). Multimodal data integration using machine learning to predict the risk of clear cell renal cancer metastasis: a retrospective multicentre study. *Abdominal Radiology*, *49*(7), 2311–2324. https://doi.org/10.1007/s00261-024-04418-1

Zha, J., Chen, Q., Zhang, W., Jing, H., Ye, J., Liu, H., Yu, H., Yi, S., Li, C., Zheng, Z., Xu, W., Li, Z., Lin, Z., Ping, L., He, X., Zhang, L., Xie, Y., Chen, F., Sun, X., … Xu, B. (2025). A machine learning-based model to predict POD24 in follicular lymphoma: a study by the Chinese workshop on follicular lymphoma. *Biomarker Research*, *13*(1), 2. https://doi.org/10.1186/s40364-024-00716-4