



Application of AI methods for multi-modal disease diagnosis

Artificial Intelligence COS5028-B
Coursework2

Student detail: Radin Mokari | 23054961

Instructure: Dr Zied MNASRI

Date of submission: 15.01.2025

Table of Contents

Introduction	3
Background	3
ML Techniques in Healthcare.....	3
Random Forest	3
Decision Tree	3
CNN.....	3
VAEs	4
Challenges in Integrating AI in Healthcare.....	4
Methodology And Data	4
Tabular Dataset	4
Binary Classification	4
Multi-Output Classification.....	5
X-ray Images dataset	5
CNN Classification.....	5
VAE	6
AI Program with User Interaction.....	6
Analysis And Discussion.....	6
RF Binary Classification	6
Decision Tree Multi-Classification.....	7
CNN Classification	8
VAE.....	9
Conclusions And Suggestions for Future Work	9
References	10

Introduction

The integration of machine learning into medical sector promoted a revolution in disease diagnosis. Machine learning empowers computer systems to learn from data, identify intricate patterns and make predictions (Basu Dev Shivahare et al. 2024). In this report, the ML models that have been built for predictions on disease presence will be discussed and analysed. Thought this report, firstly an overview of the ML techniques used for the models will be discussed. Then the building blocks of each model will be covered. And finally the performance of each model will be analysed.

Background

The healthcare industry makes substantial use of artificial intelligence (AI), particularly AI based on deep learning (DL) models (Chaddad et al. 2023). In this study, the machine learning techniques that used for the built models are Random Forest, Decision Tree, Convolutional Neural Network, and Variational Autoencoder.

ML Techniques in Healthcare

Random Forest

The random forest (RF) algorithms is a very suggested ML model technique that summarizes medical data. Rf learns from past information patterns, and it performs by building multiple decision trees, that each tree votes depending on data characteristics and make a prediction by the majority vote. It is effective for regression and classification. One key application of RF is heart disease diagnosis that it will be covered in this report (Dhanka and Maini 2021).

Decision Tree

Another widely used supervised ML techniques is decision tree. It utilizes a tree-based construction and plots investigations regard to item to conclusions about the target value of the item. The tree structure has internal nodes, leaf nodes and a root node. Each internal node refers to a test on a feature, all leaf nodes is assigned to a class and each branch corresponds to the test consequence (Jadhav and Channe 2016).

CNN

Convolutional neural network (CNN) is one of the most popular deep learning models applied for image classification, especially in medical industry. These networks perform a binary classification of detecting if a condition exists or not, therefore they are used primarily for disease detection from X-ray images. These models typically contain a sequence of convolutional layers that try to learn features from a digitized X-ray image.

These layers extract useful visualizations and patterns, and then they are applied to categorization (Behera et al. 2024).

VAEs

Variational autoencoder is a deep learning technique that reconstructs data. Once trained on some data, a VAE would take a sample from the latent space and generate a new sample that closely resembles the input data. The two main components of VAEs are:

Encoders that encode input data into the latent space and decoders that decode the latent representation back into the input space. Essentially, the task of the encoder network is to learn to take a data input and convert it into some low-dimensional latent space. This description of input into the model is translated into the latent space, then reconstructed by a decoder network. VAEs are used to different types of tasks such as semi-supervised learning, text/image generation (Pu et al. 2016).

Challenges in Integrating AI in Healthcare

In the medical field preprocessing and data quality hold great importance for machine learning (ML), especially for deep learning (DL). Building precise and trustworthy ML models that help with tasks like disease diagnosis needs very high-quality data. Based on this truth data challenges to tackle include data scarcity, meaning the real-world data needed for DL training is small, and data imperfection, which means unclear, biased or corrupted information. These issues lead to biased or inaccurate models (Vidhya et al. 2023).

Methodology And Data

Tabular Dataset

Binary Classification

The tabular dataset contains a range of features about health characteristics and 3 diseases (heart disease or attack, diabetes, stroke). The data is either numerical, categorical or binary. For each disease, feature importance has been discovered, and a separate random forest model has been built.

Data Preparation and Normalization and model development

In the csv file, the column “Diabetes” has 3 values (0, 1 and 2). However, since both 1 and 2 represent “true”, all 2 is replaced with 1 with the following code:

```
dataset['Diabetes'] = dataset['Diabetes'].replace(2, 1)
```

On the other hand, since there are numerical and categorical features in the dataset, a normalization is required to be apply before splitting the data into training and testing. The following code block is used to normalize it to a range between 0 and 1.

```
numerical_indices = ['BMI', 'GenHlth', 'MentHlth', 'PhysHlth', 'Age', 'Education', 'Income']  
scaler = MinMaxScaler()  
dataset[numerical_indices] = scaler.fit_transform(dataset[numerical_indices])
```

After that, the SMOTE function is applied to balance the dataset by adding synthetic data points to the minority class.

Now the dataset is normalized and prepared, it's time to split it into training and testing. Finally, the hyperparameters are modified and adjusted to prevent overfitting and increase the robustness of the model.

For all disease, the same steps have been applied with a difference in changing the disease column that defines the target.

Additionally, at the end of each model, a prediction function has been built to make a prediction based on the user inputs.

Multi-Output Classification

For building a model that predicts the presence of each disease simultaneously, the same normalization has been applied to the dataset, but, since we have 3 separate outputs, the resampling is applied for each target to handle multi-output prediction. In this model, a Decision Tree classifier is used as the base model for multi output classification. Also, similar to the RF models, a user interaction has been added to get data from user.

X-ray Images dataset

For this dataset, a CNN classification for the presence of pneumonia, and a variational autoencoder have been created.

CNN Classification

The data preparation step of this model is done by collecting and preprocessing, including resizing images to a uniform dimension, pixel values normalization, and augmentation with transformations such as rotation and flipping to improve generalization. It is importance to mention that the model has been built in the same notebook file as the dataset for an easier implementation and debugging.

The architecture of this model is built by convolutional layers for extracting the features, followed by max-pooling layers for dimensionality reduction and dense layers for classification. Also, non-linearity has been specified by ReLU activation. And finally, the Sigmoid activation function is utilized in the last layer for binary classification.

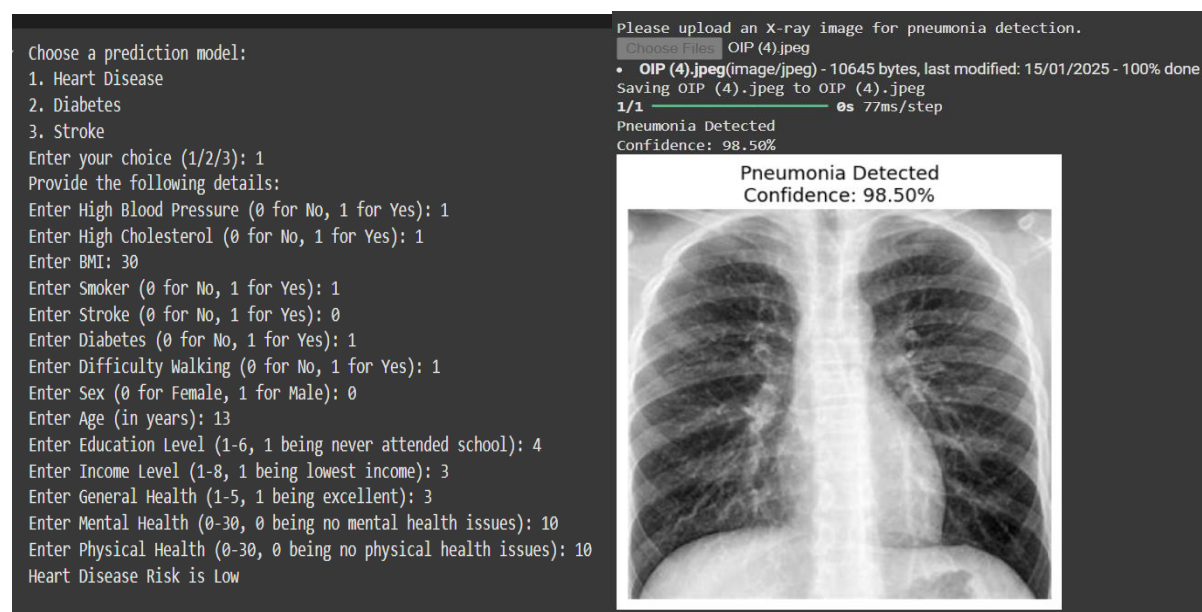
Although the model has a part for make a prediction by using an uploaded image, a separate program has been built that uses the pre-trained model.

VAE

As it discussed before, this model is for reconstructing images. Same as the CNN model, the Pneumonia MNIST dataset was used with data preprocessing that includes resizing images to 28×28 pixels, normalization and batching. The model architecture has two main parts: encoder and decoder. The encoder has convolutional layers for feature extraction, computing latent space parameters and sampling. On the other hand, the decoder has transposed convolutional layers for reconstructing images from latent variables.

AI Program with User Interaction

By saving and downloading the pre-trained models that we covered, a program has been built to make disease diagnosis based on the user data. The program has 2 components: a diagnosis for heart attack, diabetes or stroke, and a pneumonia diagnosis for the uploaded image from user. An example for working with each one is demonstrated below:



Analysis And Discussion

RF Binary Classification

Now it is apparent to dive into model analysis and discuss their performance. The performance of the classification models can be demonstrated by classification report and confusion matrix. For RF models that were built by the tabular data, the performance for all disease diagnosis is approximately similar.:

RF Heart Disease Diagnosis

[[36470 9488]					
[4906 41051]]					
	precision	recall	f1-score	support	
0	0.88	0.79	0.84	45958	
1	0.81	0.89	0.85	45957	
accuracy			0.84	91915	
macro avg	0.85	0.84	0.84	91915	
weighted avg	0.85	0.84	0.84	91915	

RF Diabetes Diagnosis

[[32636 10105]					
[6859 35882]]					
	precision	recall	f1-score	support	
0	0.83	0.76	0.79	42741	
1	0.78	0.84	0.81	42741	
accuracy			0.80	85482	
macro avg	0.80	0.80	0.80	85482	
weighted avg	0.80	0.80	0.80	85482	

RF Stroke Diagnosis

[[39155 9523]					
[5780 42898]]					
	precision	recall	f1-score	support	
0	0.87	0.80	0.84	48678	
1	0.82	0.88	0.85	48678	
accuracy			0.84	97356	
macro avg	0.84	0.84	0.84	97356	
weighted avg	0.84	0.84	0.84	97356	

Decision Tree Multi-Classification

According to the above images, heart disease and stroke diagnosis models have better accuracy and F1-scores than the diabetes model. Also, heart disease has the highest recall score for class 1.

Turning to the performance of the multi output classification model, as represented below, the heart disease diagnosis model has a coherent precision, recall and F1-scores. For the diabetes diagnosis, the model overpredicts positive cases, and this leads to false positives. On the other hand, the stroke detection has the strongest performance, which can be seen from excellent precision and recall for both classes.

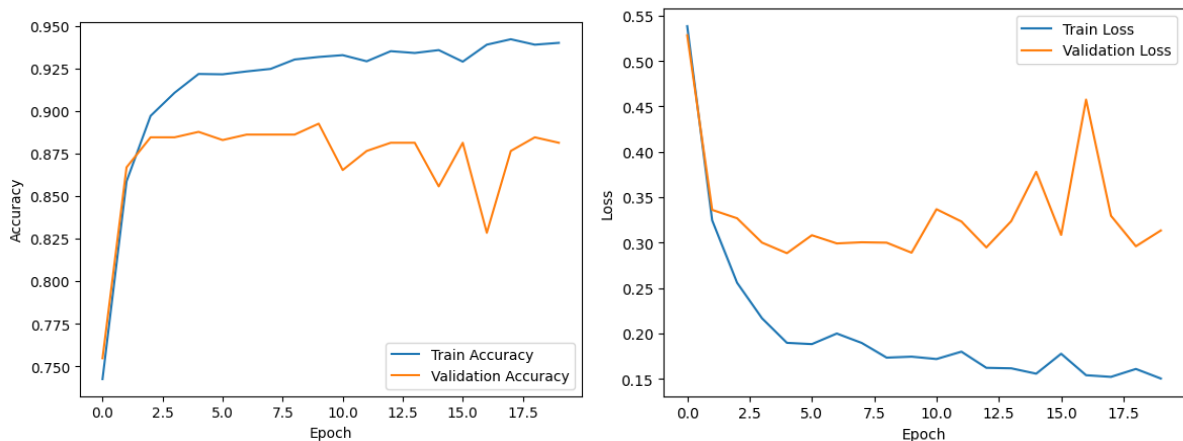
Multi-classification report and confusion matrix

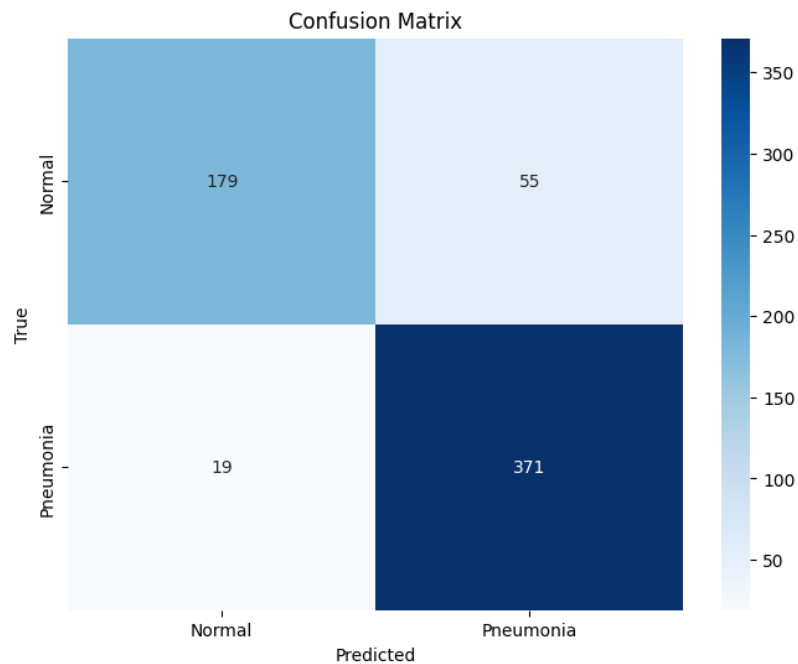
HeartDiseaseorAttack Report:					
	precision	recall	f1-score	support	
0	0.82	0.81	0.82	51472	
1	0.79	0.81	0.80	45884	
accuracy			0.81	97356	
macro avg	0.81	0.81	0.81	97356	
weighted avg	0.81	0.81	0.81	97356	
Diabetes Report:					
	precision	recall	f1-score	support	
0	0.79	0.71	0.75	54661	
1	0.68	0.76	0.72	42695	
accuracy			0.74	97356	
macro avg	0.73	0.74	0.73	97356	
weighted avg	0.74	0.74	0.74	97356	
Stroke Report:					
	precision	recall	f1-score	support	
0	0.83	0.87	0.85	48689	
1	0.86	0.82	0.84	48667	
accuracy			0.85	97356	
macro avg	0.85	0.85	0.85	97356	
weighted avg	0.85	0.85	0.85	97356	

CNN Classification

As it can be seen in the following pictures, the model obtains high accuracy on the test set, which displays its ability to detect pneumonia cases correctly. However, the fluctuations in validation loss refers to some noise sensitivity or overfitting.

```
20/20 - 0s - 18ms/step - accuracy: 0.8814 - loss: 0.3134
Test Loss: 0.31
Test Accuracy: 88.14%
```





VAE

Regard to the performance of the VAE model, the absolute root mean square error shows that the reconstructed images significantly close to the original images. And the relative root mean square error is 18.96%, which indicates that the model captures key data distributions.

Absolute RMSE: 0.1069
Relative RMSE: 0.1896

Conclusions And Suggestions for Future Work

In conclusion, the trained models are usable for disease diagnosis, particularly the binary classification models that have been built for the tabular dataset. However, several refinements and improvements can be applied to advance the performance of the models. For instance, a balanced random forest classifier would be a better choice for multi-output classifier.

References

- Basu Dev Shivahare, Singh, J., Ravi, V., Radha Raman Chandan, Tahani Jaser Alahmadi, Singh, P. and Manoj Diwakar (2024) Delving into Machine Learning's Influence on Disease Diagnosis and Prediction. *The Open public health journal* 17 (1), Bentham Science Publishers.
- Behera, N., Das, S., Singh, A.P., Swain, A.K. and Rout, M. (2024) CNN-Based Medical Image Classification Models for the Identification of Pneumonia and Malaria . 2024 *International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)* IEEE.
- Chaddad, A., Peng, J., Xu, J. and Bouridane, A. (2023) Survey of Explainable AI Techniques in Healthcare. *Sensors* 23 (2), 634.
- Dhanka, S. and Maini, S. (2021) *Random Forest for Heart Disease Detection: A Classification Approach*. 1–3<https://ieeexplore.ieee.org/abstract/document/9699506> Accessed.
- Jadhav, S.D. and Channe, H.P. (2016) Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques. *International Journal of Science and Research (IJSR)* 5 (1), 1842–1845.
- Pu, Y., Gan, Z., Henao, R., Yuan, X., Li, C., Stevens, A. and Carin, L. (2016) Variational Autoencoder for Deep Learning of Images, Labels and Captions. *arXiv (Cornell University)* Cornell University.
- Vidhya, G., Nirmala, D. and Manju, T. (2023) Quality challenges in Deep Learning Data Collection in perspective of Artificial Intelligence. *Journal of Information Technology and Computing* 4 (1), 46–58.