HIV DETECTION- Using Deep Learning

Abstract

The human immunodeficiency virus (HIV) is a hazardous virus that affects the body by corrupting the immune system, particularly the CD4 cells, which are vital as a line of immune defense. In the absence of any treatment, HIV can evolve this way; the person can come down with Acquired Immunodeficiency Syndrome (AIDS) and such cases merely cause fatal problems. This project will employ a deep learning technique to construct a highly sensitive and specific stand-alone diagnostic tool that offers real-time detection and a user-friendly web interface. Feed Forward Neural Networks (FNNs) are employed, based on which the system analyzes different data of age, marital status, HIV test in past years, AIDS education many others, and successfully detects the presence of HIV. A comprehensive dataset kicked off the model training to gain a high degree of accuracy when it comes to diagnosis. When the data is digested and the discussion is done, the outputs will be displayed on a secure and easy-to-navigate webpage that can be acquired by both clinicians and patients without a maze of tests. This novel method not only increases the speed and accuracy of HIV detection but also provides advanced diagnosis tools that are more commonly used by many people, thereby increasing successful patient outcomes, and disease management being done more effectively.

1. INTRODUCTION

Human Immunodeficiency Virus destroys the very foundation of the immune system by infecting CD4 cells that are vital to the survival of the body's defense mechanism against infections. HIV progresses into Acquired Immunodeficiency Syndrome (AIDS) if left untreated, further leading to severe impairment of health and shortened life expectancy. Early and accurate detection of HIV is therefore paramount to effective intervention and management. Deep learning belongs to the subset of machine learning, where utilizing neural networks with several layers to model complex patterns in large datasets often comes into play.

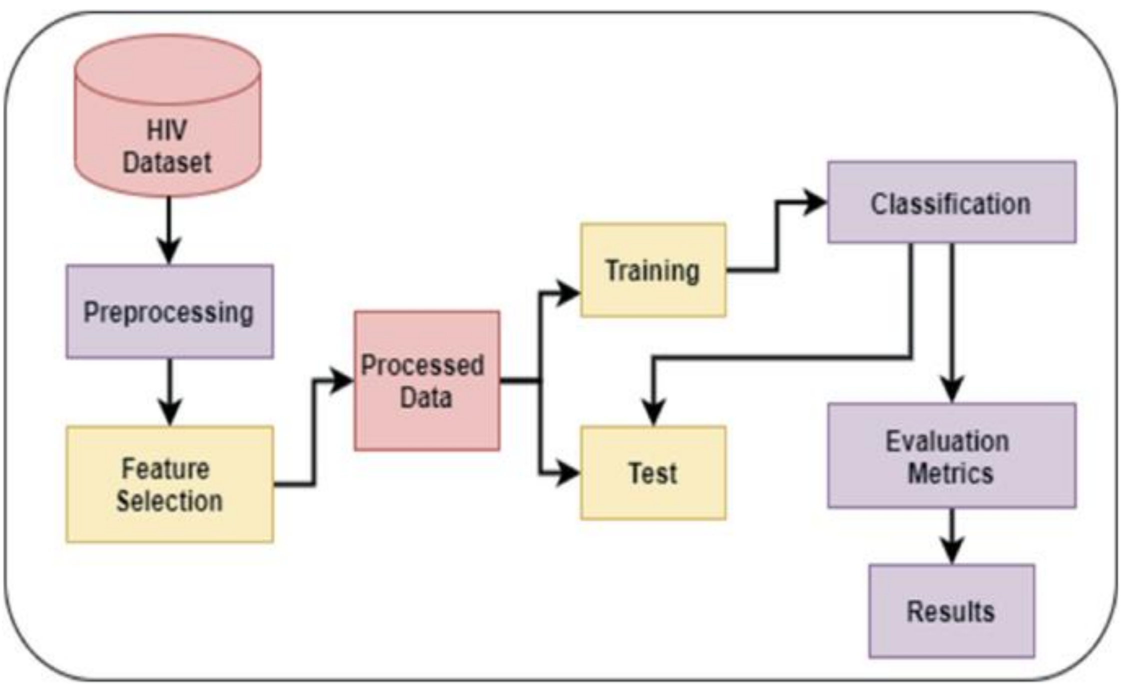
In HIV detection, deep learning analyses can be performed over large and complex datasets, picking up on subtle correlations and patterns in conventional methods that tend to miss out. Our project utilizes a hybrid deep architecture that uses the FeedForward Neural Network (FNN). This is one of the fundamental types of artificial neural networks in which data flows in one direction only from the input nodes, through the hidden nodes, to the output nodes-there are no cycles or loops and, thus, FNN's particularly well suited for supervised learning tasks, especially classification and regression tasks.

In the HIV detection application, FNNs can therefore be used to analyze the complex clinical and sociodemographic data to determine the probabilities for HIV infection.

The hidden and output layers in FNNs will, however process the input data, identify the patterns, and correlate them to ensure that they might indicate the presence of HIV and give a prediction of the status of HIV- indicating the likelihood.

FNNs improve the predictability of the diseases thus enhancing the early detection efforts. This may help someone set the clock in motion towards timely medical intervention of the viruses before they spread further. In addition to this, FNNs are relatively simple and fast, which makes them suitable candidates for deployment in clinical settings where quick and accurate diagnoses are crucial. Preprocessing techniques, such as data normalization, managing missing values, and feature engineering, enable the quality of the data and model performance by using the measures of accuracy like recall and F1 score to determine the accuracy and also by showing the transparency of model decision-making processes that can help health professionals gain trust in the results and identify key risk factors to direct interventions in appropriate ways.

It not only improves the accuracy of the diagnosis but, through deep learning techniques, contributes to a much deeper understanding of factors influencing HIV transmission. Using genomic and immunological markers in multimodal data, our project aims to build a more accurate, real-time diagnostic system.



2. IMPLEMENTATION

This section illustrates the implementation process of the HIV detection system using deep learning, namely a feedforward neural network. The proposed system is based on a dataset of 698 entries and 10 columns covering different features relevant to the diagnosis of HIV. Implementation in the model building, training, and evaluation approaches is considered.

Data Preprocessing

Preprocessing Initially, the dataset is preprocessed for training purpose. Handling of missing values, normalization, and encoding of categorical variables. It checks through all columns in the dataset for inconsistencies that may affect the performance of the model. Missing values are imputed using proper strategies such as mean imputation for numerical columns, while mode imputation is used for categorical columns. Data normalization ensures that all features make equal contributions to the model.

The techniques of standardization and scaling vary based on the data distribution.

In general, for categorical variables, one-shot encoding or label encoding is applied to transform them into a numeric form that can be fed through the deep learning model.

Model Building

A feed-forward neural network FNN is constructed for this HIV detection project. Such a network has an input layer, multiple numbers of hidden layers, and an output layer. The number of features in the dataset is represented as 10 columns of this input layer with several added hidden layers to capture the presence of complex patterns in the data.

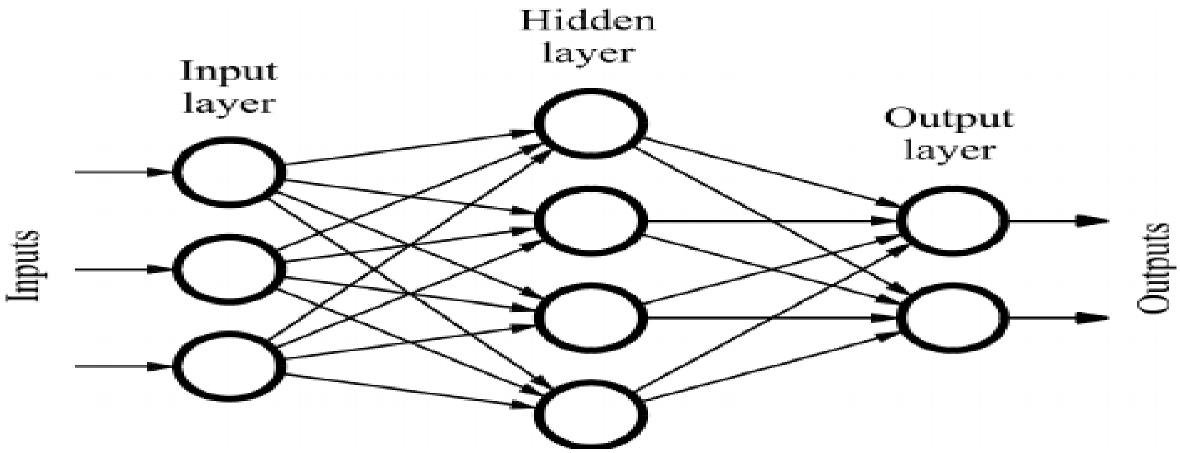


Fig 2.1 FNN layers in the HIV Detection

- The input layer accepts the 10 features from the dataset.
- Every hidden layer will have a specified number of neurons and activation functions such as ReLU or Rectified Linear Unit, to introduce non-linearity. The network may contain multiple hidden layers each of which may, optionally include dropout layers that prevent overfitting by randomly suppressing a fraction of the neurons during training.
- The output layer uses a sigmoid activation for binary classification, where HIV status is to be classified as being either positive or negative.

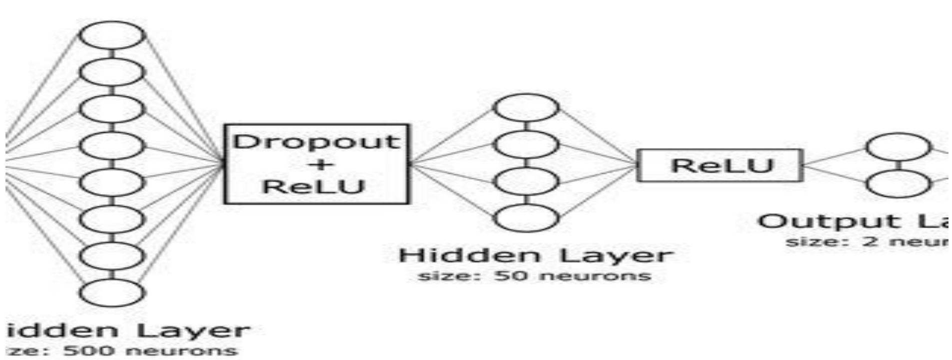


Fig 2.2 FNN along with ReLU in the HIV Detection

Model Training

The dataset is split into training and test sets to test the capability of the model on unseen data. Usual splits are 80-20, meaning 80% for training and 20% is the test set. The FNN model is compiled under a binary cross-entropy loss and the optimizer set as Adam adjusts the learning rate during the training process.

It feeds in the training data batch-wise and iteratively updates the model weights to make the loss function as small as possible. Multiple hyperparameters such as learning rate, batch size, and number of epochs have been tuned during training to achieve optimal performance during the training session.

Using early stopping and model checkpointing stops training once the model performance has stopped improving on a validation set, preventing overfitting.

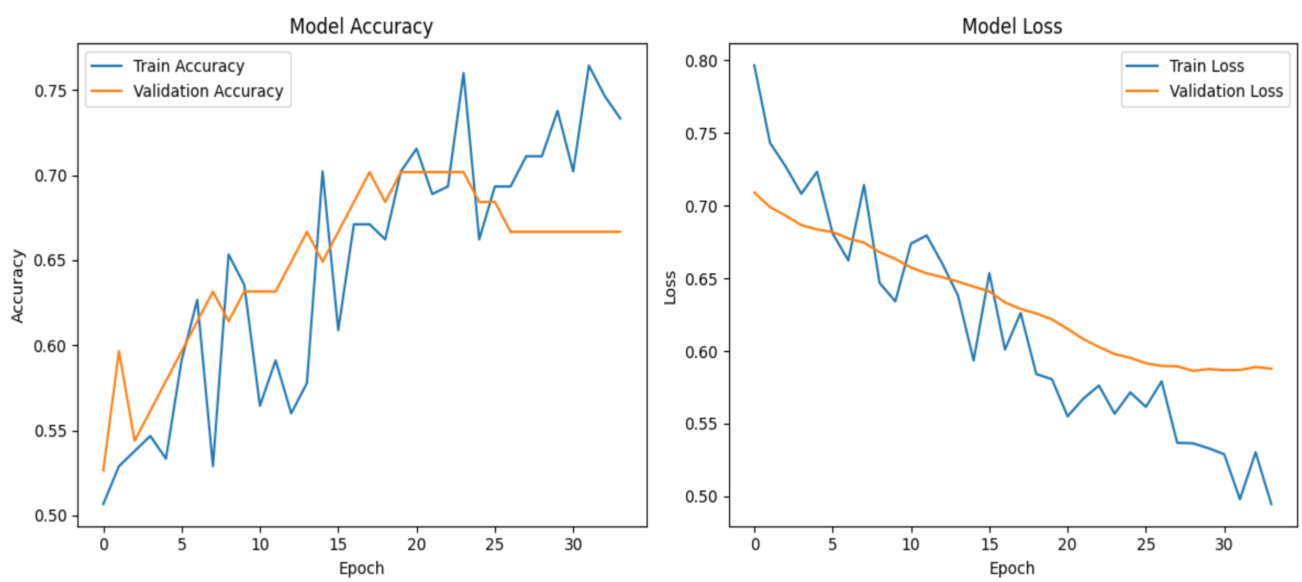


Fig 2.3 Model accuracy and model loss after training the HIV Detection model

Testing

The performance of the FNN model is tested by the testing set after training. Accuracy, precision, recall, F1-score, and AUC-ROC are used to estimate which test the model well. Confusion matrices are also built for graphically true positives, false positives, true negatives, and false negatives.

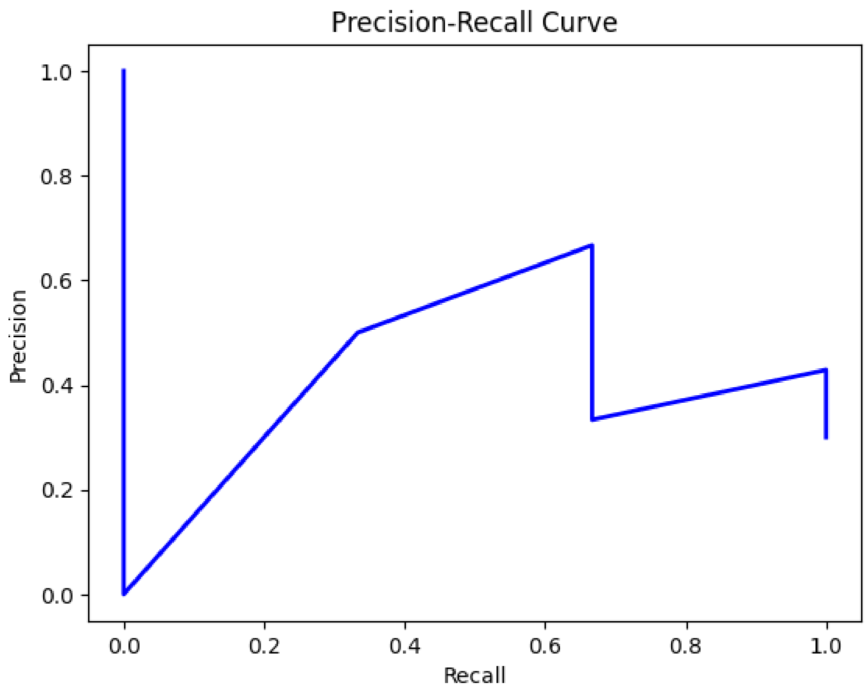


Fig 2.4 Precision-Recall Curve for HIV Detection

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.63	0.67	0.65	33
1	0.69	0.66	0.68	38
Accuracy			0.66	71
Macro avg	0.66	0.66	0.66	71
Weighted avg	0.66	0.66	0.66	71

Table 2.5. Classification Report of HIV Detection

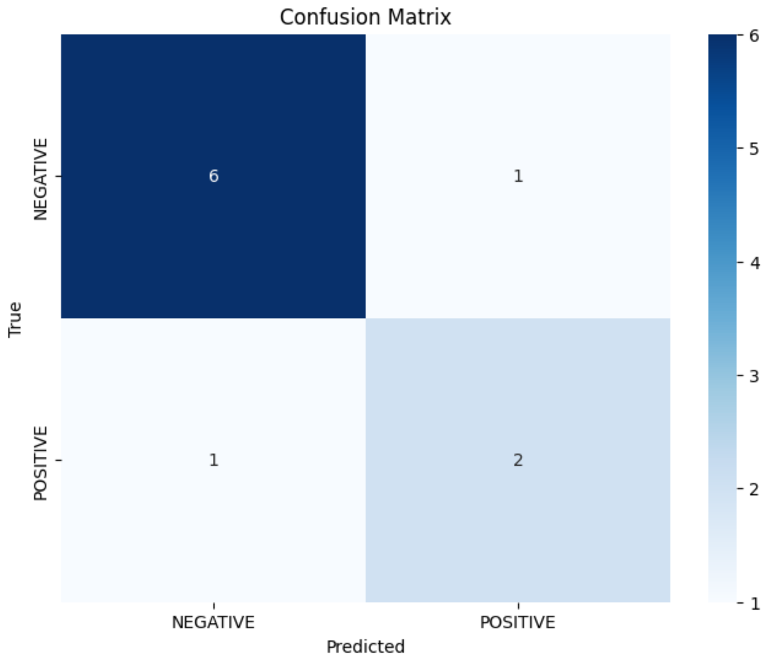


Fig 2.6. Confusion Matrix Analysis for HIV Detection

High sensitivity and specificity are important so that the model classifies positive cases with high accuracy at the same time minimizing false positives and false negatives. The actual labels in the testing set will be compared against the model's predictions to determine its general performance.

Construction of an HIV detection system that relies on a feedforward neural network (FNN) requires structuring the process of preprocessing data and model evaluation.

Properly structured datasets and advanced architectures of neural networks can be used in the improvement of the accuracy and reliability of the diagnosis system, making it more robust and effective for real-world applications. The model remains under constant evaluation and fine-tuning.

3. ALGORITHMS

The model likely uses binary cross entropy as the loss function, which is standard for binary classification tasks.

Binary cross-entropy, also known as log loss, is a crucial loss function used in binary classification problems where the output can take one of two possible values.

In the context of HIV detection using deep learning, binary cross-entropy plays a vital role in training our model to accurately classify whether an individual is HIV-positive or HIV-negative based on the input features.

Binary cross-entropy measures the performance of a classification model whose output is a probability value between 0 and 1. The function calculates the difference between the actual and predicted probabilities of a binary outcome.

Mathematically, it is defined as:

Binary Cross Entropy$=-1/N \sum^{N}_{i==1} [y_i \log(p_i) + (1-y_i) \log (1-p_i)]$

- N is the number of samples.
- y_i is the actual label (1 for HIV-positive and 0 for HIV-negative).
- P_i is the predicted probability of the sample being HIV-positive.

Receiver Operating Characteristic (ROC) is a graphical representation used to evaluate the performance of a binary classification model, such as those employed in deep learning for HIV detection. It illustrates the trade-off between sensitivity (True Positive Rate) and specificity (1 - False Positive Rate) at various threshold settings.

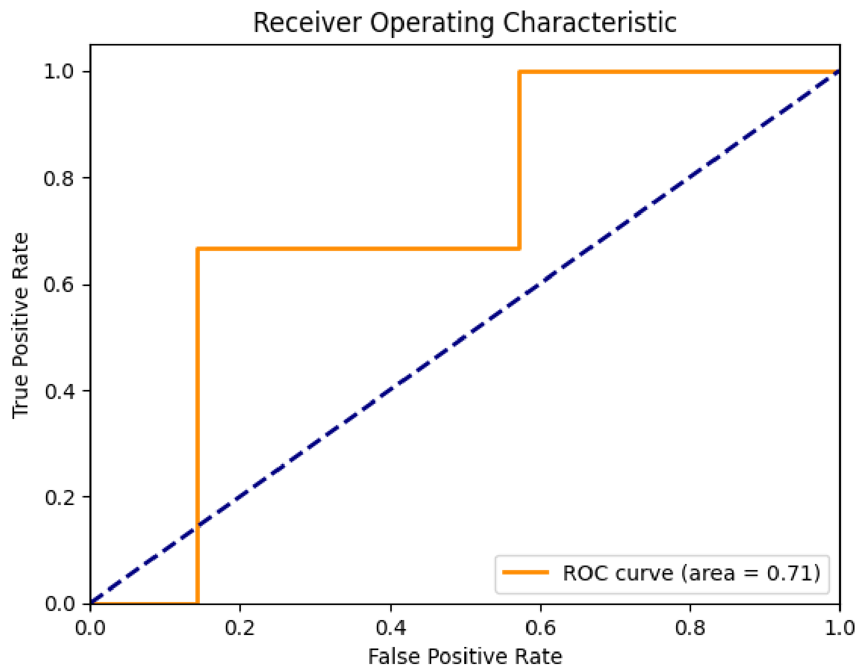


Fig 3.1 Receiver Operating Characteristic for HIV Detection

4. CONCLUSION

Deep learning in HIV Detection has shown enormous potential in the upgradation of the HIV diagnostics domain. Advanced architectures of neural networks have been well incorporated into Feedforward Neural Networks, which ensures that this project has better accuracy and sensitivity to identify cases of HIV infections.

This model comprehensively processes enormous sociodemographic and behavioral data to give a proper assessment of the risk of HIV; hence, the outcomes regarding detection are reliable and precise. One of the highlights of this project is the user-friendly web interface that it develops, which brings in accessibility and scalability, especially in remote and underserved areas.

The real-time capability provides instantaneous feedback, which is considered as utmost importance for effective medical intervention or decision-making. This rapid turnaround is crucial in clinical settings where the need for quick, accurate information directs all possible action for the benefit of patients. It allows for early diagnosis, thus starting the timely antiretroviral therapy, improving the outcome of a patient by not exposing the person-to-person transmission of HIV.

The deep learning-based HIV Detection project presents a new landmark in public health technology. It comes up with an impactful tool in diagnostics on support for high precision, real-time results, and complete usage of data, further expanding and promising better management and reduction in incidence through more ways of managing HIV in the public. Thus, deep learning can be looked at as one major change agent in altering the detection and management of the disease.